A DENOISING APPROACH TO MULTICHANNEL SIGNAL ESTIMATION

Anil M. Rao and Douglas L. Jones

Coordinated Science Laboratory
University of Illinois
1308 W. Main Street
Urbana, IL 61801

ABSTRACT

Multichannel sensor array processing has received considerable attention in many important areas of signal processing. Almost all data recorded by multisensor instruments contain various amounts of noise, and much work has been done in developing optimal processing structures for estimating the signal source from the noisy multichannel observations. The techniques developed so far assume the signal and noise processes are at least wide-sense-stationary so that optimal linear estimation can be achieved with a set of linear, time-invariant filters. Unfortunately, nonstationary signals arise in many important applications and there is no efficient structure with which to optimally deal with them. While wavelets have proven to be useful tools in dealing with certain nonstationary signals, the way in which wavelets are to be used in the multichannel setting is still an open question. Based on the structure for optimal linear estimation of nonstationary multichannel data and statistical models of spatial signal coherence, we propose a method to obtain an efficient multichannel estimator based on the wavelet transform.

1. INTRODUCTION

The estimation of signals in the presence of noise is an important problem in communications and signal processing. The use of measurements taken from an array of receiving sensors or multichannel measurements can considerably enhance signal estimation. In the array setting, beamforming is performed depending on the placement of the sensors with the goal of achieving a gain in the signal power over the noise variance via partially coherent, coherent, or noncoherent combining [1]. In the multichannel setting, the gain in the quality of the signal estimate is due to the properties of the statistical characteristics of the channel (independent noise realizations on each channel or independent fading communication channels, for example). Optimal linear estimation in the sense of minimizing mean square error is achieved by the Wiener filter, which requires the second-order statistics (both temporal and spatial) of the signal and noise. If the signal and noise processes are wide-sense stationary (WSS), then the optimal linear estimator can be realized efficiently with time-invariant, linear filters. Unfortunately, in many situations of interest the WSS assumption is not valid. For example, the desired signal may correspond to a friendly satellite signal in the presence of hostile jammers, and it is very likely that the second-order characteristics may change with time. This can happen because of physical motion, deliberate on-off jamming strategies, or other intentional nonstationary interference generated by the smart opponent [1]. In such situations, estimation techniques based on WSS signal assumptions are clearly inadequate.

Wavelets have rapidly become an indispensable tool in dealing with nonstationary signals in an efficient manner. One of the key properties underlying the success of wavelets is that they form unconditional bases for a large class of signals [4]. Consequently, wavelet expansions tend to concentrate the signal energy into a relatively small number of large coefficients, making it particularly attractive in signal estimation. The use of wavelet transforms in denoising single channel observations is well studied, but there are issues that arise in applying denoising techniques in the case of multichannel or array measurements. Clearly the way in which the wavelet transform is used will depend on how the desired signal component is related between the measurements. The signal component may not be perfectly coherent between sensors; in fact, the signal may have reduced coherence due to the complexity in the propagation of the signal from the source to spatially separated receivers [7].

In this paper we illustrate the use of wavelet-based estimation techniques in the case of multichannel data. We consider several types of multisensor environments including those in which the desired signal (possibly nonstationary) has full, partial, and no coherence between sensors by generalizing a model for partial signal coherence from the area of array signal processing. In Section 2 we briefly review estimation with wavelets and the connection to Wiener filtering in the single channel setting. In Section 3 we discuss optimal linear estimation for partially coherent measurements in the multichannel setting and in Section 4 we propose a method a multichannel denoising method. We show that under certain conditions, the multichannel estimator can be realized efficiently with only discrete Fourier transforms and wavelet transforms.

2. SINGLE CHANNEL WAVELET-BASED ESTIMATION

The standard estimation problem is to recover the discrete-time signal $s(k)$, $k = 1, 2, \ldots, N$ from the noise-corrupted observation

$$x(k) = s(k) + n(k), \quad k = 1, 2, \ldots, N \tag{1}$$

where $n(k)$ is zero-mean white Gaussian noise of variance σ^2. Let $\mathbf{x}$, $\mathbf{s}$, $\mathbf{n}$ denote $N \times 1$ column vectors containing the samples of $x(k)$, $s(k)$, and $n(k)$ respectively, and let $\mathbf{W}$ denote an $N \times N$ orthonormal wavelet transform matrix [2]. In the wavelet domain

This work was supported by the Office of Naval Research, contract no. N00014-95-1-0674

(1) becomes

$$\Theta = \gamma + z \tag{2}$$

with $\Theta = \mathbf{W}\mathbf{x}$, $\gamma = \mathbf{W}\mathrm{s}$, and $\mathrm{z} = \mathbf{W}\mathbf{n}$. Note that an orthonormal wavelet transform will map $\mathbf{n}$ to a z that is likewise zero-mean white Gaussian with variance σ^2 while compacting typical signals s into a small number of large wavelet coefficients in γ. A reasonable approach to wavelet-based signal estimation is to zero out the small entries of Θ which are most likely due to the noise (and where the signal is not) while retaining the large entries which are most likely due to the signal; such an approach is known as wavelet denoising. The motivation for processing the coefficients individually is that the wavelet transform tends to decorrelate the data; that is, it acts like a Karhunen-Loeve (KL) transform. The retain/discard operation can be viewed as a diagonal filtering operation in the wavelet domain. Let us represent the filter by

$$\mathbf{H} = diag\,[h(1), h(2), \ldots, h(N)]\,. \tag{3}$$

The signal estimate based on the wavelet domain filtering is then given by

$$\hat{\mathrm{s}} = \mathbf{W}^{-1}\mathbf{H}\mathbf{W}\mathbf{x}. \tag{4}$$

The wavelet-based signal estimate in (4) can be viewed as an approximation to the optimal linear estimator, the Wiener filter, which employs the signal-dependent KL transform (instead of the general purpose wavelet transform) and uses optimal weighting depending on the signal statistics (instead of simple thresholding) [6].

There are two different flavors for the filter: hard and soft thresholds [4]. The technique described in this paper employs hard thresholding, in which case the filter coefficients are given by

$$h(i) = \begin{cases} 1 & \text{if } |\theta(i)| > \tau \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

There are many choices for choosing the threshold; we will use the "optimal" threshold proposed in [5]:

$$\tau = \sqrt{2\sigma^2 \ln(N)} \tag{6}$$

where N is the number of samples collected and σ^2 is the noise variance.

3. MULTICHANNEL WIENER FILTERING

Let us now consider the case where we have measurements from M spatially separated sensors. Further, let us assume for simplicity that there is only a single signal component and that the sensor observations have been appropriately time-aligned. In a linear array configuration with uniform spacing d, velocity of wave propagation c, and a signal arrival at angle θ, we would need to time shift the ith sensor observation by $D_i = i\frac{d}{c}sin(\theta)$ to time-align the observations. Methods for estimating the angle of arrival are described in [8] in the WSS setting and [9] in the nonstationary setting. Now dealing only with the aligned sensor observations, let us denote the N-sample measurement at the ith sensor by the $x_i(k)$, $k = 1, 2, \ldots, N$. In general, each sensor observation will consist of a component due to the desired signal and a component due to undesired noise; that is, $x_i(k) = s_i(k) + n_i(k)$ where s_i is the signal and n_i is the noise. In many instances, the noise may be modeled as zero-mean white Gaussian noise and is assumed to be uncorrelated between sensors. It is convenient to arrange the sensor measurements in vector form with the stacked $MN \times 1$ column vector

$$\mathbf{x}^T = [x_1^T\; x_2^T \ldots x_M^T] \tag{7}$$

where each x_i is an $N \times 1$ column vector. Similarly, let s and $\mathbf{n}$ denote the $MN \times 1$ column vectors of the signal component and noise component, respectively.

Simultaneously enhancing the signal and suppressing the noise in an optimal fashion by intelligently combining the sensor measurements and applying an appropriate set of filters requires knowledge of the second-order statistics of the signal and noise. While we will not require an exact second-order statistical characterization of the signal and noise, the structure of their correlation matrices will be important. Let the $MN \times MN$ matrix $\mathbf{Q}_s = E[\mathrm{s}\mathrm{s}^T]$ denote the correlation matrix of the desired signal. Since the $MN \times 1$ vector s contains the N-sample signal components as its M blocks, $\mathbf{Q}_s$ can be viewed as an $M \times M$ block matrix with the (i, j)th subblock being the $N \times N$ cross-correlation matrix between the signal component at the ith and jth sensors, which we denote by $\mathbf{R}_{s_i s_j}$. That is,

$$\mathbf{Q}_s = \begin{bmatrix} \mathbf{R}_{s_1 s_1} & \mathbf{R}_{s_1 s_2} & \mathbf{R}_{s_1 s_3} & \cdots & \mathbf{R}_{s_1 s_M} \\ \mathbf{R}_{s_2 s_1} & \mathbf{R}_{s_2 s_2} & \mathbf{R}_{s_2 s_3} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ \mathbf{R}_{s_M s_1} & \cdots & \cdots & \cdots & \mathbf{R}_{s_M s_M} \end{bmatrix} \tag{8}$$

If we assume that the noise is zero-mean Gaussian with variance σ^2 and white both temporally and spatially, the $MN \times MN$ noise correlation matrix across the sensors becomes $\mathbf{Q}_n = E[\mathrm{n}\mathrm{n}^T] = \sigma^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix. The optimal linear estimator $\hat{\mathrm{s}} = \mathbf{G}\mathbf{x}$ that minimizes the mean square error $\mathrm{E}||\mathrm{s} - \hat{\mathrm{s}}||^2$ is the Wiener filter [3]

$$\mathbf{G}_w = \mathbf{Q}_s(\mathbf{Q}_s + \sigma^2\mathbf{I})^{-1}. \tag{9}$$

4. WAVELET-BASED MULTICHANNEL ESTIMATION

As given, the Wiener filter in (9) requires knowledge of both the spatial and temporal correlation structure of the signal. Even if these were known, for nonstationary signals the matrix $\mathbf{Q}_s$ is not Toeplitz-symmetric and there is no efficient, structured implementation for (9). Wavelets have become increasingly popular tools to deal with nonstationary signals in an efficient manner. Recall that a key property of the wavelet transform is that it approximates the KL transform for a large class of signals. Therefore, to see how we may use wavelets in the multichannel setting, we need to examine the spatio-temporal structure of the multichannel KL transform.

The multichannel KL transform is defined through the eigen-expansion of $\mathbf{Q}_s$:

$$\mathbf{Q}_s = \mathbf{U}\Lambda\mathbf{U}^T = \sum_{k=1}^{N_q} \lambda_k \mathrm{u}_k \mathrm{u}_k^T, \tag{10}$$

where $N_q \leq MN$ is the rank of $\mathbf{Q}_s$, $\mathrm{u}_k's$ are the eigenfunctions, and $\lambda_k's$ are the corresponding eigenvalues of $\mathbf{Q}_s$. The $MN \times N_q$ matrix $\mathbf{U}$ denotes the eigenvectors stacked next to each other; that is, $\mathbf{U} = [\mathrm{u}_1, \mathrm{u}_2, \ldots, \mathrm{u}_{N_q}]$. The matrix $\mathbf{U}^T$ is the KL transform for the multichannel signal s since it decorrelates, or diagonalizes, the correlation function $\mathbf{Q}_s$ and concentrates the signal energy into the smallest possible subspace. We may refer to such a KL transform as the spatio-temporal KL transform because it decorrelates the multichannel signal both in time and space.

To provide a structured implementation of the multichannel estimator, we need to examine the interplay between the spatial and temporal contributions to the structure of the spatio-temporal KL transform. Coherence loss will be accounted for by introducing a decorrelation between the time-aligned sensor signals. An exponential power law model has been suggested in the literature [7], whereby the signal cross-correlation matrix between the ith and jth sensors will be scaled by the coefficient

$$c_{ij} = e^{-\frac{|i-j|}{L}}, \tag{11}$$

where L is a dimensionless characteristic correlation length expressed in element spacing units that can be chosen to model a specific environment. That is, $E[s_i s_j^T] = \mathbf{R}_{s_i s_j} = c_{ij}\mathbf{R}_s$ where $\mathbf{R}_s$ is the correlation matrix of the signal source. This intuitively satisfying model gives an exponential decrease in cross-correlation with increasing sensor separation. We will not restrict ourselves to the exponential model; rather, we will allow the decorrelation coefficients to be arbitrary with the only restriction that $c_{ij} \leq 1$. We may arrange the signal decorrelation coefficients in matrix form as $\mathbf{C} = \{c_{ij}\}$. Observe that this model for partial coherence includes as special cases the coherent environment ($c_{ij} = 1\ \forall i, j$) and noncoherent environment ($\mathbf{C} = \mathbf{I}$). With this assumed decorrelation structure, we may write the block form of the correlation matrix of the time-aligned multichannel signal s given in (8) as

$$\mathbf{Q}_s = \begin{bmatrix} c_{11}\mathbf{R}_s & c_{12}\mathbf{R}_s & c_{13}\mathbf{R}_s & \cdots & c_{1M}\mathbf{R}_s \\ c_{21}\mathbf{R}_s & c_{22}\mathbf{R}_s & c_{23}\mathbf{R}_{s_2 s_3} & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ c_{M1}\mathbf{R}_s & \cdots & \cdots & \cdots & c_{MM}\mathbf{R}_s \end{bmatrix} \tag{12}$$

This may be expressed more compactly as a Kronecker product: $\mathbf{Q}_s = \mathbf{C} \otimes \mathbf{R}_s$. We can obtain spectral decompositions for the $M \times M$ spatial decorrelation matrix $\mathbf{C}$ and the $N \times N$ signal source correlation matrix $\mathbf{R}_s$ with

$$\mathbf{C} = \sum_{k=1}^{N_c} \mu_k \mathbf{u}_{c_k}\mathbf{u}_{c_k}^T \tag{13}$$

$$\mathbf{R}_s = \sum_{k=1}^{N_s} \gamma_k \mathbf{u}_{s_k}\mathbf{u}_{s_k}^T \tag{14}$$

where $N_c \leq M$ is the rank of $\mathbf{C}$[1] and $N_s \leq N$ is the rank of $\mathbf{R}_s$.

Let us stack the eigenvectors of $\mathbf{C}$ in the $M \times N_c$ matrix $\mathbf{U}_C = [\mathbf{u}_{c_1}\ \mathbf{u}_{c_2} \ldots \mathbf{u}_{c_{N_c}}]$ and the eigenvectors of $\mathbf{R}_s$ in the $N \times N_s$ matrix $\mathbf{U}_S = [\mathbf{u}_{s_1}\ \mathbf{u}_{s_2} \ldots \mathbf{u}_{s_{N_s}}]$. The matrix $\mathbf{U}_S$ defines the temporal KL transform for the signal source $s(k)$, $k = 1, 2, \ldots, N$. Using elementary properties of Kronecker products [10], the eigenvectors of $\mathbf{Q}_s$ in (10) are given by

$$\mathbf{u}_i = \mathbf{u}_{c_j} \otimes \mathbf{u}_{s_k} \tag{15}$$

for $i = 1, 2, \ldots, N_c N_s$, $j = 1, 2, \ldots, N_c$, $k = 1, 2, \ldots, N_s$. We can also write the stacked matrix of eigenvectors for $\mathbf{Q}_s$ in Kronecker form as $\mathbf{U} = \mathbf{U}_C \otimes \mathbf{U}_S$. This states that the spatio-temporal KL transform may be decomposed via a Kronecker product into the spatial KL transform and the temporal KL transform.

Now it is most likely that the matrix $\mathbf{C}$ (the spatial characteristics of the sensors) is either known *a priori* from characteristics of the medium or can be reliably estimated with transmission of a pilot signal. However, the temporal correlation structure of the signal source is not known in many important problems. In addition, if the signal is nonstationary, then its correlation matrix cannot be estimated from a single realization of the process as can be done in the WSS case. Since the wavelet transform approximately decorrelates and concentrates the signal energy in a relatively small subspace, it can serve as an approximate KL basis for a broad class of signals [2]; that is, we can substitute an appropriate $N \times N$ orthonormal wavelet transform matrix $\mathbf{W}$ for the signal source's temporal KL transform $\mathbf{U}_S^T$.

Recall that the first step in the Wiener filter is to process the multichannel observation $\mathbf{x}$ with the spatio-temporal KL transform for s via $\mathbf{U}^T = \mathbf{U}_C^T \otimes \mathbf{U}_S^T$. Let us substitute the wavelet transform matrix for the signal source's temporal KL basis and denote the transformation with

$$\mathbf{\Theta} = (\mathbf{U}_C^T \otimes \mathbf{W})\mathbf{x} \tag{16}$$

In effect, we spatially decorrelate exactly and temporally decorrelate approximately, allowing us to denoise the multichannel observation $\mathbf{x}$. If we let the set of $N_c \times 1$ column vectors $\{\mathbf{v}_k\}_{k=1}^M$ denote the columns of $\mathbf{U}_C^T$ (the rows of $\mathbf{U}_C$), then recalling the block form of $\mathbf{x}^T = [x_1^T\ x_2^T \ldots x_M^T]$ we may express $\mathbf{\Theta}$ as

$$\mathbf{\Theta} = \sum_{k=1}^{M} \mathbf{v}_k \otimes (\mathbf{W}x_k). \tag{17}$$

Letting u_{ij} denote the (i, j)th element of $\mathbf{U}_C$, we may express $\mathbf{\Theta}$ in the form of N_c blocks as $\mathbf{\Theta} = [\theta_1\ \theta_2 \ldots \theta_{N_c}]$ with each θ_i of size N samples. From (17) we have

$$\theta_j = \mathbf{W}\Big(\sum_{i=1}^{M} u_{ij}x_i\Big)\ j = 1, 2, \ldots, N_c \tag{18}$$

We recognize the term $\sum_{i=1}^M u_{ij}x_i$ as the jth subblock of the matrix product $\mathbf{U}_C^T\mathbf{x}$. Since $\mathbf{U}_C^T$ is a diagonalizing matrix for $\mathbf{C}$, we see from (18) that the structure of the wavelet approximation to the KL transform for s is to first decorrelate the M-channel observation $\mathbf{x}$, producing N_c channels which we then wavelet transform with $\mathbf{W}$. The next step in the wavelet estimation technique is to process each θ_i with a thresholding filter as in (5) and then inverse transform back to the time domain. Knowing the wavelet approximation to the forward KL transform $\mathbf{U}^T$, the structure of the wavelet approximation to the inverse KL transform is straightforward. The structure of the wavelet-based estimator is shown in Figure 1. The question that remains is how to choose the thresholds for the N_c diagonal filters. Recall from (6) that the value of the threshold should be chosen based on the noise variance and the number of samples. Due to the unitary nature of $\mathbf{U}_C^T$, the noise variance on each channel after spatial decorrelation remains at σ^2; therefore, the thresholds on each of the filters should be chosen as in (6).

It is interesting examine how the structure of the wavelet-based estimator simplifies in the cases of a perfectly coherent and completely noncoherent signal between sensors. In the noncoherent

[1] In general $N_c = M$ except in the case of a perfect signal coherence between sensors in which case $N_c = 1$.

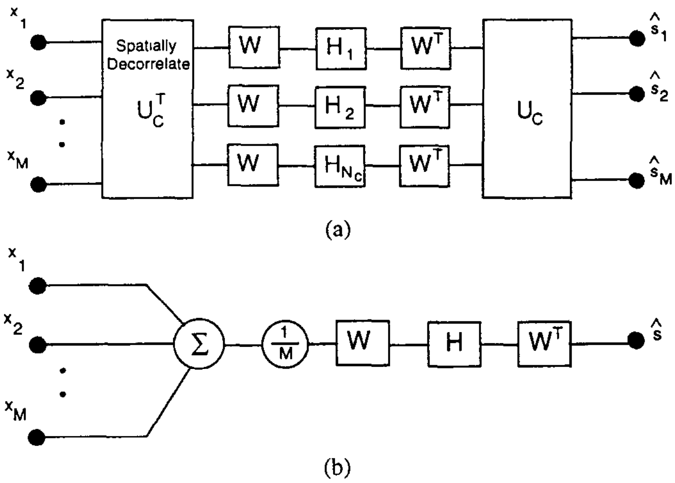

Figure 1: Wavelet-Based Multichannel Estimators: (a) Partially Coherent and Noncoherent ($\mathbf{U}_C = \mathbf{I}$ and $N_C = M$) Environment (b) Coherent Signal Environment

case, the signal component is uncorrelated between the multichannel measurements, implying that $\mathbf{C} = \mathbf{I}$, the $M \times M$ identity matrix. In this case the matrix whose columns are the eigenvectors of $\mathbf{C}$ is also the identity matrix. Therefore, no matrix preprocessing of the sensor observations are required before taking wavelet transforms. In the completely coherent setting, the signal component is statistically identical at each sensor. In this case, $\mathbf{C}$ is a rank-1 matrix of all ones, and the matrix $\mathbf{U}_C$ is an $M \times 1$ column vector where all the elements are equal to $\frac{1}{\sqrt{M}}$. Therefore, spatially decorrelating (and including the scaling factor from the recorrelation) reduces to simply averaging the observations. There is then only one channel to wavelet transform and threshold. These estimation structures are also shown in Figure 1.

It is worth noting an approximation we can make to efficiently realize the first step in our proposed estimator, the spatial decorrelation via $\mathbf{U}_C^T$. For this we first note that $\mathbf{U}_C$ diagonalizes the decorrelation matrix $\mathbf{C}$ which itself is Toeplitz-symmetric. If we assume that $\mathbf{C}$ is a finite-order Toeplitz matrix in the sense that $c_{ij} = 0$ for $|i - j|$ large enough, then it is known that the eigenvalue distribution of such a matrix asymptotically approaches the discrete Fourier transform (DFT) of its first row as the dimensionality M grows large [11]. Physically, this says that if we have a large number of sensors and the decorrelation characteristics are such that there is complete coherence loss beyond a certain sensor separation, then the DFT matrix can be substituted for $\mathbf{U}_C^T$ (and the inverse DFT matrix can be substituted for $\mathbf{U}_C$), approximately decorrelating the sensor observations spatially; the advantage being of course that efficient algorithms exist to compute the DFT and that the exact form of the decorrelation structure between sensors need not be known.

5. SUMMARY

In this paper we proposed an efficient multichannel wavelet-based estimator applicable to nonstationary signals. By examining the structure of the optimal multichannel linear estimator (the multichannel Wiener filter) and models of the spatial coherence of the signal, we identified the contributions of the spatial and temporal signal correlation to the structure of the estimator. Specifically, we found that we were able to decompose the spatio-temporal KL transform for the multichannel signal into the spatial KL transform and the temporal KL transform via a Kronecker product. An appropriate wavelet basis was used to approximate the signal source's temporal KL basis, the advantage being that the wavelet basis is independent of the specific signal statistics and that fast algorithms exist to compute the discrete wavelet transform. Our proposed multichannel estimator first removes spatial redundancy by spatially decorrelating the signal across the sensors which, under certain conditions, can be approximately carried out via a discrete Fourier transform. The resulting decorrelated sensor observations are then each wavelet transformed and passed through a diagonal filter which discards coefficients below a threshold that depends on the noise variance while retaining the coefficients above the threshold. The thresholded coefficients are then inverse wavelet transformed and spatially recorrelated to produce the multichannel signal estimates.

6. REFERENCES

[1] S. Pillai, *Array Signal Processing*. New York: Springer-Verlag, 1989.

[2] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Pretice Hall, 1995.

[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[4] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *App. Comp. Harmonic Anal.*, pp. 100–115, December 1993.

[5] D. Donoho, "De-noising by soft thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613–627, May 1995.

[6] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved Wavelet Denoising via Empirical Wiener Filtering," *Proceedings of SPIE*, San Diego, July 1997.

[7] D. Morgan, "Coherence effects on the detection performance of quadratic array processors, with applications to large-array matched-field beamforming," *Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 737–747, February 1990.

[8] G. Carter, *Coherence and Time Delay Estimation*. New York: Institute of Electrical and Electronics Engineers Inc., 1993.

[9] A. M. Rao, "Optimal Quadratic Array Processing Using Time-Frequency and Time-Scale Representations," submitted to *IEEE Transactions on Signal Processing*, 1998.

[10] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 772–776, September 1978.

[11] R. M. Gray, "On the asympotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, IT-18, pp. 725–730, September 1972.