REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1995	Feb. 1, 1993 - January 31, 1995

4. TITLE AND SUBTITLE

Multiscale Surveillance Information Acquisition and Fusion

5. FUNDING NUMBERS

AFOSR-F49620-93-1-0151

2304/ES

6. AUTHOR(S)

Dr. A. H. Tewfik

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(S)

Dept. of Electrical Engineering
University of Minnesota
Room 4-174 EE/CSci Bldg
Minneapolis, MN 55455

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 95 - 0151

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Dr. Jon Sjogren,
AFOSR/NM
Bldg. 410
Bolling Air Force Base, Washington, D.C.

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The goals of this research effort were to develop robust and adaptive array processing techniques and study adaptive radar waveform selection strategies for optimal range-Doppler imaging in a given time horizon. Specific achievements include the development of two novel wavelet based least mean squares algorithm (LMS) filtering and beamforming procedures. These procedures exploit the sparse structure of the wavelet transform of a wide class of autocorrelation matrices to derive a data dependent transformation of the input process. Their numerical complexities are comparable to those of the fast Fourier transform and discrete cosine transform based LMS techniques. However, their convergence rates are appreciably faster than those of time domain and other transform domain LMS procedures.
Another achievement of this research effort is a new waveform selection strategy for optimal radar range-Doppler imaging. In particular, we have designed an optimal waveform selection strategy for the case where the radar targets to be imaged are known to belong to one of two classes. We have also identified an extension of this approach to the case where the target can belong to more than two classes. A complete study of the numerical and performance issue involved in that extension is currently under way.

14. SUBJECT TERMS

wavelets, radar, SAR, ISAR, beamforming, adaptive algorithms, array processing, adaptive filtering, signal processing

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Technical Report for Grant AFOSR-F49620-93-1-0151: "Multiscale Surveillance Information Acquisition and Fusion"

Ahmed H Tewfik, PI
Department of Electrical Engineering
University of Minnesota, Minneapolis, MN 55455
and
Dr. Jon Sjogren, Program Monitor
AFOSR/NM Bldg. 410
Bolling Air Force Base
Washington, DC 20332-6448

Period covered: Feb. 1993-Jan. 1995

May 1995

1 Executive Summary

The two main goals of this research effort were to study novel approaches to array signal processing and construct an optimal theory of radar range-Doppler imaging. These research goals were met by:

1. *Developing fast, yet computationally efficient, adaptive beamforming techniques that exploit the wavelet domain structure of the observed field.* Adaptive beamforming and filtering algorithms display slow convergence when the correlation matrix of blocks of the input process is highly ill-conditioned. Our new approaches compute an adaptive transformation of the data to ensure that the transformed data has a well-conditioned correlation matrix. The adaptive transformations are computed either via a fast sparse Cholesky factorization of the estimated correlation matrix of the wavelet transform of the input data, or via a sparse QR factorization of the wavelet domain input data matrix.

2. *Constructing a new methodology for optimal design and operation of fast high resolution range-Doppler imaging radars, including synthetic aperture and inverse synthetic aperture radars (SAR and ISAR).* Current radar imaging approaches are based on a radar mode of operation that is designed for the detection of point targets. As a consequence, images obtained in this mode of operation are filtered low resolution rendering of the actual range-Doppler reflectivity map. We have shown that for optimal imaging of a given target range-Doppler reflectivity function with N waveforms, one must use any N waveforms that form an orthogonal basis for a space of waveforms that is matched to the target range-Doppler reflectivity function. We have developed a novel optimal waveform selection strategy for the case where the radar targets to be imaged are known to belong to one of two classes. We have also identified an

extension of this approach to the case where the target can belong to more than two classes. A complete study of the numerical and performance issue involved in that extension is currently under way.

In the remainder of this report we provide a brief description of the results that we have obtained under this grant. These results fall into two distinct areas: fast wavelet domain adaptive filtering procedures and waveform selection for optimal radar range-Doppler imaging. More detailed descriptions of the results that we review here may be found in a number of publications that have appeared in the literature or are currently under review or preparation. See, e.g., [1], [2], [3], [4].

2 Wavelet domain adaptive FIR filtering and beamforming

Adaptive Finite Impulse Response (FIR) filters are widely used in a variety of signal processing applications such as adaptive noise cancelation, channel equalization, system identification and echo cancelation. Typically in these applications, filters are implemented in the time domain as a tapped delay line filter. The filter lengths are usually chosen to be very large, to achieve the desired level of performance [5]. An alternative to using such long adaptive FIR filters is to implement adaptive Infinite Impulse Response (IIR) filters. Though IIR filters require a much smaller number of coefficients than FIR filters, adaptive IIR filters are not widely used due to problems of instability, slow convergence and local minima [5].

The most popular algorithm used to adapt the parameters of FIR filters is the Widrow-Hopf LMS algorithm. Its popularity is due to its low computational complexity and its robustness to implementation errors. Consider an N tap adaptive filter. Let $h_k(n)$ be the k^{th} tap of the filter at time n. Denote by $x(n)$, $d(n)$ and $y(n)$ the input to the filter, the desired response and the output of the adaptive filter at time n respectively. Furthermore, denote by $e(n)$ the output error, i.e., $e(n) = d(n) - z(n)$. Note that $z(n)$ is given by

$$z(n) = \mathbf{x}^T(n)\mathbf{h}(n). \tag{2.1}$$

Here $\mathbf{x}(n)$ is the input signal vector and $\mathbf{h}(n)$ is the adaptive weight vector, i.e.,

$$\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T, \tag{2.2}$$

$$\mathbf{h}(n) = [h_0(n), h_1(n), \cdots, h_{N-1}(n)]^T, \tag{2.3}$$

where $(\cdot)^T$ denotes the transpose operator. The standard LMS algorithm updates $\mathbf{h}(n)$ using the recursion [6]

$$\mathbf{h}(n+1) = \mathbf{h}(n) - 2\mu\hat{\nabla}_x = \mathbf{h}(n) + 2\mu\mathbf{x}(n)e(n), \tag{2.4}$$

where μ is called the step size.

The convergence properties of the LMS algorithm are well known [6], [7]. In particular, the main drawback is its convergence speed. Let $\mathbf{R}_x$ be the autocorrelation matrix of the input vector $\mathbf{x}(n)$, i.e.,

$$\mathbf{R}_x = E[\mathbf{x}(n)\mathbf{x}^T(n)].$$

For a fixed step size μ, the convergence speed of the LMS algorithm depends on the condition number, $\chi(\mathbf{R}_x)$, of this input autocorrelation matrix [7]. When all the eigenvalues of the input correlation matrix are equal, i.e., the condition number $\chi(\mathbf{R}_x) = 1$, the algorithm converges fastest. As the eigenvalue spread $\chi(\mathbf{R}_x)$ increases (i.e., as the input correlation matrix becomes more ill conditioned), the algorithm converges more slowly.

To improve the convergence properties of the algorithm, we can transform the input vector so that the input correlation matrix in the transformed domain has a lower eigenvalue spread. In particular, if $\mathbf{R}_x$ is non-singular, we can whiten the input process $\mathbf{x}(n)$ by pre-multiplying $\mathbf{x}(n)$ with the matrix $\mathbf{G}_x^{-1}$, where $\mathbf{G}_x$ is the lower triangular Cholesky factor of the autocorrelation matrix $\mathbf{R}_x$, i.e.,

$$\mathbf{G}_x\mathbf{G}_x^T = \mathbf{R}_x. \tag{2.5}$$

This is equivalent to using an LMS-Newton type algorithm [6]. However, the main drawback of this approach is that it requires $O(N^2)$ flops and therefore is computationally expensive.

Whitening, of course, cannot be used when $\mathbf{R}_x$ is singular. In such a case, we can use an alternative procedure that we describe in [8].

The increase in computational complexity due to whitening may be alleviated if the autocorrelation matrix $\mathbf{R}_x$ has some special sparse structure. In particular, if the correlation matrix is diagonal, this increase is minimal. This observation has motivated the development of a variety of transform domain algorithms, e.g., [9] - [11] (for a more complete list of work in transform domain filtering refer to [5]), and more recently [12] - [14]. By properly transforming the input prior to the implementation of the LMS algorithm, one hopes that the input to the adaptive FIR filter would have a nearly diagonal correlation matrix in the transform domain. A LMS-Newton type filter can then be implemented in the transform domain with much smaller computational burden and improved convergence performance. For example, if the original autocorrelation matrix is Toeplitz with Hermitian symmetry, when the matrix size is large (large filter lengths), it can be approximated by a circulant matrix [10]. Therefore, a discrete Fourier transform (DFT) (implemented using fast Fourier transform (FFT) algorithms) or a discrete cosine transform (DCT), will nearly diagonalize the correlation matrix. The cost of such a transformation is a minimal $O(N \log N)$ flops. Adaptive filtering in the sub-bands has also been discussed to improve the convergence speed with low computational burden [15].

It should be noted that the use of these various orthogonal transforms will not result in an optimal convergence speed for all types of input signals. The degree to which a given transformation succeeds in near-diagonalizing an autocorrelation matrix obviously depends on the type of input signal. For example, the DCT is known to approximate the Karhunen-Loéve Transform (KLT) when the signal is a low pass process but is known to perform poorly when the signal is a high pass process [11], [16].

We developed two novel *fast* wavelet domain Newton type LMS algorithms. These algorithms do *not* attempt to diagonalize the correlation matrix of the input process. Rather, they exploit the special sparse structure of the wavelet transform of wide classes of correlation matrices and that of the Cholesky factors of these matrices [17]. Specifically, the algorithms identify a whitening transformation in the wavelet domain to minimize computational complexity. As a result, they take fewer computations to identify this transformation than to compute a KLT or update an estimate of the inverse of the input covariance matrix. The procedures estimate *structured* estimates of the wavelet domain correlation matrix of the input process. They differ in the exact estimates they use and in the way they identify the data dependent transformation. The first approach explicitly computes a sparse estimate of the wavelet domain correlation matrix of the input process. It then computes the Cholesky factor of this matrix and uses its inverse to whiten the input. The complexity of this approach is $O(N \log^2(N))$. In contrast, the second approach computes a sparse approximation to the Cholesky factor of the wavelet domain correlation matrix of the input process *directly*. It does so by computing a sparse QR decomposition of the wavelet domain data matrix. While this approximation does not seem to converge to the true sparse Cholesky factor of the wavelet domain input covariance matrix, it is effective in improving the condition number of the

input covariance matrix. The complexity of this second approach is $O(N\log(N))$. Its computational complexity is, therefore, no higher than that of the DCT or the FFT based transform domain LMS algorithms. Yet, its performance is vastly superior to that of these latter algorithms. Thus, both techniques lead to low complexity LMS algorithms with a convergence performance similar to that of the more complex KLT based or Newton type LMS algorithms.

Note that the discrete wavelet transform (DWT) provides a good approximation to the KLT of several processes, [17] - [18]. The sparse structure of the DWT of wide classes of matrices has been used to speed up matrix-matrix and matrix-vector multiplications [19], [20] and positive definite matrix inversion [17]. They have also been used to derive fast estimation procedures [21].

LMS type algorithms in the wavelet transform domain have been suggested recently [12] - [14]. However, these algorithms do *not* exploit the structure of the wavelet transform of a correlation matrix. They make the assumption that the correlation matrix in the wavelet transform domain is nearly diagonal. Unfortunately, this is not always the case (see Section 5 in [1] for examples). Therefore, the behavior of these algorithms would be similar to that of other transform domain LMS algorithms. In particular, our study (c.f. [1]) shows that the proposed DWT LMS algorithms which exploit the full structure of the wavelet domain correlation matrices outperform these algorithms as well as other transform domain algorithms that rely on near-diagonalizing the autocorrelation matrix. Furthermore, they are more robust than the other transform domain algorithms: their convergence behavior is not sensitive to changes in the values of the parameters of the correlation matrix of the input process.

2.A Wavelet transform domain adaptive filtering

Let us now briefly describe the two new fast wavelet domain techniques for whitening the input data. Refer to [1] for a more complete discussion of these procedures and an analysis of the effect of the approximations that we make on the performance of the resulting wavelet domain LMS algorithms. These two techniques rely on the special sparse structure of the wavelet transform of wide classes of correlation matrices and that of the Cholesky factors of these matrices [17].

Specifically, suppose that we compute an M-band wavelet decomposition of an input stochastic process $x(n)$. Denote by $R_w^{j,k}(l,m)$ the cross-correlation between the l^{th} component of the stochastic process in the j^{th} band and the m^{th} component of the process in the k^{th} band. It is shown in [17] that under very mild conditions on the covariance of $R(l,m)$ of $x(n)$, $|R_w^{j,k}(l,m)|$ decays asymptotically to zero at a rate much faster than $|R(l,m)|$. In particular, denote by $\mathbf{R}_y$ the autocorrelation matrix of the wavelet transform $\mathbf{y}(n)$ of the vector $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-N+1)]^T$. Let $\hat{\mathbf{R}}_y$ be the approximation to $\mathbf{R}_y$ that is obtained by setting all entries of $\mathbf{R}_y$ of magnitude less than a small constant δ_R to zero. Furthermore, let $\mathbf{R}_y^{j,k}$ denote the sub-block of $\mathbf{R}_y$ that represents the cross covariance between the outputs of filters $w_j(n)$ and $w_k(n)$. The results of [17] indicate that each sub-block $\hat{\mathbf{R}}_y^{j,k}$ of $\hat{\mathbf{R}}_y$ is banded. Furthermore, each row or column of $\hat{\mathbf{R}}_y$ will have $O(\log N)$ non-zero elements.

It is also shown in [17] that the Cholesky factor of $\hat{\mathbf{R}}_y$ has the same sparse banded structure as that of the original matrix (for this proof and the choice of δ_R, refer to [17]). The banded structure and the bandwidths of $\hat{\mathbf{R}}_y(k)$ depend on the wavelet transform filters used and the input process characterization.

The class of processes for which the above results hold true is quite large. For example it includes all stationary processes that can be obtained by passing white noise through a rational filter. The covariance functions of such processes decay exponentially fast to zero for fixed n as m tends to infinity. It also includes nonstationary process such as a sampled version of a fractional Brownian motion [22], [23]. The fractionally differenced white noise process [24], [25] and samples

of the underwater background acoustic noise in shallow water taken along any straight line [26] are other examples of nonstationary processes in that class.

2.A.1 Fast Wavelet domain LMS algorithm via Cholesky factorization of an estimated covariance matrix

Denote by $\mathbf{y}(n) = \mathbf{Q}\mathbf{x}(n)$ the $N \times N$ discrete wavelet transform of $\mathbf{x}(n)$. We update the transformed weight vectors $\mathbf{w}(n) = \mathbf{Q}\mathbf{h}(n)$, in the proposed algorithm using the equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\hat{\mathbf{g}}(n). \tag{2.6}$$

Here,

$$\hat{\mathbf{R}}_y(n)\hat{\mathbf{g}}(n) = \hat{\nabla}_y(n), \tag{2.7}$$

where $\hat{\mathbf{R}}_y(n)$ is a sparse estimate of $\mathbf{R}_y$ at step n. In particular, we compute $\mathbf{R}_y$ by estimating *only the expected non-negligible entries* of $\mathbf{R}_y$. All other entries of $\hat{\mathbf{R}}_y(n)$ are set to zero (see below). $\hat{\nabla}_y(n)$, the instantaneous gradient, is given by

$$\hat{\nabla}_y(n) = -2e(n)\mathbf{y}(n), \tag{2.8}$$

where $e(n)$ is output the error given as $e(n) = d(n) - \mathbf{w}(n)^T\mathbf{y}(n)$.

We now use the sparse Cholesky factorization of $\hat{\mathbf{R}}_y(n)$ to solve (2.7) by rewriting it as

$$\hat{\mathbf{L}}_y(n)\hat{\mathbf{D}}_y(n)\hat{\mathbf{L}}_y^T(n)\hat{\mathbf{g}}(n) = \hat{\nabla}_y(n). \tag{2.9}$$

We can solve the above equation in three steps as

$$\hat{\mathbf{L}}_y(n)\mathbf{a} = \hat{\nabla}_y(n), \tag{2.10}$$

$$\hat{\mathbf{D}}_y(n)\tilde{\mathbf{a}} = \mathbf{a} \quad \text{and} \tag{2.11}$$

$$\hat{\mathbf{L}}_y^T(n)\hat{\mathbf{g}}(n) = \tilde{\mathbf{a}}. \tag{2.12}$$

Since $\hat{\mathbf{D}}_y(n)$ is a diagonal matrix and $\hat{\mathbf{L}}_y$ has at most $\log N$ non-zero element per row [17], evaluating $\hat{\mathbf{g}}(n)$ using this approach requires $O(N \log N)$ operations only.

In order to use the above algorithm, we need to:

1. compute the wavelet transform $\mathbf{y}(n)$ of $\mathbf{x}(n)$,
2. maintain and update the sparse matrix approximate $\hat{\mathbf{R}}_y(n)$ of the wavelet domain correlation matrix,
3. compute the approximate Cholesky factor $\hat{\mathbf{L}}_y(n)$ of $\hat{\mathbf{R}}_y(n)$,
4. solve equations (2.10)-(2.11).

Computing the wavelet transform $\mathbf{y}(n)$ of $\mathbf{x}(n)$ is inexpensive: it requires only $O(N)$ operations.

To compute $\hat{\mathbf{R}}_y$ we proceed as follows. Note that the correlation matrix $\mathbf{R}_y$ consists of sub-blocks $\mathbf{R}_y^{j,k}$, where $\mathbf{R}_y^{j,k}$ denotes the cross correlation matrix of the output of the j^{th} band and that of the k^{th} band. From the discussion of [17], the elements of $\mathbf{R}_y^{j,k}$ away from its main diagonal decay to zero faster than $R_x(l,m)$. The sparse approximate $\hat{\mathbf{R}}_y^{j,k}$ of $\mathbf{R}_y^{j,k}$ obtained by setting all elements of $\mathbf{R}_y^{j,k}$ below a threshold δ_R to zero, is therefore a band diagonal matrix. We form $\hat{\mathbf{R}}_y$ by replacing $\mathbf{R}_y^{j,k}$ in $\mathbf{R}_y$ by their corresponding sparse band diagonal approximates $\hat{\mathbf{R}}_y^{j,k}$.

Given the threshold δ_R, the decimation factor M, the type of the characterization of the input process (partial knowledge), we can determine the rate of decay of the elements of $\mathbf{R}_y^{j,k}$, away from its main diagonal (c.f. [17]). This in turn determines the approximate bandwidths of the matrices $\hat{\mathbf{R}}_y^{j,k}$. With this in mind we define a sparsity map $\mathbf{S}$ such that its $(l,m)^{th}$ element $S(l,m)$ is given by

$$S(l,m) = \begin{cases} 1 & l = m \text{ or } R_y(l,m) \text{ is predicted to be significant} \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Specifically, we choose $S(l,m) = 1$ whenever entry (l,m) is "near" the main diagonal of one of the matrices $\mathbf{R}_y^{j,k}$. For the type of input process in our experiments, we have chosen $S(l,m) = 1$ for each l whenever entry (l,m) was three or four entries or less away from the main diagonal of a matrix $\mathbf{R}_y^{j,k}$ and some other elements, which give these bands a jagged appearance. The width of these bands and the exact jagged nature depend on the type of input process.

We can now develop a method for updating $\hat{\mathbf{R}}_y(n+1)$ at time $n+1$ in terms of its estimate at time n, by updating only those entries for which $S(l,m)$ is non zero. The update equation for each element of $\hat{\mathbf{R}}_y(n+1)$ is given as

$$\hat{R}_y(l,m:n+1) = \begin{cases} (1-\gamma_n)\hat{R}_y(l,m:n) + \gamma_n y_l(n) y_m(n), & S(l,m) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

where $R_y(l,m:n)$ denotes the $(l,m)^{th}$ element of $\mathbf{R}_y(n)$ and γ_n is a weighting factor. Parameter γ_n is usually chosen to be a constant < 1 independent of n or as $\frac{1}{n}$. It allows us to "forget" past input values.

Note that we update only those elements of $\mathbf{R}_y(n)$, which are significant. Since $\hat{\mathbf{R}}_y$ has only $O(N \log N)$ non-negligible elements, this update also requires $O(N \log N)$ operations. Our up-date scheme is *not* equivalent to updating $\hat{\mathbf{R}}_y(n)$ by an outer product. Hence, the update of $\hat{\mathbf{R}}_y(n)$ is *not a simple rank one update.*

Solving (2.10)-(2.11) requires $O(N \log N)$ operations since $\hat{\mathbf{L}}_y(n)$ has at most $O(\log N)$ non-zero element per row or column.

The lower triangular matrix $\hat{\mathbf{L}}_y(n)$ and the diagonal matrix $\hat{\mathbf{D}}_y(n)$ may be computed using the iteration [27]:

For $k = 1, 2, \cdots, N$

$$\hat{D}_y(k,k:n) = \hat{R}_y(k,k:n) - \sum_{p=1}^{k-1} \hat{L}_y^2(k,p:n)\hat{D}_y(p,p:n) \tag{2.15}$$

For $i = k+1, \cdots, N$

$$\hat{L}_y(i,k:n) = (\hat{R}_y(i,k:n) - \sum_{p=1}^{k-1} \hat{L}_y(i,p:n)\hat{L}_y(k,p:n))/\hat{D}_y(k,k:n), \tag{2.16}$$

where $\hat{R}_y(i,j:n)$ denotes the $(i,j)^{\text{th}}$ element of $\hat{\mathbf{R}}_y(n)$. The computation in each sub-block proceeds until an element of magnitude ϵ or less is computed, where ϵ is a pre-selected threshold (see [17] for guidance on how to select ϵ). At this point all remaining elements in the column being computed within that sub-block are set to zero. Since both $\hat{\mathbf{R}}_y(n)$ and $\hat{\mathbf{L}}_y(n)$ have at most $O(\log N)$ non-zero element per row or column, this iteration requires $O(N \log^2 N)$ operations. Therefore, the main computational bottleneck of our algorithm is the computation of $\hat{\mathbf{L}}_y(n)$ from $\hat{\mathbf{R}}_y(n)$.

In the initial stages of the algorithm, the matrix $\hat{\mathbf{R}}_y(n)$ may not be positive definite, this would then imply that the direction calculated using equations 2.10-2.12 might not be the descent direction. This problem has been studied in conjunction with the Newton method for optimization,

where the Hessian might turn out to be indefinite or singular. Several modified Newton methods have been proposed in literature, which modify the Hessian whenever the Hessian is not positive definite and reverting to the Newton directions as the minimum is approached, thus taking advantage of the faster convergence of Newton's method.

An early approach to this problem were the directional discrimination methods (cf. [28]). However these methods involve an eigenvalue decomposition. Another way of modifying the Hessian approximate, $\hat{\mathbf{R}}_y$ is to add another positive definite matrix to it. The simplest positive definite matrix that could be added to $\hat{\mathbf{R}}_y(n)$ is a small diagonal matrix, $\kappa_n \mathbf{I}$ [29], [30]. κ_n is chosen to ensure that $\hat{\mathbf{R}}_y(n)+\kappa_n\mathbf{I}$ is positive definite. Depending on the size of κ_n, the method behaves more like the simple LMS ($\kappa_n \rightarrow \infty$) or LMS-Newton type method ($\kappa_n \rightarrow 0$). It is thus desirable that κ_n become zero as the minimum is approached to take advantage of the fast convergence properties of the LMS-Newton method. In practice, κ_n is chosen as $\frac{c}{n}$ where c is a constant for the first 100 to 200 iterations till the estimate for $\hat{\mathbf{R}}_y(n)$ becomes sufficiently positive definite after which we can discontinue diagonal loading. This assures that the addition of $\kappa_n\mathbf{I}$ does not contribute to the excess MSE, and the speed of convergence is not affected much.

Another method of modifying the Hessian makes use of the of the fact that during the Cholesky $(\hat{\mathbf{L}}_y(n)\hat{\mathbf{D}}_y(n)\hat{\mathbf{L}}_y^T(\mathrm{n}))$ decomposition of $\hat{\mathbf{R}}_y(n)$ one or more diagonal elements of $\hat{\mathbf{D}}_y(n)$ become non-positive whenever the $\hat{\mathbf{R}}_y(n)$ becomes nonpositive definite. A simple modification to the Cholesky decomposition procedure then produces a Cholesky factorization of $\hat{\mathbf{R}}_y(n) + \mathbf{\Phi}(n)$, where $\mathbf{\Phi}(n)$ is diagonal with positive diagonal elements. This factorization is computed in a single pass and avoids the need for several trial $\mathbf{\Phi}(n)$ matrices (the interested reader is referred to [31]).

Another simple engineering modification suggests that whenever $\hat{\mathbf{R}}_y(n)$ becomes non-positive definite, the Cholesky factorization is stopped (as soon as one encounters a negative diagonal element in $\hat{\mathbf{D}}_y(n)$) and for that iteration, the descent direction is taken to be the same as that of a standard transform domain LMS algorithm. In practice such a modification affects the rate of convergence very little and it remains nearly the same as that of the unmodified LMS-Newton algorithm with the exact positive definite $\mathbf{R}_y(n)$.

2.A.2 Fast Wavelet domain LMS algorithm via direct estimation of the Cholesky factor of the input covariance matrix

The main computational bottleneck of the algorithm discussed above is its Cholesky factorization. Also, in the initial stages of the algorithm, the matrix $\hat{\mathbf{R}}_y(n)$ may not be positive definite. We therefore need a procedure that computes sparse triangular factors of $\hat{\mathbf{R}}_y(n)$ at a lower computational cost and also works in situations where the matrix $\hat{\mathbf{R}}_y(n)$ may not be positive definite. Such an algorithm may constructed using a QR decomposition of the transform domain data matrix.

Consider the data matrix $\mathbf{Y}(n)$ obtained by stacking up the weighted input vectors of the adaptive filter from time 0 to n as rows

$$\mathbf{Y}(n) = \begin{pmatrix} \sqrt{\gamma(1-\gamma)^n}\mathbf{y}^T(0) \\ \sqrt{\gamma(1-\gamma)^{(n-1)}}\mathbf{y}^T(1) \\ \vdots \\ \sqrt{\gamma}\mathbf{y}^T(n) \end{pmatrix} \doteq \begin{pmatrix} \sqrt{(1-\gamma)}\mathbf{Y}(n-1) \\ \sqrt{\gamma}\mathbf{y}^T(n) \end{pmatrix}, \tag{2.17}$$

where $\gamma < 1$ is the forgetting factor. Let us compute the upper triangular factor $\mathbf{U}(n)$ of $\mathbf{Y}(n)$ by performing a QR decomposition of $\mathbf{Y}(n)$. That factor converges (up to a sign difference in the

rows) to the Cholesky factor of the correlation matrix $\mathbf{R}_y(n)$ defined as

$$\mathbf{R}_y(n) = \sum_{k=0}^{n} \gamma(1-\gamma)^{n-k}\mathbf{y}(k)\mathbf{y}^T(k) \doteq (1-\gamma)\mathbf{R}_y(n-1) + \gamma\mathbf{y}(n)\mathbf{y}^T(n). \tag{2.18}$$

Thus, it is reasonable to estimate $\hat{\mathbf{R}}_y(n)$ as $\hat{\mathbf{U}}_y^T(n)\hat{\mathbf{U}}_y(n)$, where $\hat{\mathbf{U}}_y(n)$ is a structured "pseudo" sparse upper triangular factor of the data matrix. In particular, we can update the structured upper triangular factor at every iteration rather than recompute it as in the previous algorithm. The structured upper triangular factor that we would obtain will be close to the upper triangular factor obtained by an exact QR decomposition of $\mathbf{Y}(n)$ but is *not* equal to this exact factor.

The update equations for the transformed weight vectors can now be written as

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\hat{\mathbf{g}}(n), \tag{2.19}$$

where,

$$\hat{\mathbf{U}}_y^T(n)\mathbf{U}_y(n)\hat{\mathbf{g}}(n) = \hat{\nabla}_y(n). \tag{2.20}$$

The instantaneous gradient $\nabla_y(n)$ is obtained as in the previous algorithm. The above equation can now be solved in two steps as

$$\hat{\mathbf{U}}_y^T(n)\mathbf{a} = \hat{\nabla}_y(n), \tag{2.21}$$

$$\hat{\mathbf{U}}_y(n)\hat{\mathbf{g}}(n) = \mathbf{a}. \tag{2.22}$$

Since, $\hat{\mathbf{U}}_y(n)$ is sparse having at most $O(\log N)$ non-zero elements per row, evaluating $\hat{\mathbf{g}}(n)$ using this approach requires $O(N \log N)$ operations.

We can compute and maintain the sparse "pseudo" upper triangular factor $\hat{\mathbf{U}}_y(n)$ in $O(N \log N)$ flops. Specifically, recall that the upper triangular factor $\mathbf{U}_y(n)$ obtained by performing a QR decomposition of the data matrix $\mathbf{Y}(n)$ converges (up to a sign difference in the rows) to the Cholesky factor $\mathbf{L}_y^T$. Therefore, the sparsity structure of both the matrices is the same. With this observation in mind, we define a new sparsity map, $\mathbf{S}$, similar to the one described in the previous section. The $(l,m)^{th}$ element of this sparsity map, $S(l,m)$, is obtained as

$$S(l,m) = \begin{cases} 1 & l = m \text{ or } L_y(m,l) \text{ is predicted to be significant} \\ 0 & \text{otherwise.} \end{cases} \tag{2.23}$$

The matrix $\hat{\mathbf{U}}_y(n)$ can now be easily updated. This is done by appending the input vector $\hat{\mathbf{y}}(n)$ as a new row to $\hat{\mathbf{U}}_y(n-1)$ and zeroing this row by appropriate Givens rotations [27]. The k^{th} Givens rotation is a multiplication by an orthogonal matrix of the form

$$\Psi_k = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & c_k & \mathbf{0} & s_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -s_k & \mathbf{0} & c_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where $c_k^2 + s_k^2 = 1$ and the $\mathbf{I}$'s represent identity matrices of appropriate dimensions. This operation can be expressed as

$$\begin{pmatrix} \hat{\mathbf{U}}_y(n) \\ 0 \end{pmatrix} = \mathbf{Z}(n) \begin{pmatrix} \sqrt{(1-\gamma)}\hat{\mathbf{U}}_y(n-1) \\ \sqrt{\gamma}\mathbf{y}^T(n) \end{pmatrix}, \tag{2.24}$$

$\mathbf{Z}(n)$ is the unitary matrix denoting the effect of the sequence of Givens rotations.

The zeroing procedure takes into account the sparse structure of $\hat{\mathbf{U}}_y(n)$ and proceeds as follows: we annihilate the first element of the added row by applying a Givens rotation to operate on the first and last row of the modified matrix. During this update, only the elements of the first row that are predicted to be non-negligible and their corresponding entries in the last row are updated. All others in the first row are left as zeros. The elements of the last row which have not yet been updated have to be multiplied by the gain c_1. However, this multiplication is not performed right away. A table is constructed to keep track of the elements which require this multiplication. The gain c_1 is also stored. Next, the second element in the last row of the transformed matrix is annihilated by applying the Givens rotation to operate on the second and last rows. The second element of the last row is first multiplied by c_1 in order to calculate c_2 and s_2. The previous procedure is repeated by updating only the non-zero elements in the second row and their corresponding elements in the last row after multiplying these by the gain c_2c_1. The element locations which are not updated at this stage and the gains they have to be multiplied with are entered into the table. Thus, at each rotation, the locations of the elements in the last row which are not updated and their corresponding gains are stored in the table. The gains are a sequence of multiplications by c_k's and can be either computed or updated from previously calculated gains. The procedure which requires fewer number of operations is chosen. As each row has at most $O(\log N)$ elements, only $O(\log N)$ operations are performed to update these elements. As the gains are computed using the fewest number of flops and due to the special sparse structure, the total number of computations required to compute all the required gains is $O(N \log N)$. The entire procedure therefore requires only $O(N \log N)$ computations.

The update process can be best explained by looking at an example. Consider an 8×8 upper triangular matrix with the sparsity structure as shown in Fig. 1. Note, that the matrix considered in this example is a banded matrix with unity bandwidths. Also, the transform chosen is a dyadic WT. This sparsity structure corresponds with that of the thresholded covariance matrix in the WT domain, $\hat{\mathbf{R}}_y$, with diagonal sub-blocks $\hat{\mathbf{R}}_y^{j,k}$. The x's in the figure indicate non-zero elements of the matrix and the r's indicate the new row. In the first Givens rotation, only the non-zero elements in the first row are updated. The first r of the new row is replaced by a zero, and the elements of this row which are affected by the non-zero elements of the first row of the matrix are updated. The remaining elements of the new row have to be multiplied by c_1 in order that the sequence of rotations remain orthogonal. The elements which are not updated are then entered into the table along with the gain they need to be multiplied with. Table 1 shows the book keeping for the example at each stage of the Givens rotation. From the table, we see how the gains can be obtained from previously calculated gains. For example, the gain $c_3c_2c_1$ in the table is calculated by multiplying the previously calculated gain c_2c_1 by c_3 requiring only one multiplication. Similarly products like $c_4c_3c_2$ can also be calculated by multiplying $c_3c_2c_1$ with c_4 and dividing the result by c_1, though in this example we could form the product by direct multiplications as they require the same number of flops.

For a matrix having the sparse structure similar to the one shown in Fig. 1 and which uses a dyadic wavelet transform, a rough estimate of the number gain computations can be obtained as

$$\text{Number of floating point operations} = \frac{N}{2} + \frac{N}{4} \sum_{i=1}^{k-1} \frac{k-i}{2^{i-1}} - (k-1) \tag{2.25}$$

where $k = \log_2 N$. This shows that the number of gain computations is $O(N \log N)$. Note also that in those cases were the bandwidths of the sub-blocks $\hat{\mathbf{R}}_y^{j,k}$ are wider than one, the number of computations required to perform the Givens update increases while the number of gain computations needed decreases. However, the Givens update still takes only $O(N \log N)$ flops. Thus the entire

Rotation #	Updated elements	Gains	Elements	Extra Computations
1	1,5,7,8	c_1	2,3,4,6	0
2	2	c_2c_1	3,4,6	0
		c_2	5,7,8	
3	3,6	$c_3(c_2c_1)$	4	1
		c_3c_2	5,7,8	
4	4	c_4	6	1
		$c_4c_3c_2$	5,7,8	
5	5,7,8	c_5c_4	6	2
6	6	c_6	7,8	1
7	7,8			0
8	8			0

Table 1: Book keeping table indicating the number of extra computations for the 8×8 matrix example.

update procedure takes only $O(N \log N)$ flops.

2.B Discussion

The first algorithm that we presented uses a structured estimate the WT domain covariance matrix. It uses this estimate to compute an exact Cholesky decomposition. Thus the error $\hat{\mathbf{L}}_y(n)\hat{\mathbf{L}}_y^T(n) - \hat{\mathbf{R}}_y$ tends to a small value as $n \rightarrow \infty$. therefore, the error $\mathbf{E}(n) = \hat{\mathbf{L}}_y(n)\hat{\mathbf{L}}_y^T(n) - \mathbf{R}_y(n)$ tends to $\mathbf{E} = \hat{\mathbf{R}}_y - \mathbf{R}_y$ for large n. This implies that for large n, the triangular factors obtained by this algorithm converge to the exact Cholesky factors of the sparse WT domain covariance matrix. For large n, the condition number of the matrix $(\mathbf{L}_y(n)\mathbf{L}_y^T(n))^{-1}\mathbf{R}_y(n)$ can be written as

$$\begin{aligned} \chi((\mathbf{L}_y(n)\mathbf{L}_y^T(n))^{-1}\mathbf{R}_y(n)) &\approx \chi((\mathbf{L}_y\mathbf{L}_y^T)^{-1}\mathbf{R}_y) = \chi(\hat{\mathbf{R}}_y^{-1}\mathbf{R}_y) = \chi((\mathbf{R}_y + \mathbf{E})^{-1}\mathbf{R}_y) \\ &= \frac{\lambda_{\max}(\mathbf{I} + \mathbf{R}_y^{-1}\mathbf{E})}{\lambda_{\min}(\mathbf{I} + \mathbf{R}_y^{-1}\mathbf{E})} \leq \frac{1 + \lambda_{\max}(\mathbf{R}_y^{-1})\sigma_{\max}(\mathbf{E})}{1 - \lambda_{\max}(\mathbf{R}_y^{-1})\sigma_{\max}(\mathbf{E})}. \end{aligned} \tag{2.26}$$

Here we have used the inequality [32],

$$\sigma_{j+k+1}(\mathbf{AB}) \leq \sigma_{j+1}(\mathbf{A})\sigma_{k+1}(\mathbf{B}), \tag{2.27}$$

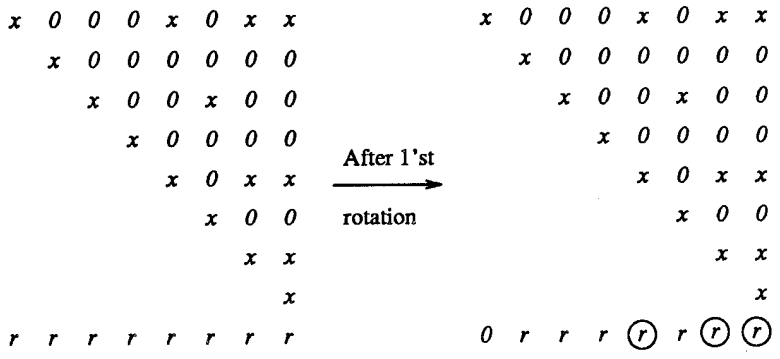

Figure 1: Zeroing of the first element in the sparse QR updation procedure.

where $\sigma_1(\mathbf{A}), \cdots \sigma_N(\mathbf{A})$ denote the singular values of $\mathbf{A}$ in non ascending order, we get

$$\lambda_{\min}(\mathbf{I} + \mathbf{R}_y^{-1}\mathbf{E}) \geq 1 - \lambda_{\max}(\mathbf{R}_y^{-1})\sigma_{\max}(\mathbf{E}). \tag{2.28}$$

However, the elements of $\mathbf{E}$ are negligible. Thus, $\lambda_{\max}(\mathbf{R}_y^{-1})\sigma_{\max}(\mathbf{E}) \ll 1$. Thus the condition number $\chi(\mathbf{L}_y\mathbf{L}_y^T)^{-1}\mathbf{R}_y)$ is close to unity.

The second algorithm uses the QR decomposition to force $\hat{\mathbf{U}}_y^T(n)\hat{\mathbf{U}}_y(n)$ to be close to the thresholded WT domain covariance matrix $\hat{\mathbf{R}}_y(n)$. In essence, this approach computes the sparse upper triangular factors of a positive definite matrix close (in some norm) to the sparse covariance matrix $\hat{\mathbf{R}}_y(n)$. Note that even though the error $\mathbf{E}(n) = \hat{\mathbf{R}}_y - \hat{\mathbf{U}}_y^T(n)\hat{\mathbf{U}}_y(n)$ is not amenable to a simple direct analysis, our simulation results indicate that condition number of the transformed and preconditioned input covariance matrix, $\chi(\hat{\mathbf{U}}_y^{-T}(n)\mathbf{R}_y(n)\hat{\mathbf{U}}_y^{-1}(n))$ converges rapidly to close to one. Thus, the sparse QR update procedure in conjunction with the wavelet domain LMS algorithm yields a Newton-LMS type algorithm with fast convergence rate and low ($O(N \log N)$) complexity.

In summary, both procedures are effective in whitening the input data. The second procedure has a lower computational complexity at the expense of a complicated bookkeeping procedure.

The superior performance of the two algorithms is clearly illustrated by Figs. 2 and 3. Fig. 2 shows the learning curves of the proposed algorithm, the DCT domain and the DWT domain modified LMS algorithms. Both the DCT and the DWT domain modified LMS algorithms use diagonal preconditioning. The learning curves are for a system identification example. The unknown system was modeled as a 32 tap FIR filter. The adaptive filters were chosen to be of the same length as the unknown modeling FIR filter. The input process to the adaptive filter was a single pole AR process. The pole location was kept at -0.98. The signal to noise ratio at the output of the unknown system was chosen to be 80dB. The learning curves were obtained as an average of 25 statistically independent experiments. The figure shows the superior performance of the proposed algorithms. As expected, the DCT domain modified LMS algorithm does not perform well when the input is a high pass process. The figure also confirms the fact that even the wavelet transform diagonalizes the correlation matrices of only certain classes of processes in the transform domain. Hence, with simple diagonal preconditioning we might not be able to obtain the optimum convergence rates. Note that μ for all algorithms was kept at 0.005. A larger μ made the DCT and DWT domain modified LMS algorithms unstable. However, a larger μ improved the convergence speed of the proposed algorithm. This is due to the fact that the condition number of the correlation matrix is large even with simple diagonal preconditioning in both the transform domains. As μ depends inversely on the largest eigenvalue of the correlation matrix, a large μ makes the DCT and the DWT domain modified LMS algorithms unstable. However, the condition number of the input correlation matrix is made close to unity by the proposed algorithm and so it does not suffer from these effects.

In order to show the robustness of the proposed algorithms to the type of input process, a system identification example was simulated. The input was again a single pole AR process with the pole located at 0.9. All other parameters in this experiment were identical to those of the previous experiment. Fig. 3 shows the learning curves of the proposed algorithm and the DCT domain modified LMS algorithm. The figure shows that the performance of the DCT domain modified LMS algorithm is now comparable to that of the proposed algorithm. Comparing Fig. 2 with Fig. 3 shows that the proposed algorithms are not sensitive to the type of input process and achieve optimal convergence speeds for all inputs. Simulations also show that the final misadjustment error for both the algorithms are nearly equal.

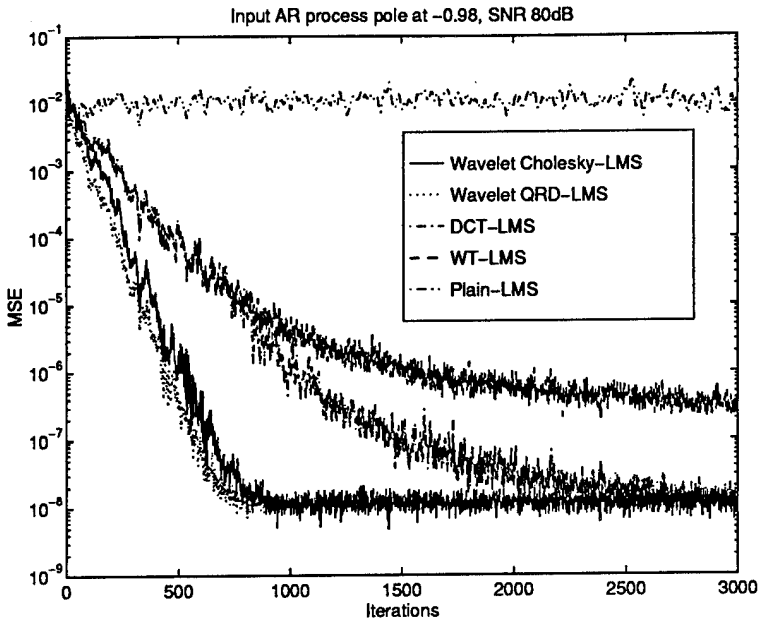

Figure 2: Plot showing the learning curves of the proposed algorithm and the DCT-LMS and DWT-LMS algorithms for a system identification example. The input signal is a single pole AR process with pole at -0.98.

3 Adaptive radar range-Doppler imaging

In [2] and [3], we describe a novel high resolution radar range-Doppler imaging that relies on an adaptive waveform selection strategy. In this approach, the transmitted radar waveforms are chosen sequentially and adaptively to match as closely as possible the optimal probing waveforms that correspond to the range-Doppler image that has to be acquired. The new radar imaging procedure yields higher resolution images for a given number of transmitted radar waveforms or a given imaging time than traditional imaging techniques that rely on a fixed set of waveforms. Alternatively, it produces an image with the desired range-Doppler resolution using minimal imaging time and a minimal number of probing waveforms. This new approach can be used in the new generation of flexible radars that are currently under testing by several manufacturers. We are also investigating an extension of these techniques for high resolution 2-D Doppler -range ultrasound imaging. By 2-D Doppler, we mean an estimate of the line of sight and transversal velocities of the target. See [33] for details.

Traditionally radar range-Doppler images (e.g., SAR and ISAR images) have been obtained by transmitting a single waveform or a train of waveforms obtained from a single pulse, repeatedly. The returns corresponding to the transmitted waveforms are then collected, demodulated, and filtered using a matched filter. The outputs of the matched filter are finally aligned and a Fast Fouier transform is computed along each range bin to generate the range-Doppler images [34]. The range-Doppler images thus obtained suffer from poor resolution and range migration effects. Range resolution increases with a decrease in the width of the transmitted pulse, while Doppler resolution increases with the number of pulses used to obtain the range-Doppler image.

In contrast with traditional range-Doppler imaging techniques, the procedure that we developed selects the transmitted radar waveforms dynamically and adaptively. The procedure *solves a simultaneous image classification and reconstruction problem.* In this approach, progressively

finer images are formed by processing the returns corresponding to the transmitted signals. The transmitted waveforms are selected sequentially and adaptively as more information is gathered. Specifically, *each transmitted waveform is selected to maximize the amount of improvement in the quality (measured using the 2 norm) of the current approximation to the desired range-Doppler image and provide guidance for the selection of the following waveform.* This strategy is based on the fact that the optimal waveforms that should be used to image a range-Doppler density function are singular functions of two kernels derived from the range- Doppler density function [2]. An added advantage of the proposed approach is that it does not suffer from range migration effects because it is based on a broadband formulation of the radar range-Doppler imaging problem.

Consider first a point reflector at a distance $r(t_0)$ from a monostatic radar at time t_0. Let s(t) be the waveform transmitted by the radar. We will assume here that the echo received by the radar at time t is given by

$$As(t - \tau(t)) \tag{3.29}$$

where A is a constant determined by the reflection properties of the point target and the propagation characteristics of the medium and $\tau(t)$ is the total delay incurred by the part of the waveform that arrives at the radar at time t. Note that

$$\tau(t) = \frac{2}{c} r(t - \frac{1}{2}\tau(t)) \tag{3.30}$$

where c is the velocity of propagation of the electromagnetic waves in the medium.

It is convenient to express $\tau(t)$ as a power series over the signal duration. In particular, if the change in the velocity of the reflector over the illumination time (duration of $s(t)$) is negligible compared to c, then we find that [8]

$$\tau(t) \approx x + \frac{2v(x/2)}{c + v(x/2)}(t - x) \tag{3.31}$$

where x is an arbitrary reference time instant and $v(x/2)$ is the velocity of the point reflector at time $x/2$ along the line of sight. Note that it follows from (3.30) that the range of the target at time $t = x/2$ is $cx/2$. Hence, the received waveform at time t will have the form $Af(y(t - x)$ where $y = (c - v(x/2))/(c + v(x/2))$. This is the exact broadband form of the echo rather than the usual narrow band approximation used in the literature. This formulation of the echo takes automatically care of range-migration effects.

Suppose now that the target consists of a continuum of point reflectors located at various ranges, moving with different velocities with respect to the radar and having different reflectivities. Assuming that reflection and propagation are linear and that all parts of the target are illuminated equally, the echo $e(t)$ due to the target will have the form

$$e(t) = \int_0^\infty dy \int_{-\infty}^\infty dx\, D(x, y) s(y(t - x)). \tag{3.32}$$

In the above equation $D(x, y)$ is the reflectivity of a point target located at range $cx/2$ and moving with velocity $(1 - y)c/(1 + y)$ at time $x/2$. Since a negative range is meaningless, $D(x, y) = 0$ for all $x < 0$. Our goal in the next section will be to reconstruct the best approximation to $D(x, y)$ by recording the echoes due to a fixed number N of transmitted waveforms.

We begin by observing that (3.32) has the same form as an inverse wavelet transform. Specifically, if $\psi(t)$ is a valid wavelet function then any square integrable signal $f(t)$ can be written as

$$f(t) = \frac{1}{C_\psi} \int_0^\infty dy \int_{-\infty}^\infty dx\, F(x, y) \sqrt{y} \psi(y(t - x)) \tag{3.33}$$

where C_ψ is a constant that depends on $\psi(t)$ and $F(x,y)$ is the continuous wavelet transform of $f(t)$ with respect to $\psi(t)$. This observation seems to suggest then that one can reconstruct $D(x,y)$ by transmitting a single valid wavelet function $\psi(t)$ and computing the wavelet transform of the echo. Unfortunately, as observed by Naparst [35] and Maass [36], following such an approach simply yields the projection of $D(x,y)$ onto the range of the wavelet transform with respect to $\psi(t)$.

Clearly the problem that we are addressing involves reconstructing a 2-D function from 1-D observations. Thus, in general, reconstructing an arbitrary $D(x,y)$ will require transmitting and recording an infinite number of waveforms and echoes. Suppose on the other hand that we are restricted to using N or less waveforms. How can we optimize the choice of these waveforms to get the most accurate approximation of $D(x,y)$ in the norm $||D(x,y)||^2 = \int_0^\infty dy \int_{-\infty}^\infty dx|D(x,y)|^2$? In [2] we showed that if $D(x,y)$ is actually known, then the optimal strategy is to transmit the N singular functions corresponding to the N largest singular values of two kernels derived from $D(x,y)$. These singular functions happen to be valid continuous wavelet functions.

The results of [2] provide a general guideline for choosing N waveforms for optimal reconstruction of a range-Doppler distribution $D(x,y)$. Specifically, the waveforms should be close approximations to the singular functions corresponding to the N largest singular values of two kernels derived from $D(x,y)$. We can use one of two strategies to solve this problem: we can either select the N waveforms to good approximations to the singular functions of the widest possible class of target kernels, or, we can select these waveforms adaptively as we acquire information about $D(x,y)$. The former approach is discussed in [2]. Here, we provide a brief description of a simple instance of the latter approach. Details may be found in [3].

3.A Discretization of the problem

We begin by briefly describing the discretization of the problem that is needed to implement the off-line and on-line stages of the imaging algorithm. The implementation of the optimal waveform selection strategy is based on a discretization of the Fourier transform of (3.32). Specifically, note that the Fourier transform of (3.32) yields:

$$E(\omega) = \int_0^\infty dy\, \Delta(\omega, y) \frac{S(\frac{\omega}{y})}{y} \tag{3.34}$$

where

$$\Delta(\omega, y) = \int_{-\infty}^\infty dx D(x,y) e^{-j\omega x}. \tag{3.35}$$

Reconstructing $\Delta(\omega,\omega/u)$ is essentially the same as reconstructing $D(x,y)$ since they are related by a one-to-one transformation. It can be shown that the operator corresponding to $\Delta(\omega,\omega/u)$ is compact and therefore admits a singular value decomposition [2]. Hence,$\Delta(\omega,\omega/u)$ can be approximated using its N singular functions corresponding to the largest singular values.

Assuming that the support of the scale parameter y is the interval $[y_{min}, y_{max}]$, (3.34) becomes

$$E(\omega) = \int_{y_{min}}^{y_{max}} dy\, \Delta(\omega, y) \frac{S(\frac{\omega}{y})}{y}. \tag{3.36}$$

Making the change of variable $u = \omega/y$, we get

$$E(\omega) = \int_{\frac{\omega}{y_{max}}}^{\frac{\omega}{y_{min}}} du\, \Delta(\omega, \omega/u) \frac{S(u)}{u}. \tag{3.37}$$

In practice $D(x,y)$ is also of finite support along the variable x. Hence $\Delta(\omega,y)$ has an infinite support. However, we shall limit ourselves to reconstructing $\Delta(\omega,y)$ within a frequency interval $[-\omega_{max},\omega_{max}]$. The choice of ω_{max} controls the ultimate achievable spatial resolution. We further normalize the ω frequency axis such that ω_{max} corresponds to π.

It remains for us to determine how we should discretize $\Delta(\omega,\omega/u)$. Two discretization schemes are possible. In the first scheme, we elect to discretize $S(\omega)$ on a uniform grid along the frequency variable ω. Unfortunately, this leads to the reconstruction of $D(\omega,y)$ on a very sparse grid in the $\omega - y$ plane. This leads to a very poor construction of $\Delta(\omega,y)$. Alternatively, we may elect to discretize $\Delta(\omega,y)$ on a uniform rectangular grid in the $\omega - y$ plane. This leads to a structured sparse matrix and a non-uniform sampling of $S(\omega)$ along the frequency variable ω. The exact structure of the discrete matrix A that corresponds to $\Delta(\omega,y)$ depends on the range of y that we need to consider. To find the value of the Fourier transform of the optimal waveform, we use an approach similar to a singular value decomposition. Specifically, we determine the N vectors $\mathbf{S_n}$, $n = 1,N$, that yield the matrix $A_N = \sum_{n=1}^{N}(A\mathbf{S_n})\mathbf{S_n^H}$ that is closest in a windowed Frobenius norm to A. The window disregards entries of A that are know to be zero a priori because of the discretization scheme that we have chosen. Here, $A\mathbf{S_n}$ represents a sampled version of the Fourier transform of the return and $\mathbf{S_n^H}$ denotes the complex conjugate transpose of $\mathbf{S_n}$.

3.B Waveform selection in the two target class case

The waveform selection strategy is based on an important observation: If we seek to minimize the windowed Frobenius norm of the error between A and A_N, we need not necessarily use the optimal vectors $\mathbf{S_n}$. It suffices to use any set of vectors $\mathbf{u_n}$ that form an orthogonal basis for the space spanned by the vectors $\mathbf{S_n}$ (c.f. [2] for the proof of this result).

In particular, we may divide the problem of reconstructing an unknown target reflectivity function into two stages. In the first stage the target reflectivity matrices of the collection of targets is clustered into distinct classes. Clustering is achieved by observing the dimension of intersection of the target matrices' right subspaces. This step is done *offline*. The second stage is done *online* and can be divided into two steps, an identification/ reconstruction step and an enhancement step. In the first step, and since the target class is unknown, vectors outside the right subspaces of the two classes have to be sent as shown in Fig.4 for the two target classes A and B. Since these vectors are not optimal in the sense that they do not lie in the appropriate spaces corresponding to either A or B, the approximation error increases from the minimum achievable by an optimal set of vectors. In the second step, once the target is identified, one proceeds by sending an appropriate set of vectors corresponding to the identified class to obtain a higher resolution of the target reflectivity matrix.

3.B.1 Reconstruction Error

In order to quantify the approximation error resulting from using an arbitrary set of N orthonormal vectors, we assume that the target is A. Let $\hat{A}_N$ be the approximation of A using a set of N orthonormal vectors $\mathbf{t}_i$ for $i = 1,\ldots,N$. Then we can write the error as

$$\|A - \hat{A}_N\|_F^2 = J_{min} + J_{excess} \tag{3.38}$$

where

$$J_{min} = \sum_{k=N+1}^{n} \sigma_k^2(A) \tag{3.39}$$

is the minimum error energy obtained when using the optimal vectors for reconstruction. It can be shown [4] that the excess error J_{excess} satisfies the bound

$$(\sigma_N^2(A) - \sigma_{N+1}^2(A))E^{\perp} \leq J_{excess} \leq (\sigma_1^2(A) - \sigma_n^2(A))E^{\perp} \tag{3.40}$$

where $E^{\perp}$ is the total energy put in $R_A^{\perp}$. These bounds are sharp whenever a significant gap exists in the set of singular values such that $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_N(A) \gg \sigma_{N+1}(A) \geq \sigma_{N+2}(A) \geq \cdots \geq \sigma_n(A)$. A similar expression also holds when the target class is B. It follows from (3.40) that J_{excess} can be reduced by reducing $E^{\perp}$. This can be achieved by selecting identification vectors that have successively the least energy possible in $R_A^{\perp}$ and $R_B^{\perp}$. This requirement can be satisfied by using the notion of principal vectors and angles between two subspaces [37]. The angles $\theta_1, \ldots, \theta_N$ between two subspaces R_A and R_B are defined recursively for $i = 1, \ldots, N$ by

$$\cos\theta_k = \max_{\mathbf{u} \in R_A} \max_{\mathbf{v} \in R_B} \mathbf{u}^T\mathbf{v} = \mathbf{u}_k^T\mathbf{v}_k \tag{3.41}$$

subject to

$$\|\mathbf{u}\| = \|\mathbf{v}\| = 1 \tag{3.42}$$

$$\mathbf{u}^T\mathbf{u}_i = 0 \qquad i = 1, \ldots, k \tag{3.43}$$

$$\mathbf{v}^T\mathbf{v}_i = 0 \qquad i = 1, \ldots, k. \tag{3.44}$$

The principal angles satisfy $0 \leq \theta_1 \leq \cdots \leq \theta_N \leq \pi/2$. The vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ are called principal vectors between R_A and R_B and both form an orthonormal basis for these subspaces. respectively. From this definition, one can show that the unit vector having the least component in $R_A^{\perp}$ and $R_B^{\perp}$ is given by

$$\mathbf{t}_1 = \alpha_1(\mathbf{u}_1 + \mathbf{v}_1) \tag{3.45}$$

where $\mathbf{u}_1$ and $\mathbf{v}_1$ are the first principal vectors of R_A and R_B , respectively and α_1 a scalar chosen such that $\|\mathbf{t}_1\| = 1$ and $\mathbf{t}_1$ bisecting the first principal angle θ_1. One can show [37] that the vector having the least component in $L_A \cap span\{\mathbf{t}_1, \ldots, \mathbf{t}_{i-1}\}^{\perp}$ and $L_B \cap span\{\mathbf{t}_1, \ldots, \mathbf{t}_{i-1}\}^{\perp}$ is given for $i = 2, \ldots, p$ by

$$\mathbf{t}_i = \alpha_i(\mathbf{u}_i + \mathbf{v}_i) \tag{3.46}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith principal vectors of L_A and L_B , respectively and α_i a scalar chosen such that $\|\mathbf{t}_i\| = 1$ and $\mathbf{t}_i$ bisecting the ith principal angle θ_i. We will show in the next section that we may actually have to use vectors of the form $\mathbf{t}_i = \alpha_i\mathbf{u}_i + \beta_i\mathbf{v}_i$ for some scalars α_i and β_i to minimize the number of transmissions needed for identification. The number of transmissions required for identification is given by p.

Once the target class is identified, we can obtain a higher resolution target reflectivity matrix by sending the principal vectors $(\mathbf{u}_{p+1}, \ldots, \mathbf{u}_N)$ or $(\mathbf{v}_{p+1}, \ldots, \mathbf{v}_N)$ depending on the target.

After K transmissions, the target density approximation is given by

$$\hat{T}_K = \sum_{i=1}^{K} (T\mathbf{t}_i)\mathbf{t}_i^T \tag{3.47}$$

where T denotes the target and $\hat{T}_K$ its approximation after K transmissions.

Equation (3.40) has another important implication. When $E^{\perp}$ is small enough for both target classes, the components in $R_A^{\perp}$ and $R_B^{\perp}$ are small so that for practical purposes their right subspaces

can be considered to coincide. In terms of principal angles, if $\theta_i, i \leq N$, is small the total energy in $R_A^{\perp}$ and $R_B^{\perp}$ due to the identification vectors $\mathbf{t}_1, \ldots \mathbf{t}_i$ will be small. If i is large enough, the excess error energy is reduced and the approximation error energy approaches the value given in (3.39). The above suggests the following clustering criterion. We say that two target densities belong to the same class if for some $i \leq N$, the ith principal angle between their right subspaces falls below some prescribed threshold θ_{th}. Obviously, the larger i, the "closer" will be the densities belonging to the same cluster. In an extreme case, when $i = N$, the right subspaces of both classes would coincide.

As a final note here, we would like to point out that when clustering matrices, each class contains in general more than one matrix. Although the matrices in a particular class are said to be "close", a response due to different matrices of the same class might lie in disjoint left subspaces as shown in Fig.6.

3.B.2 Waveform Selection

Classes are identified by the subspaces of the optimal probing vectors. These are the the subspaces spanned by the right singular vectors corresponding to the N largest singular values. The singular values play no role in this categorization. To see why this is so, consider an extreme case where the right and left subspaces of two target matrices coincide but have different singular values. In the clustering step, they would belong to the same class despite the difference in their singular values.

For minimum excess error, we would like to devise a test to determine the target by observing the return from a single transmission. For simplicity, we will consider first the case when the left subspaces of A and B are not intersecting. Denoting the left space observation by $\mathbf{r}$ and the transmitted vector by $\mathbf{x}$, we can set up our identification problem as a composite hypotheses testing problem that is given by

$$\begin{aligned} H_o \ \text{(A target)} \ &: \quad \mathbf{r} = A_N\mathbf{x} + \mathbf{w} \\ H_1 \ \text{(B target)} \ &: \quad \mathbf{r} = B_N\mathbf{x} + \mathbf{w} \end{aligned}$$

where A_N and B_N are the optimal rank-N approximations of A and B, respectively, and $\mathbf{w}$ is the uncertainty due to the component in the subspaces $R_A^{\perp}$ or $R_B^{\perp}$. This uncertainty is present because of the absence of information about the singular values of A and B. Since no stochastic information is available about the uncertainty, we form the following generalized likelihood ratio

$$\Lambda(\mathbf{r}) = \frac{\max\limits_{\mathbf{x}} \|\mathbf{r} - B_N\mathbf{x}\|^2}{\max\limits_{\mathbf{x}} \|\mathbf{r} - A_N\mathbf{x}\|^2}. \tag{3.48}$$

The maximization leads to a least-square estimate of $\mathbf{x}$ and the likelihood ratio becomes

$$\begin{aligned} \Lambda(\mathbf{r}) &= \frac{\|(I - B_N(B_N^TB_N)^{-1}B_N^T)\mathbf{r}\|^2}{\|(I - A_N(A_N^TA_N)^{-1}A_N^T)\mathbf{r}\|^2} \\ &= \frac{\|(I - P_B)\mathbf{r}\|^2}{\|(I - P_A)\mathbf{r}\|^2} \end{aligned} \tag{3.49}$$

where P_A and P_B are projection matrices on L_A and L_B, respectively. Hence, our rule becomes measuring the energy in $L_A^{\perp}$ and $L_B^{\perp}$ and deciding on the target that has a smaller energy, i.e., that is closer in the 2-norm to its respective left subspace assuming equally likely targets.

We now examine the conditions under which a test based on (3.49) will provide correct decisions. We rewrite (3.46) as

$$\mathbf{t}_i = \alpha \mathbf{u}_i + \sqrt{1-\alpha^2}\,\mathbf{u}_i^{\perp} \qquad i = 1, \ldots, p \tag{3.50}$$

where $\mathbf{u}_i^{\perp}$ is a vector perpendicular to $\mathbf{u}_i$ and lies in the subspace spanned by $\mathbf{u}_i$ and $\mathbf{v}_i$ as shown in Fig.5. In equation (3.50) we allowed ourselves the flexibility to control the components of $\mathbf{t}_i$ in R_A and in $R_A^{\perp}$ using the scalar α. Unlike (3.46), $\mathbf{t}_i$ need not bisect the principal angle θ_i.

Let the target be A and let $E_A = \alpha^2$. It can be shown that for a correct decision we have the following condition [4]

$$1 \geq E_A \geq \frac{1}{1+\rho(A)} = \eta_A \tag{3.51}$$

where

$$\rho(A) = \frac{\sin^2\theta_{min}\sigma_N^2(A)}{\sin^2\theta_{max}\sigma_{N+1}^2(A)}. \tag{3.52}$$

In (3.52) θ_{min} and θ_{max} are the minimum and maximum principle angles between the left subspaces of the target matrices A and B. Similarly, one can show that if the target is B then

$$1 \geq E_B \geq \frac{1}{1+\rho(B)} = \eta_B \qquad \text{and} \tag{3.53}$$

$$\rho(B) = \frac{\sin^2\theta_{min}\sigma_N^2(B)}{\sin^2\theta_{max}\sigma_{N+1}^2(B)} \tag{3.54}$$

where E_B is the energy of $\mathbf{t}_i$ in R_B.

The inequalities in (3.51) and (3.53) show that $\mathbf{t}_i$ cannot be placed arbitrarily between the principal vectors and that erroneous decisions might result from a choice such as the one given in (3.46). Therefore, it might be necessary to sacrifice the reconstruction performance of one transmission to identify the target. It is desirable to have η_A and η_B take as small values as possible so that both (3.51) and (3.53) are simultaneously satisfied. This situation is depicted in Fig. 7. It shows how $\mathbf{t}_i$ can vary without violating the inequalities. It shows also that in this case the choice given in (3.46) is acceptable. Small values for η_A and η_B are obtained whenever $\sigma_N(A)/\sigma_{N+1}(A) \gg |sin\theta_{max}/sin\theta_{min}|$ and $\sigma_N(B)/\sigma_{N+1}(B) \gg |sin\theta_{max}/sin\theta_{min}|$. For this to occur it is desirable to have a significant gap in the distribution of the singular values of both target matrices. If the left subspaces intersect, or if θ_{min} is too small so that (3.51) and (3.53) cannot be simultaneously satisfied, we have to modify our strategy. Specifically, rather than using (3.50), we let $\mathbf{t}_i = \mathbf{u}_i$. The return will either lie in L_A if the target is A, or will have components in L_B and $L_B^{\perp}$ if the target is B. If the right subspaces are close, the component in $L_B^{\perp}$ could be so small that one could consider the target to be B. In this case, at least another transmission would be necessary to determine the target. This scenario could occur whenever there exists an intersection between the left subspaces of the two classes. For practical purposes, the dimension of this intersection is measured by the number of principal angles that falls below some prescribed threshold. If the dimension of this intersection is p, then we need at most p transmissions to identify the target. This is the same p as in (3.46).

3.C Waveform selection in the general $M \geq 2$ target class case

The extension of the above procedure to the case where the target may belong to one of M classes (rather than just two) is conceptually straightforward. However, the implementation is more complex. The added complexity is due to the fact that unlike the two class case, the procedure typically takes longer to identify the class to which the target belongs. Furthermore, selecting the waveform to be transmitted at any point requires that we solve a max-min problem. This max-min problem must be solved in real-time during actual imaging unless enough memory is available for storing all the possible waveform selections that can be made during actual imaging. We have developed a procedure that solves the resulting max-min problem efficiently by exploiting the geometry of the libraries of waveforms that correspond to the M classes (c.f. [3] for details).

Fig. (8) shows the reduction in the Frobenius norm of the error between an estimated target reflectivity function and the actual target reflectivity as a function of the number of waveforms a typical experiment. The three curves compare the case where we know the class to which the target belongs (lowest curve), where the target is known to belong to one of two classes (middle curve) and the case where the target is known to belong to one of three classes (upper curve). Note that as the uncertainty in class membership increase, we pay a price for our ignorance in a reduced reconstruction quality.

4 Publications in Technical Journals

1. "Parametrization of Compactly Supported Orthonormal Wavelets," H. Zou and A. H. Tewfik, *IEEE Trans. on Signal Proc.*, vol. 41, no. 3, pp. 1428-1431, March 1993.

2. "An SVD Approach to Edge Detection," A. H. Tewfik and M. Deriche, *IEEE Trans. on Image Proc.*, vol. 2, no. 3, pp. 353-368, July 1993.

3. " Signal Modeling with Filtered Discrete Fractional Noise Processes," M. Deriche and A. H. Tewfik, *IEEE Trans. on Signal Proc.*, vol. 41, no. 9, pp. 2839-2849, Sept. 1993.

4. " Maximum Likelihood Estimation of the Parameters of Discrete Fractional Gaussian Noise Processes," M. Deriche and A. H. Tewfik, *IEEE Trans. on Signal Proc.*, vol. 41, no. 10, pp. 2977-2989, Oct. 1993.

5. "Low Bit Rate Transparent Audio Compression Using Adapted Wavelets," D. Sinha and A. H. Tewfik, *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3463-3479, Dec. 1993.

6. "Fast Multiscale Statistical Signal Processing Algorithms," A. H. Tewfik and M.-Y. Kim, *IEEE Trans. on Signal Proc.*, vol. 42, no. 3, pp. 572- 585, March 1994.

7. "Recent Progress in the Application of Wavelets in Surveillance Systems," A. H. Tewfik, S. Hosur and S. Sowelam, invited paper, *Optical Engineering,* vol. 33, no. 8, pp. 2509-2519, August 1994.

8. "Time Delay Estimation Using Wavelet Transform for PW Ultrasound," X.-L. Xu, A. H. Tewfik and J. F. Greenleaf, invited paper, to appear in *Annals of Biomedical Engineering,* 1995.

9. "Multiscale Difference Equation Signal Models: Part I Theory" M. Ali and A. H. Tewfik, to appear in *IEEE Trans. on Signal Proc.*, Oct. 1995.

10. "Completeness and Stability of Partial Wavelet Domain Signal Representations" A. H. Tewfik and H. Zou, to appear in *IEEE Trans. on Signal Proc.*, Dec. 1995.

11. "A Wavelet Transform Domain Adaptive Algorithm," S. Hosur and A. H. Tewfik, submitted to *IEEE Trans. on Signal Proc.*, April 1994, revised May 1995.

12. "Binary Wavelet Decomposition of Binary Images," M. Swanson and A. H. Tewfik, in preparation, submitted to *IEEE Trans. on Image Proc.*, Jan. 1995.

13. "Generalized ULV Adaptive Algorithm," S. Hosur, A. H. Tewfik and D. Boley, submitted to *IEEE Trans. on Signal Proc.*, June 1995.

14. " Waveform Selection in High Resolution Radar Range-Doppler Imaging: Part I theory" A. H. Tewfik and S. Sowelam , in preparation, to be submitted to *IEEE Trans. on Image Proc.*, June 1995.

15. "Waveform Selection in High Resolution Radar Range-Doppler Imaging: Part II adaptive algorithm" S. Sowelam and A. H. Tewfik, in preparation, to be submitted to *IEEE Trans. on Image Proc.*, Sept. 1995.

5 Professional Personnel

- Principal Investigator
  1. Prof. A. H. Tewfik
- Graduate Ph.D. theses completed with partial funding from this grants
  1. Deepen Sinha: Thesis on "Adaptive High Quality Wavelet Encoding of Audio Signals," Feb. 1993.
  2. Hehong Zou: Thesis on "Generalized Wavelet Representations," April 1993.
- Graduate students
  1. Srinath Hosur
  2. Bin Zhu
  3. Murtaza Ali

6 Interactions

- Plenary Talks, Keynote Addresses and invited tutorials:

  1. "Potentials and Limitations of Wavelets in Signal Acquisition and Processing," A. H. Tewfik, *15th Annual Int. Conf. of the IEEE EMBS*, Oct. 1993.
  2. "Digital Fractals," A. H. Tewfik, *1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
  3. "Wavelets: Theory and Applications," *1994 IEEE Time-Frequency and Time-Scale Symp.*, Philadelphia, PA, Oct. 1994.

- Invited Papers presented at meetings, conferences and seminars:

  1. "Estimation of Range-Doppler Radar Images," A. H. Tewfik, invited paper, in *Proc. of the 1993 IEEE Symp. on Circuits and Systems*, Chicago, IL, May 1993.
  2. "Acoustical Applications of Wavelets: Sonar and Audio Coding," A. H. Tewfik, invited paper, in *Proc. 125th Meeting of the Acoustical Society of America*, Ottawa, Canada, May 1993.
  3. "Enhanced Wavelet Based Audio Coder," A. H. Tewfik and M. Ali, invited paper, in *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, Monterey, CA., Nov. 1993.
  4. "Recent Developments in Wavelet Theory and Applications," A. H. Tewfik, invited paper, *NSF Sponsored International Conference on Mathematical Analysis and Signal Processing*, Cairo, Egypt, Jan. 1994.
  5. "Recent Progress in the Application of Wavelets in Surveillance Systems," A. H. Tewfik and S. Hosur, in *Proc. SPIE Conf. on Wavelet Applications*, SPIE Proc. Vol. 2242, April 1994.
  6. "M-target Class Adaptive Radar Range-Doppler Imaging in Clutter: Theory and Experimental Results," A. H. Tewfik and S. Sowelam, in *Proc. SPIE Conf. on Wavelet Applications for Dual Use*, April 1995.
  7. "Wavelets: A Passing Wave or a Truly Useful Tool?," A. H. Tewfik, in *Conf. Digital Processing Technology (Critical Review)*, April 1995.

- Papers presented at meetings, conferences and seminars:

  1. "Fast Magnetic Resonance Imaging via Frequency Domain Wavelet Transforms," A. H. Tewfik, H. Garnaoui and X. Hu , in *Proc. of the 1993 IEEE Conf. on Acoust. Speech and Signal Proc.*, Minneapolis, MN, April 1993.
  2. "Completeness and Stability of Partial Wavelet Domain Signal Representations" H. Zou, A. H. Tewfik and W. Xu , in *Proc. of the 1993 IEEE Conf. on Acoust. Speech and Signal Proc.*, Minneapolis, MN, April 1993.
  3. "Wavelet Transform Domain LMS Algorithm" S. Hosur and A. H. Tewfik, in *Proc. of the 1993 IEEE Conf. on Acoust. Speech and Signal Proc.*, Minneapolis, MN, April 1993.

4. "Low Bit Rate Transparent Audio Compression Using a Dynamic Dictionary and Optimized Wavelets" D. Sinha and A. H. Tewfik, in *Proc. of the 1993 IEEE Conf. on Acoust. Speech and Signal Proc.*, Minneapolis, MN, April 1993.
5. "Filtered Fractals in Signal Modeling" M. Deriche and A. H. Tewfik, in *Proc. of the 1993 IEEE Symp. on Circuits and Systems*, Chicago, IL, May 1993.
6. "Real-time Synthesis of CD-Quality Audio Using Adaptive Wavelet Based Coding Algorithm," M. Ali, A. H. Tewfik, D. Sinha, W. Anderson and J. Rowlands, in *Proc. 1993 DSPx Conference*, San Jose, CA, Oct. 1993.
7. "Robust Multi-Resolution Integrated Target Sensing and Recognition," A. H. Tewfik, in *Proc. 1st Joint ATR Workshop*, Lexington, MA, Nov. 1993.
8. "Perfect Reconstruction Filter Banks with Arbitrary Regularity,"A. H. Tewfik, in *Proc. SPIE Conf. on Wavelet Applications*, SPIE Proc. Vol. 2242, April, 1994.
9. "Second Generation Audio Information Coding," A. H. Tewfik, M. Ali and V. Viswanathan, in *Proc. SPIE Conf. on Wavelet Applications*, SPIE Proc. Vol. 2242, April, 1994.
10. "Generalized URV Subspace Tracking LMS Algorithm" S. Hosur, A. H. Tewfik and D. Boley, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
11. "Multiscale Difference Equation Signal Modeling and Analysis" M. Ali and A. H. Tewfik, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
12. "Wavelet Domain Bearing Estimation in Unknown Correlated Noise" A. H. Tewfik, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
13. "Robust Multiresolution Integrated Target Sensing And Recognition," *MIT Lincoln Labs.*, ARPA Joint ATR Workshop, Nov. 1993.
14. "Wavelets and Radar Signal Processing," *Naval Weapons Center*, China Lake, CA, December 1993.
15. "Wavelets: Theory and Applications," *University of Minnesota, Physics Colloquium*, Feb. 1994.
16. "Perfect Reconstruction Filter Banks with Arbitrary Regularity,"A. H. Tewfik, in *Proc. SPIE Conf. on Wavelet Applications*, SPIE Proc. Vol. 2242, April, 1994.
17. "Second Generation Audio Information Coding," A. H. Tewfik, M. Ali and V. Viswanathan, in *Proc. SPIE Conf. on Wavelet Applications*, SPIE Proc. Vol. 2242, April, 1994.
18. "Generalized URV Subspace Tracking LMS Algorithm" S. Hosur, A. H. Tewfik and D. Boley, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
19. "Multiscale Difference Equation Signal Modeling and Analysis" M. Ali and A. H. Tewfik, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
20. "Wavelet Domain Bearing Estimation in Unknown Correlated Noise" A. H. Tewfik, in *Proc. of the 1994 IEEE Conf. on Acoust. Speech and Signal Proc.*, Adelaide, Australia, April 1994.
21. "ECG Coding by Wavelet Transform Extrema," A. E. Cetin, A. H. Tewfik and Y. Yardimci, *1994 IEEE Symp. Time-Freq. and Time-Scale,* Oct. 1994.

22. "Optimal Waveform Selection in Range-Dopler Imaging," S. Sowelam and A. H. Tewfik, *1994 IEEE Int. Conf. Image Proc.*, Nov. 1994.
23. "Wavelet Decomposition of Binary Finite Images," M. Swanson and A. H. Tewfik, *1994 IEEE Int. Conf. Image Proc.*, Nov. 1994.
24. ""Waveform Selection for High Resolution Range-Dopler Imaging," A. H. Tewfik, *1995 ONR Wideband RF Science and Technology Workshop.*, Jan. 1995.
25. "Low Bit Rate Transparent Image Coding With Optimized Mixed Representations," A. H. Tewfik and B. Zhu, in *Proc. SPIE Conf. on Wavelet Applications for Dual Use*, April 1995.
26. "Image Coding with Mixed Representations and Visual Masking" B. Zhu, A. H. Tewfik and O. Gerek, in *Proc. of the 1995 IEEE Conf. on Acoust. Speech and Signal Proc.*, Detroit, MI, May 1995.
27. "Detection of Weak Signals Using Adaptive Stochastic Resonance" A. Asdi and A. H. Tewfik, *Proc. of the 1995 IEEE Conf. on Acoust. Speech and Signal Proc.*, Detroit, MI, May 1995.

- Consultative and advisory functions to USAF and DOD
  - Visit to MIT Lincoln Laboratories, Nov. 1993.
  - Visit to NAWCWPNS, China Lake, CA, Dec. 1993.
  - Visit to NraD, SanDiego, CA, Jan. 1995.

- Patents
  None filed in 1993-1995.

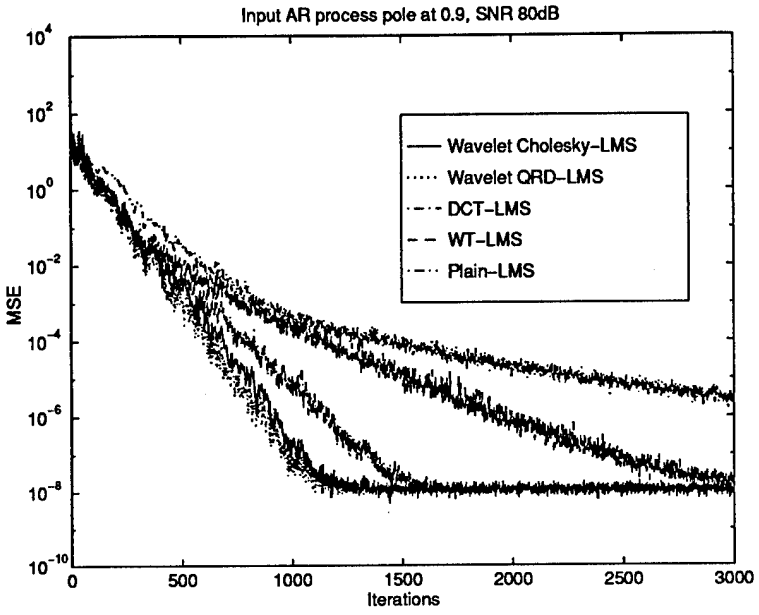

Figure 3: Plot showing the learning curves of the proposed algorithm, and the DCT-LMS for a system identification example. The input signal is a single pole AR process with pole at 0.9.

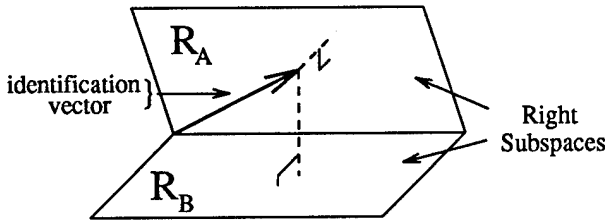

Figure 4: Identification vectors relative to the right subspaces.

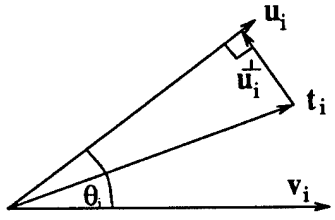

Figure 5: Transmitted vector as a combination of the ith principal vectors.

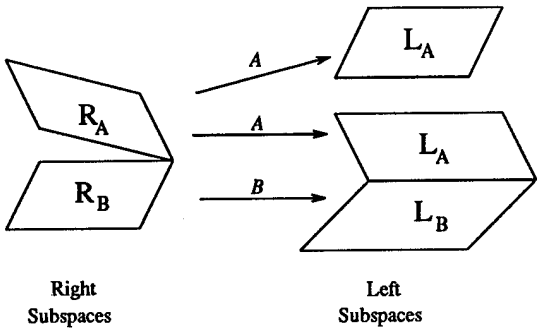

Figure 6: Left subspace corresponding to two different target classes.

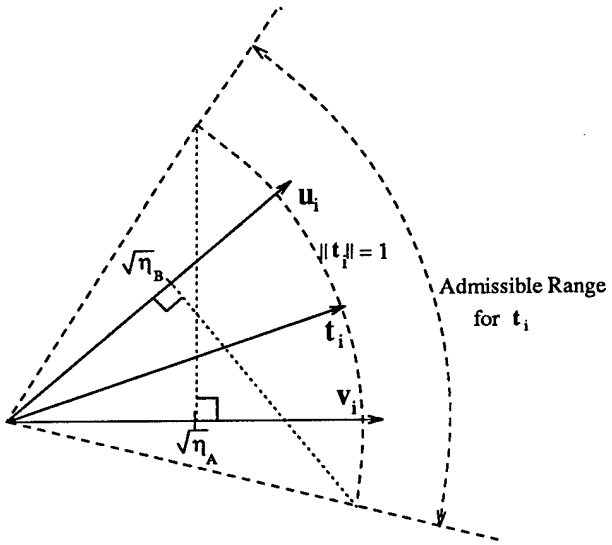

Figure 7: Range of the transmitted vector, $\mathbf{t}_i$, for correct decision.

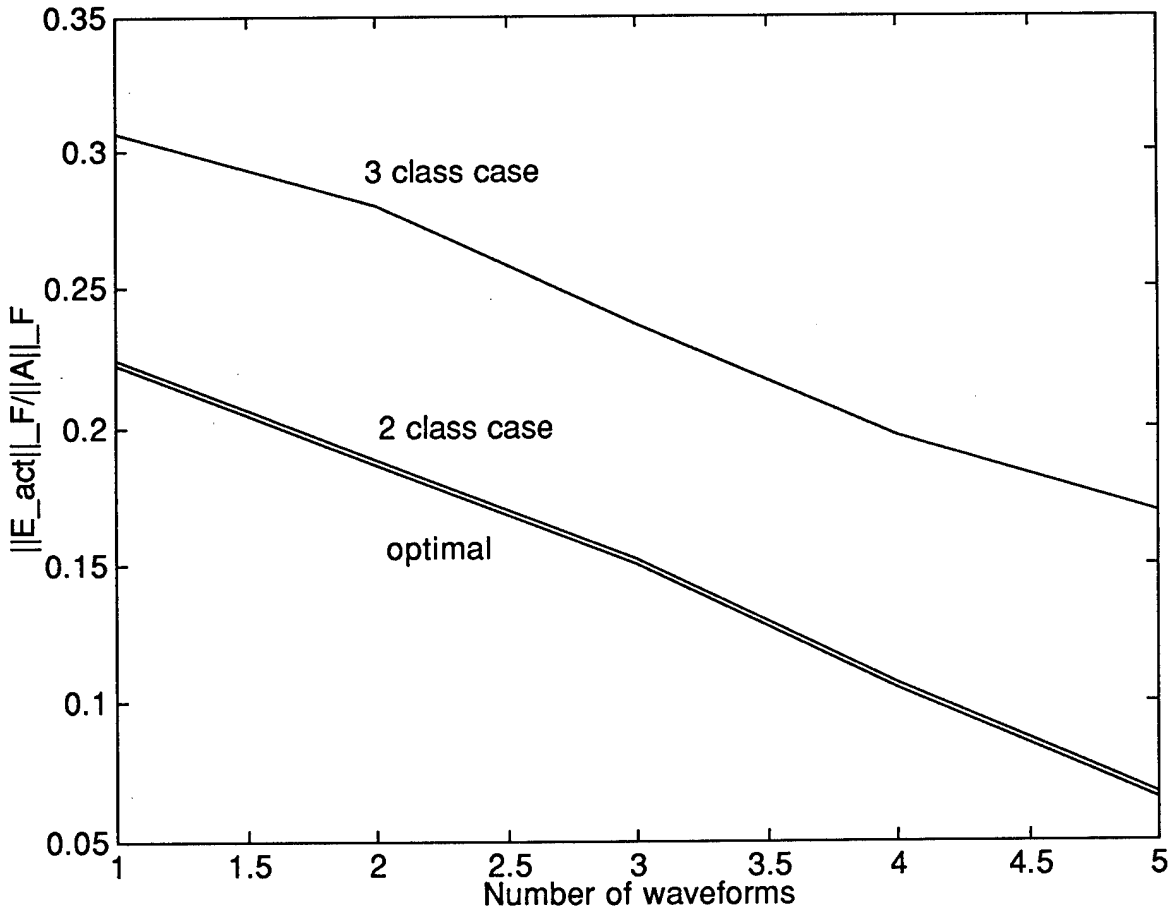

Figure 8: Improvement of the reconstructed $D(x, y)$ in a typical experiment as a function of the number of waveforms used in three scenarios: known target class, two possible target classes and three possible target classes.

References

[1] S. Hosur and A. H. Tewfik, "A Wavelet Transform Domain Adaptive Algorithm," submitted to *IEEE Trans. on Signal Proc.*, April 1994, revised May 1995.

[2] A. H. Tewfik, ' Waveform Selection in High Resolution Radar Range-Doppler Imaging: Part I theory," in preparation, to be submitted to *IEEE Trans. on Image Proc.*, June 1995.

[3] A. H. Tewfik, "Waveform Selection in High Resolution Radar Range-Doppler Imaging: Part II adaptive algorithm," in preparation, to be submitted to *IEEE Trans. on Image Proc.*, Sept. 1995.

[4] A. H. Tewfik, S. Hosur, and S. Sowelam, "Recent progress in the application of wavelets in surveillance," *Optical Engineering*, pp. 2509-2519, August 1994.

[5] J. J. Shynk, "Frequency domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–39, 1992.

[6] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Engelwood Cliffs, N.J.: Prentice Hall, 1985.

[7] S. Haykin, *Adaptive Filter Theory.* Engelwood Cliffs, N.J.: Prentice Hall, 1991.

[8] S. Hosur, A. H. Tewfik, and D. L. Boley, "Generalized URV subspace tracking LMS algorithm." Under Preperation, See also *Proc. ICASSP*, Adelaide, Australia, April 1994.

[9] B. Walzman and M. Schwartz, "Automatic equalization using the discrete frequency domain," *IEEE Trans. Info. Theory*, vol. 19, no. 1, pp. 59–68, 1973.

[10] J. C. Lee and C. K. Un, "Performance of transform domain LMS adaptive digital filters," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 34, no. 3, pp. 499–510, 1986.

[11] D. F. Marshall, W. K. Jenkins, and J. J. Murphy, "The use of orthogonal transforms for improving performance of adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, no. 4, pp. 474–483, 1989.

[12] S. Hosur and A. H. Tewfik, "Wavelet transform domain LMS algorithm," in *Proc. Int. Conf. Acosut., Speech and Signal Proc.*, vol. III, pp. 508–510, April 1993.

[13] N. Erdöl and F. Basbug, "Performance of wavelet tramsform based adaptive filters," in *Proc. Int. Conf. Acosut., Speech and Signal Proc.*, vol. III, pp. 500–503, April 1993.

[14] M. Doroslovacki and H. Fan, "Wavelet-based adaptive filtering," in *Proc. Int. Conf. Acosut., Speech and Signal Proc.*, vol. III, pp. 488–491, April 1993.

[15] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 40, no. 8, pp. 1862–1875, 1992.

[16] S. Gazor and B. Farhang-Boroujeny, "Quantization effects in transform-domain normalized LMS algortihms," *IEEE Trans. Circuits Syst. II: Analog and Digital Signal Processing*, vol. 39, no. 1, pp. 1–7, 1992.

[17] A. H. Tewfik and M. Kim, "Fast multiscale signal processing algorithms," to appear in *IEEE Trans. Signal Processing*, March 1994.

[18] R. Dijkerman, V. Badrinath, and R. R. Mazumdar, "Multiscale representation of stochastic processes using compactly supported wavelets," in *Proc. IEEE-SP Int. Symp. Time-Frequency Anal.*, pp. 185–188, Oct. 1992.

[19] R. C. G. Beylkin and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," *Commun. Pure and Appl. Math.*, vol. XLIV, pp. 141–183, 1991.

[20] M. V. Wickerhauser, "Non standard matrix multiplication." Preprint, *Dept. of Math.*, Yale University, 1990.

[21] G. W. Wornell and A. V. Oppenheim, "Wavelet based representations for a class of self-similar signals with applications to fractal modulation," *IEEE Trans. Info. Theory*, vol. 38, no. 2, pp. 785–800, 1992.

[22] B. B. Mandelbrot and J. W. V. Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.

[23] R. J. Barton and H. V. Poor, "Signal detection in fractional gaussian noise," *IEEE Trans. Info. Theory*, vol. 34, no. 5, pp. 943–959, 1988.

[24] J. R. M. Hosking, "Fractional differencing," *Biometrica*, vol. 68, no. 1, pp. 165–176, 1981.

[25] M. Deriche and A. H. Tewfik, "Signal modeling with filtered discrete fractional noise processes," *IEEE Trans. Signal Processing*, vol. 41, no. 9, pp. 2839–49, 1993.

[26] W. S. Burdic, *Underwater Acoustic System Analysis.* Prentice Hall, 1984.

[27] G. H. Golub and C. F. Van Loan, *Matrix Computations.* The John Hopkins University Press, 1983.

[28] Y. Bard, *Nonlinear Parameter Estimation.* New York.: Academic Press, 1974.

[29] R. T. Compton Jr., *Adaptive Antennas : Concepts and Performance.* Prentice Hall, 1988.

[30] S. M. Goldfeld et. al., "Maximization by quadratic hill climbing," *Econometrica*, no. 34, pp. 541–551, 1966.

[31] P. E. Gill et. al., "Newton-type methods for unconstrained and linearly constrained optimization," *Math. Prog.*, pp. 311–350, 1974.

[32] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge, UK: Cambridge University Press, 1991.

[33] Y. Kadah and A. H. Tewfik, "True-Velocity Duplex Imaging Using A Single Transducer ," to appear in *Proc. IEEE Int. Conf. on Image Proc.*, Oct. 1995.

[34] D. R. Wehner, *High Resolution Radar.* Norwood, MA: Artech House, Inc., 1987.

[35] H. Naparst, "Dense target signal processing," *IEEE Trans. Info. Theory*, vol. IT-37, no. 2, pp. 317–327, 1991.

[36] P. Maass, *Wideband Approximation and Wavelet Transform.* New York: Springer-Verlag, 1992.

[37] D. S. Watkins, *Fundamentals of Matrix Computations*, John Wiley & Sons, Inc., New York, 1991.