Nonlinear processing of a shift invariant DWT for noise reduction

M. Lang*, H. Guo†, J. E. Odegard‡, and C. S. Burrus §
Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251-1892

R. O. Wells¶
Department of Mathematics, Rice University, Houston, TX 77251-1892

Paper accepted for the Proceedings of SPIE, *Mathematical Imaging: Wavelet Applications for Dual Use*, in SPIE Symposium on OE/Aerospace Sensing and Dual Use Photonics, 17-21 April 1995, Orlando, FL.

ABSTRACT

A novel approach for noise reduction is presented. Similar to Donoho, we employ thresholding in some wavelet transform domain but use a nondecimated and consequently redundant wavelet transform instead of the usual orthogonal one. Another difference is the shift invariance as opposed to the traditional orthogonal wavelet transform. We show that this new approach can be interpreted as a repeated application of Donoho's original method. The main feature is, however, a dramatically improved noise reduction compared to Donoho's approach, both in terms of the l_2 error and visually, for a large class of signals. This is shown by theoretical and experimental results, including synthetic aperture radar (SAR) images.

Keywords: noise reduction, shift invariance, redundancy, wavelet transform, SAR

1 INTRODUCTION

Recently a novel approach for noise reduction due to Donoho and Johnstone[12,13] has been established. It employs thresholding in the wavelet domain and can be shown to be asymptotically near optimal for a wide class of signals corrupted by additive white Gaussian noise (AWGN). Moreover, the same method can be used in a wide variety of related problems such as linear inverse problems,[10] data compression and statistical estimation.[11]

Donoho's method for noise reduction has been proved to work well for 1D medical[10] and geophysical signals[30] as well as for 2D geophysical[30] and synthetic aperture radar (SAR) signals[18,24] where the AWGN assumption is a viable approximation to the real noise properties.[1] Surprisingly, the method can be successfully applied even in cases where the error is not AWGN. For example it is possible to remove blocking artifacts in images of JPEG decoded signals by applying wavelet thresholding, where the error is neither white nor Gaussian.[16] The

*This research was partially supported by the Alexander von Humboldt foundation and by TI
†This research was partially supported by TI
‡This research was supported by AFOSR under grant F49620-1-0006 funded by ARPA
§This research was supported by BNR
¶This research was supported by AFOSR under grant F49620-1-0006 funded by ARPA

interesting feature in all these applications is that the resulting signals are essentially noise free with little loss of image detail. We will present SAR image examples below.

In the following summary of some of the recent literature, papers on compression, coding, and detection are also included, since all these schemes can be interpreted as thresholding in the wavelet domain. The advantage, but also the problem, of the wavelet thresholding scheme is the degree of freedom offered by the large number of parameters that has to be specified. They are: the particular transform (e.g., wavelet, 2 band/M band, best basis, modulated cosine, biorthogonal, time or space varying), the analysis and synthesis filters, the particular thresholding scheme, the determination of the threshold, and the number of scales used. Most of the recent papers focus on the choice of a threshold parameter and/or the determination of the wavelet transform.[12,30,26,34] The two predominant thresholding schemes are soft thresholding ("shrink or kill") and hard thresholding ("keep or kill"). Donoho has shown that soft thresholding is the l_2 optimal nonlinear function in the wavelet domain to apply if one requires the resulting function to be at least as smooth as the original, noise free one. Furthermore the resulting error is within a logarithmic factor of the so-called ideal risk — a performance measure of some ideal scheme. This concept will be discussed later in detail. Hard thresholding, on the other hand, yields better l_2 performance but does not guarantee the smoothness property cited above. This means in practice that the resulting signals might exhibit spurious oscillations. It can be shown that hard thresholding is also within a logarithmic factor of the ideal risk.[13]

There exists a variety of proposals for estimating the threshold. Donoho gives some minimax thresholds for several threshold schemes as well as a "universal threshold". These explicitly depend on the standard deviation σ of the noise* and the number of data points. Experimental studies in[16,18] show that for some applications the optimal threshold can be easily computed as $c\sigma$ where c is a constant. An alternative approach for selecting the threshold, independently proposed by Nason and Weyrich[26,34] uses cross-validation. The goal is to minimize the least squares error between the original (unknown) function and its estimate based on the noisy observation. Although this approach yields smaller l_2 error for some cases, it is worse than some minimax thresholds in other cases.[26] A third approach presented by Saito,[30] employs an information theoretic error measure, the minimum description length (MDL). The method implicitly computes a (hard) threshold by determining the number of largest (in modulus) wavelet coefficients to be kept.

The set of orthogonal bases usually contains a two band wavelet analysis with Daubechies wavelets or Coiflets, the wavelet packets, local trigonometric bases. There are two common methods for adapting the wavelet transform to the signal at hand. One computes all transforms of the signal and chooses that which yields a minimum error according to some performance measure, for example the MDL[30] or the entropy.[5,6] The other approach is the so-called matching pursuit algorithm[23,8] for signal representation by Mallat. In contrast to the algorithm mentioned first, the matching pursuit algorithm yields a nonorthogonal expansion. The idea is to select the largest (in modulus) expansion coefficient out of a redundant transform, e.g. a set of several orthogonal transforms, subtract the corresponding component from the signal, compute the coefficients of all transforms of the residual, select the largest coefficient and so forth.

At this point it might appear that there is not much to gain in performance once the wavelet filters, the threshold and the number of scales used are optimized. However, there is a dramatic improvement possible by giving up orthogonality. We propose the usage of a redundant and shift invariant transform that is closely related to an orthogonal wavelet analysis. The computational complexity is $O(N \log N)$, where N denotes the length of the input signal.

The paper is organized as follows. In the next section we review the traditional wavelet transform and Donoho's method for noise reduction. In section 3 several algorithms for computing a shift invariant wavelet transform are presented. A combination of Beylkin's algorithm and wavelet denoising is proposed in section 4. An analysis of the ideal risk for the new denoising algorithm is included. Also it is shown that the actual risk is within a logarithmic factor of the ideal risk similar to Donoho's method. Several one and two dimensional examples as

*Here σ is assumed to be known. In practice, it can be easily estimated using the methods proposed in.[13,18]

well as a comparison of the ideal risks and the actual risks for Donoho's and the new method support our opinion that the proposed algorithm offers a considerable improvement over denoising with orthogonal wavelets.

2 DENOISING BY THRESHOLDING — A REVIEW

Let

$$y_i = x_i + \epsilon n_i, \quad i = 1, \ldots, N \tag{1}$$

be a finite length signal of observations of the signal x_i that is corrupted by i.i.d. zero mean, white Gaussian noise n_i with standard deviation ϵ, i.e., $n_i \overset{iid}{\sim} \mathcal{N}(0,1)$. The goal is to recover the signal x from the noisy observations y. Here and in the following, v denotes a vector with the ordered elements v_i if the index i is omitted. Let W be a left invertible wavelet transformation matrix of the discrete wavelet transform (DWT). Then Eq. (1) can be written in the transformation domain

$$Y = X + N, \quad \text{or,} \quad Y_i = X_i + N_i, \tag{2}$$

where capital letters denote variables in the transform domain, i.e., $Y = Wy$. Then the inverse transform matrix M exists and we have

$$MW = I. \tag{3}$$

The following presentation is oriented at Donoho's approach[12,13,10,11] that assumes an orthogonal wavelet transform with a square W, i.e., $W^{-1} = W^T$. We will use the same assumption throughout this section.

Let $\hat{X}$ denote an estimate of X, based on the observations Y. We consider diagonal linear projections

$$\Delta = \text{diag}(\delta_1, \ldots, \delta_N), \quad \delta_i \in \{0,1\}, \quad i = 1, \ldots, N, \tag{4}$$

which give rise to the estimate

$$\hat{x} = M\hat{X} = M\Delta Y = M\Delta W y. \tag{5}$$

The estimate $\hat{X}$ is obtained by simply keeping or killing the individual wavelet coefficients. Since we are interested in the l_2 error we define the risk measure

$$\mathcal{R}(\hat{X}, X) = E\left[\|\hat{x} - x\|_2^2\right] = E\left[\|M(\hat{X} - X)\|_2^2\right] = E\left[\|\hat{X} - X\|_2^2\right]. \tag{6}$$

Notice that the last equality in Eq. (6) is a consequence of the orthogonality of W. The optimal coefficients in the diagonal projection scheme are $\delta_i = 1_{X_i > \epsilon}$,[†] i.e., only those values of Y where the corresponding elements of X are larger than ϵ are kept, all others are set to zero. This leads to the ideal risk

$$\mathcal{R}_{id}(\hat{X}, X) = \sum_{n=1}^{N} \min(\hat{X}^2, \epsilon^2). \tag{7}$$

The ideal risk can not be attained in practice, since it requires an oracle that has knowledge of the (unknown) wavelet transform X. However, it gives us a lower limit for the l_2 error.

Donoho proposes the following scheme for denoising:

1. compute the DWT $Y = Wy$

[†] It is interesting to note that allowing arbitrary $\delta_i \in \mathbb{R}$ improves the ideal risk by at most a factor of 2.[14]

2. perform thresholding in the wavelet domain, according to so-called hard thresholding

$$\hat{X} = T_h(Y,t) = \begin{cases} Y, & |Y| \geq t \\ 0, & |Y| < t \end{cases} \tag{8}$$

or according to so-called soft thresholding

$$\hat{X} = T_S(Y,t) = \begin{cases} \mathrm{sgn}(Y)(|Y| - t), & |Y| \geq t \\ 0, & |Y| < t \end{cases} \tag{9}$$

3. compute the inverse DWT $\hat{x} = M\hat{X}$

This simple scheme has several interesting properties. It's risk is within a logarithmic factor ($\log N$) of the ideal risk for both thresholding schemes and properly chosen thresholds $t(N, \epsilon)$. If one employs soft thresholding then the estimate is with high probability at least as smooth as the original function. The proof of this proposition relies on the fact that wavelets are unconditional bases for a variety of smoothness classes and that soft thresholding guarantees (with high probability) that the shrinkage condition $|\hat{X}_i| < |X_i|$ holds. The shrinkage condition guarantees that $\hat{x}$ is in the same smoothness class as is x. Moreover, the soft threshold estimate is the optimal estimate that satisfies the shrinkage condition. The smoothness property guarantees an estimate free from spurious oscillations, other than hard thresholding or Fourier methods. Also, it can be shown that it is not possible to come closer to the ideal risk than within a factor $\log N$. Not only has Donoho's method nice theoretical properties but also works very well in practice.

Some comments have to be made at this point. Similar to traditional approaches, e.g., low pass filtering, there is a tradeoff between suppression of noise and oversmoothing of image details, although to a smaller extend. Also, hard thresholding yields better results in terms of the l_2 error. That is not surprising since the observation value y_i itself is clearly a better estimate for the real value x_i than a shrunk value in a zero mean noise scenario. However, the estimated function obtained from hard thresholding typically exhibits undesired, spurious oscillations as predicted by the theory.

3 SHIFT INVARIANT DISCRETE WAVELET TRANSFORMS

As is well-known, the discrete wavelet transform is not shift invariant, i.e., there is no "simple" relationship between the wavelet coefficients of the original and the shifted signal[‡]. Indeed, let be $X = Wx$ the (orthogonal) DWT of x and S_R be a matrix performing a circular right shift by R, $R \in \mathbb{Z}$, then

$$X_s = Wx_s = WS_Rx = WS_RMX, \tag{10}$$

which establishes the connection between the wavelet transforms of two shifted versions of a signal, x and x_s, by the orthogonal matrix WS_RM. As an illustrative example consider Fig. 1.

The first and obvious way of computing a shift invariant discrete wavelet transform (SIDWT) is simply computing the wavelet transform of *all* shifts. Usually the two band wavelet transform is computed as follows: 1) filter the input signal by a low pass and a high pass filter, respectively, 2) downsample each filter output, and 3) iterate the low pass output. Because of the downsampling, the number of output values at each stage of the filter bank (corresponding to coarser and coarser scales of the DWT) is equal to the number of the input values. Precisely N values have to be stored. The computational complexity is $O(N)$. Directly computing the wavelet transform of all shifts therefore requires the storage of N^2 elements and has computational complexity $O(N^2)$.

[‡] Since we deal with finite length signals, we really mean circular shift.

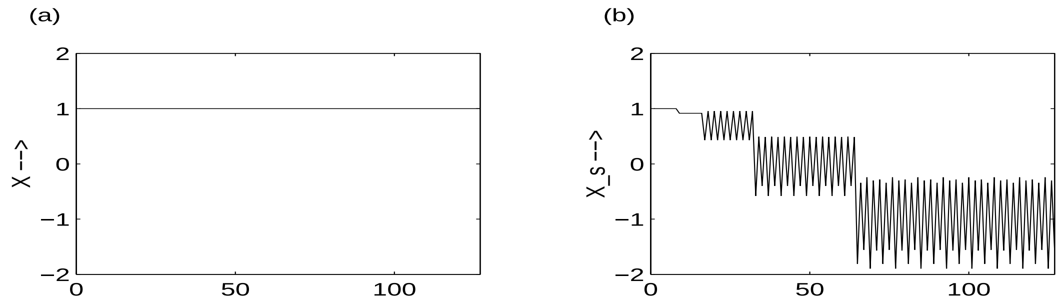

Figure 1: Shift variance of the wavelet transform. (a) Wavelet transform X of a signal x. (b) Wavelet transform X_s of x_s that is a shifted version of x (left shift by one).

Beylkin,[4] Shensa,[31] and our group[§] independently realized that 1) there are only $N \log N$ different coefficient values among those corresponding to all shifts of the input signal and 2) those can be computed with computational complexity $N \log N$. This can be easily seen by considering one stage of the filter bank. Let be

$$y = [y_0\, y_1\, y_2\, \ldots\, y_N]^T = hx \tag{11}$$

where y is the output of either the high pass or the low pass filter in the analysis filter bank, x the input and the matrix h describes the filtering operation. Downsampling of y by a factor of two means keeping the even indexed elements and discarding the odd ones. Consider the case of an input signal shifted by one. Then the output signal is shifted by one as well and sampling with the same operator as before corresponds to keeping the odd indexed coefficients as opposed to the even ones. Thus, the set of data points to be further processed is completely different. However, for a shift of the input signal by two, the downsampled output signal differs from the output of the nonshifted input only by a shift of one. This is easily generalized for any odd and even shift and we see that the set of wavelet coefficients of the first stage of the filter bank for arbitrary shifts consists of only $2N$ different values. Considering the fact that only the low pass component (N values) is iterated, one recognizes that after L stages exactly LN values result. Using the same arguments as in the shift variant case, one can prove that the computational complexity is $O(N \log N)$. The derivation for the synthesis is analogous.

Mallat proposes a scheme for computing an approximation of the continuous wavelet transform[22] that turns out to be equivalent to the method described above. This has been realized and proved by Shensa.[31] Moreover, Shensa shows that Mallat's algorithm exhibits the same structure as the so-called algorithm à trous.[19,15] Interestingly, Mallat's intention in[22] was not in particular to overcome the shift variance of the DWT but to get an approximation to the continuous wavelet transform.

In the following we shall refer to the algorithm for computing the SIDWT as the Beylkin algorithm[¶] since this is the one we have implemented. Alternative algorithms for computing a shift invariant wavelet transform[9,21,2] are based on the scheme presented in.[4] They explicitly or implicitly try to find an optimal, signal dependent shift of the input signal. Thus, the transform becomes shift invariant and orthogonal but signal dependent.

We mention that the generalization of the Beylkin algorithm to the multidimensional case, to an M band multiresolution analysis, and to wavelet packets is straightforward.

[§] Those are the ones we are aware of.
[¶] However, it should be noted that Mallat published his algorithm earlier.

4 COMBINING THE SHENSA-BEYLKIN-MALLAT-À TROUS ALGORITHM AND WAVELET DENOISING

We acknowledge a private conversation of one of the authors with Dr. Coifman who stated that the application of Donoho's method to several shifts of the observation combined with averaging yields a considerable improvement.‖ This statement first lead us to the following algorithm: 1) apply Donoho's method not only to "some" but to *all* circular shifts of the input signal 2) average the adjusted output signals. As has been shown in the previous section, the computation of all possible shifts can be effectively done using Beylkin's algorithm. Thus, instead of using the algorithm just described, one simply applies thresholding to the SIDWT of the observation and computes the inverse transform.

Before going into details we want to briefly discuss the differences between using the traditional orthogonal and the shift invariant wavelet transform. Obviously, by using more than N wavelet coefficients we introduce redundancy. Several authors stated that redundant wavelet transforms, or frames, add to the numerical robustness[7,3,25] in case of adding white noise in the transform domain, e.g., by quantization. This is, however, different from the scenario we are interested in since 1) we have correlated noise due to the redundancy and 2) we try to remove noise in the transform domain rather than considering the effect of adding some noise. It is worthwhile mentioning that Starck, Murtagh and Bijaoui[32] have applied thresholding to astronomical signals transformed by the algorithm à trous which is closely related to the SIDWT. However, they seem not to be aware of the ongoing research in denoising by thresholding the wavelet coefficients.

4.1 Performance analysis

The analysis of the ideal risk for the SIDWT is similar to that by Guo.[17] Define the sets A and B according to

$$A = \{i \mid |X_i| \geq \epsilon\} \tag{12}$$
$$B = \{i \mid |X_i| < \epsilon\} \tag{13}$$

and an ideal diagonal projection estimator, or oracle,

$$\tilde{X} = \begin{cases} Y_i = X_i + N_i & i \in A \\ 0 & i \in B. \end{cases} \tag{14}$$

The pointwise estimation error is then

$$\tilde{X}_i - X_i = \begin{cases} N_i & i \in A \\ -X_i & i \in B. \end{cases} \tag{15}$$

In the following a vector or matrix indexed by A (or B) indicates that only those rows are kept that have indices out of A (or B). All others are set to zero. With these definitions and Eq. (6) the ideal risk for the SIDWT can be derived

$$\begin{aligned} \mathcal{R}_{id}(\tilde{X}, X) &= E\left[||M(\tilde{X} - X)||_2^2\right] & (16)\\ &= E\left[||M(N_A - X_B)||_2^2\right] & (17)\\ &= E\left[(N_A - X_B)^T \underbrace{M^T M}_{C_M}(N_A - X_B)\right] & (18) \end{aligned}$$

‖A similar remark can be found in,[29] p. 53.

$$
\begin{aligned}
&= E\left[N_A^T M^T M N_A\right] - 2X_B^T C_M E\left[N_A\right] + X_B^T C_M X_B & (19)\\
&= \mathrm{tr}\left[E\left[M N_A N_A^T M^T\right]\right] + X_B^T C_M X_B & (20)\\
&= \mathrm{tr}\left[M E\left[W_A \epsilon n \epsilon n^T W_A^T\right] M^T\right] + X_B^T C_M X_B & (21)\\
&= \epsilon^2 \mathrm{tr}\left[M W_A W_A^T M^T\right] + X_B^T C_M X_B. & (22)
\end{aligned}
$$

$\mathrm{tr}(X)$ denotes the trace of X. For the derivation we have used the fact that $N_A = \epsilon W_A n$ and consequently the N_{Ai} have zero mean. Notice that for orthogonal W the Eq. (22) immediately specializes to Eq. (7). Eq. (22) depends on the particular signal X_B, the transform, M, and the noise level ϵ.

It can be shown that when using the SIDWT introduced above and the thresholding scheme proposed by Donoho (including his choice of the threshold) then there exists the same upper bound for the actual risk as for case of the orthogonal DWT. That is the ideal risk times a logarithmic (in N) factor. We give only an outline of the proof. Johnstone and Silverman state[20] that for colored noise the oracle chooses $\delta_i = 1_{X_i \geq \epsilon_i}$, where ϵ_i is the standard deviation of the ith component. Since Donoho's method applies uniform thresholding to all components, one has to show that the diagonal elements of C_M (the variances of the components of N) are identical. This can be shown by considering the reconstruction scheme of the SIDWT. With these statements the rest of the proof can be carried out in the same way as the one given by Donoho.[13]

5 EXAMPLES

For the examples 1 and 2 that deal with one dimensional signals we use the following set of parameters. The scaling filter is a Daubechies filter of length 6 with a maximum number of vanishing moments, the signals are of length 512, and the number of scales (or filter bank stages) used is 7. The threshold is chosen as the product of the median absolute deviation estimate for the standard deviation[13] and a fixed number, i.e., 3 for the Donoho method (soft thresholding) and 3.6 for the new method (hard thresholding). The signals are generated by Donoho's MATLAB routine `MakeSignal` from his software package `TeachWave`.

Example 1: The signal 'Doppler' depicted in Fig. 2(a) is denoised using Donoho's and our new method. Figs. 2(b-d) (signal to noise ratio, or SNR, of 0dB) and 2(e-g) (SNR of 20dB) show the noisy signal, the signal denoised with Donoho's method, and with the new method, respectively. In both cases (0dB and 20dB), the new method yields by far better results. One feature that seems to be of special interest is that although we apply hard thresholding to the new method (getting better estimates in the l_2 sense) the signal is considerably more smooth than that resulting from Donoho's approach.

Example 2: We compare the resulting SNRs (denoted SNR_{out}) of Donoho's denoising scheme and the new one for different SNRs (denoted SNR_{in}) of the noisy signal. For a given SNR_{in}, we have used 100 noise realizations to get an estimate for the expected value of the resulting SNR_{out}. Fig. 3(a-c) shows the result for three signals, 'Bumps', 'Blocks', and 'Doppler' (cf. Example 1). The x-axis denotes the given SNR_{in}, the y-axis the resulting SNR_{out}. The dotted line corresponds to no processing.

For low SNR (< -10dB) there is essentially no difference in performance for both methods. For larger SNR the new method clearly outperforms Donoho's method (more than 10dB difference for the signal 'Bumps' and SNR_{in} of 20dB) with the tendency of an increasing gap between the SNRs. For high SNR Donoho's method yields results that are even worse than the unprocessed signals.[**] The new method, however, always yields better results and seems to approach a line parallel to the one corresponding to the unprocessed case. This is especially interesting in a scenario where the SNR is unknown. It appears that the new method is robust with respect to a large range of values of the SNR and a variety of signals.

[**] This might be due to the fact that soft thresholding is used instead of hard thresholding.

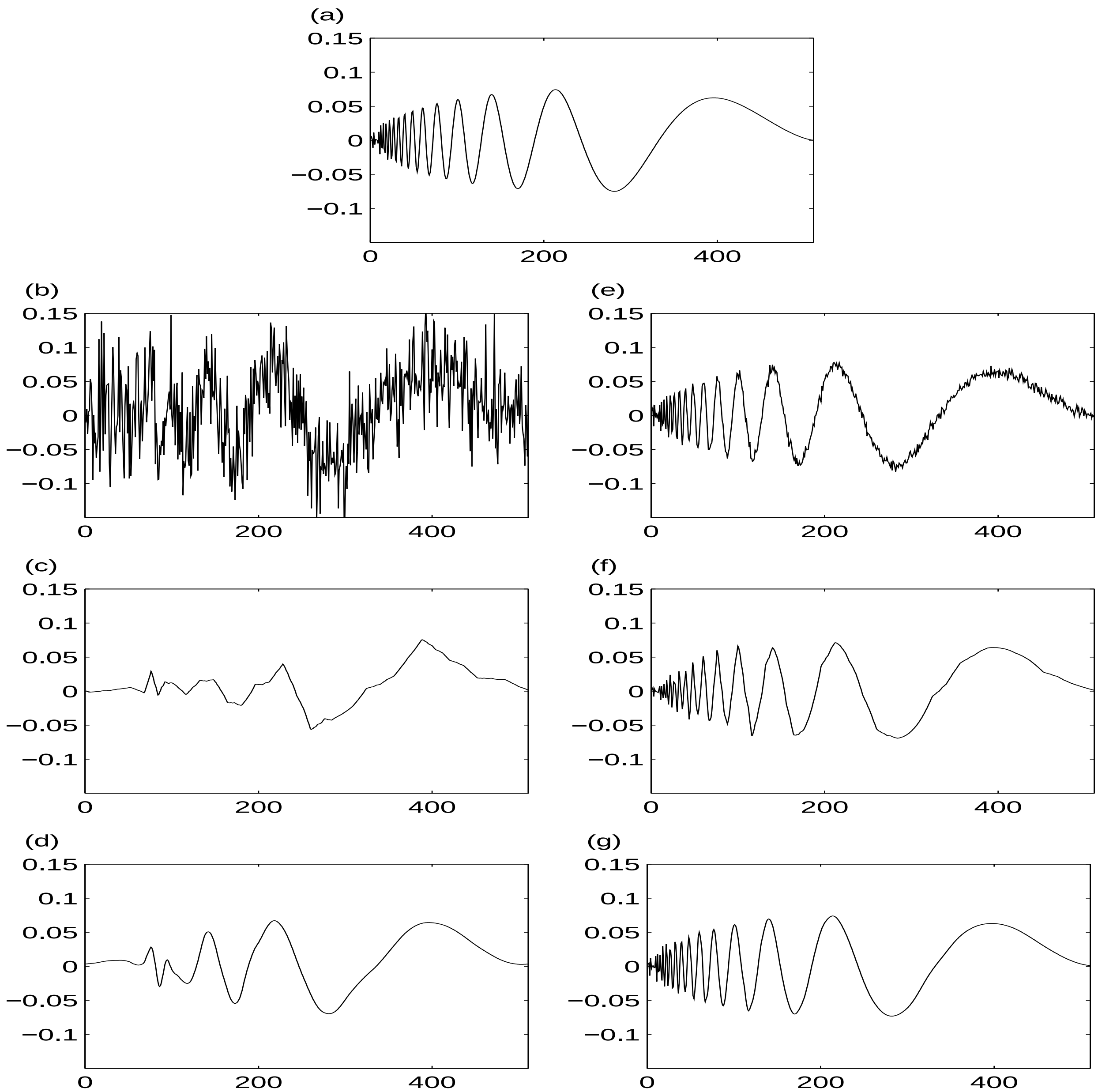

Figure 2: Denoising of signal 'Doppler': (a) Original image. (b) Noisy signal (SNR of 0dB). (c) Signal corresponding to (b) using Donoho's method. (d) Signal corresponding to (b) using the new method.(e) Noisy signal (SNR of 20dB). (f) Signal corresponding to (e) using Donoho's method. (g) Signal corresponding to (e) using the new method.

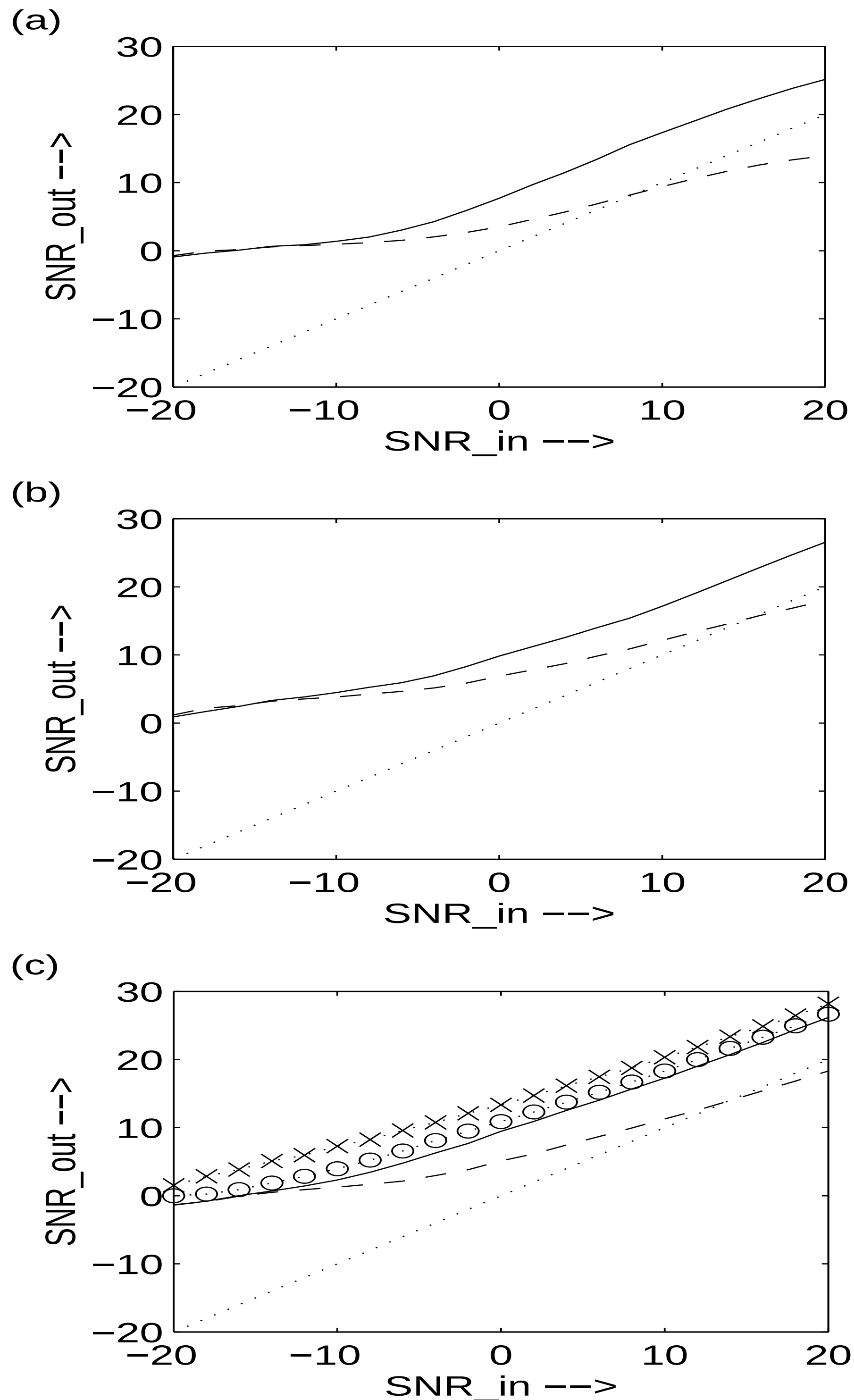

Figure 3: Improvement of the signal to noise ratio for three different signals[13]: dotted line SNR of noisy signal, dashed line SNR resulting from thresholding the DWT, solid line SNR resulting from thresholding the SIDWT. (a) Bumps. (b) Blocks. (c) Doppler. The dotted lines marked by o and x depict the SNRs corresponding to the ideal risks of Donoho's and the new method.

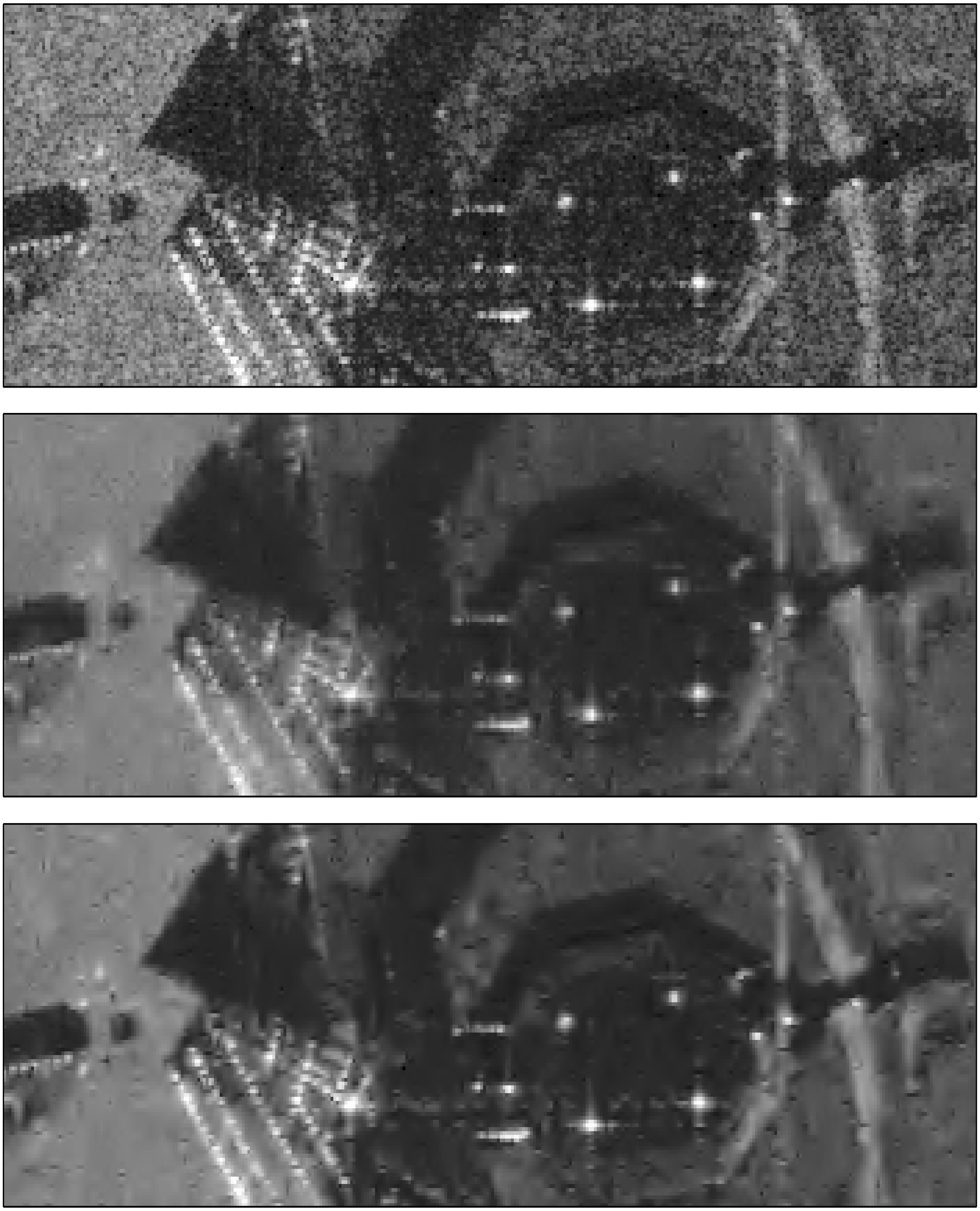

Figure 4: Denoising of a SAR image, courtesy Lincoln Lab. From the top: original image, denoised by thresholding the DWT, denoised by thresholding the SIDWT.

Fig. 3 additionally includes the SNRs corresponding to the ideal risk according to the Eq. (22). The difference between the ideal risks of Donoho's and the new method is nearly independent of the SNR and about 3dB. In contrast to Donoho's method the new one closely follows the ideal risk.

Example 3: Images resulting from a coherent imaging process, e.g. SAR images, are known to exhibit so-called speckle noise. The AWGN model is a very good approximation to the noise properties if one considers the logarithm of the absolute pixel values.[1] Thus SAR images perfectly fit the assumptions made for denoising. The top of Fig. 4 shows a SAR image after processing with a polarimetric whitening filter.[27] The resulting image using Donoho's and the new method are given in the middle and at the bottom of Fig. 4. The new algorithm yields a much sharper image (check the brighter spots around the black oval to the left) and preserves the intensity of the bright spots much better.

6 SUMMARY AND FUTURE WORK

We have presented a new method for denoising one or two dimensional signals. It combines Beylkin's algorithm for computing the shift invariant, redundant wavelet transform and Donoho's idea of applying thresholding to the (orthogonal) wavelet transform. We show that several of the theoretical results of Donoho, proved for the case of an orthogonal DWT, also apply for the new algorithm. Several examples show that the performance is superior to that of Donoho's method.

However, there are several open questions for future research: Why is the resulting signal smoother despite the fact that we use hard thresholding? Is it possible to prove a tighter upper risk bound than $\log N$ times the ideal risk (cf. Fig. 3(c))? For which class of signals and transforms is the ideal risk of the new method smaller than that of Donoho's? Is there a considerable advantage of using all SIDWT coefficients rather than those of the DWT of the best shift? Can the performance be considerably improved by using wavelet packets? Does using an information theoretic scheme such as the MDL affect the performance? The new method also promises improvements for coding and compression schemes that are currently investigated.[28,33]

Software is available from the World Wide Web at `http://www-dsp.rice.edu` or by anonymous ftp from `cml.rice.edu` in the directory `/pub/software`.

7 REFERENCES

[1] H. H. Arsenault and G. April. Properties of speckle integrated with a finite aperture and logarithmically transformed. *J. Opt. Soc. Am.*, 66:1160–1163, November 1976.

[2] F. Bao and N. Erdol. Optimal initial phase wavelet transform. In *Sixth Digital Signal Processing Workshop*, pages 187–190, October 1994.

[3] J. J. Benedetto and S. Li. Subband coding and noise reduction in multiresolution analysis frames. In *Wavelet Applications in Signal and Image Processing*, volume 2242, pages 154–165, San Diego, CA, July 1994. SPIE.

[4] G. Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM J. Numer. Anal.*, 29(6):1716–1740, 1992.

[5] R. R. Coifman and F. Majid. Adapted waveform analysis and denoising. In Y. Meyer and S. Roques, editors, *Progress in Wavelet Analysis and Applications (Proceedings of the International Conference "Wavelets and Applications", Toulouse, France, June, 1992)*, pages 63–76. Editions Frontieres, B.P. 33, 91192 Gif-sur-Yvette, Cedex, France, 1993.

[6] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, 38(2):1713–1716, 1992.

[7] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[8] G. Davis and S. Mallat. Wavelet vector quantization with matching pursuit. In *Workshop on Information Theory and Statistics*, page 55. IEEE, october 1994.

[9] S. del Marco, J. Weiss, and K. Jager. Wavepacket-based transient signal detector using a translation invariant wavelet transform. In *Wavelet Applications in Signal and Image Processing*, volume 2242, pages 154–165, San Diego, CA, July 1994. SPIE.

[10] D. L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics*, volume 00, pages 173–205. American Mathematical Society, 1993.

[11] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1(1):100–115, December 1993.

[12] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 1994. Also Tech. Report 409, Department of Statistics, Stanford University.

[13] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. To appear, Also Tech. Report 400, Stanford University, Department of Statistics, July, 1992, revised April 1993.

[14] D. L. Donoho and I. M. Johnstone. Ideal denoising in an orthonormal basis chosen from a library of bases. Technical report, September 1994.

[15] P. Dutilleux. An implementation of the "algorithme à trou" to compute the wavelet transform. In J. M. Combes, A. Grossman, and Ph. Tchamitchian, editors, *Wavelets Time-Frequency Methods and Phase Space*, pages 2–20, Berlin Heidelberg, 1989. Springer-Verlag.

[16] R. A. Gopinath, M. Lang, H. Guo, and J. E. Odegard. Enhancement of decompressed images at low bit rates. In *SPIE Math. Imagaging: Wavelet Applications in Signal and Image Processing*, volume 2303, pages 366–377, San Diego, CA, July 1994. Also Tech report CML TR94-05, Rice University, Houston, TX.

[17] H. Guo. Redundant wavelet transform and denoising. Technical report, Rice University, Houston, TX, 1995.

[18] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. Selesnick, and C. S. Burrus. Speckle reduction via wavelet shrinkage with application to SAR based ATD/R. In *SPIE Math. Imagaging: Wavelet Applications in Signal and Image Processing*, volume 2303, pages 333–344, San Diego, CA, July 1994. Also Tech report CML TR94-03, Rice University, Houston, TX.

[19] M. Holschneider, R. Kronland-Martinet, J. Morlet, and Ph. Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In J. M. Combes, A. Grossman, and Ph. Tchamitchian, editors, *Wavelets Time-Frequency Methods and Phase Space*, pages 286–297, Berlin Heidelberg, 1989. Springer-Verlag.

[20] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. Technical report, University of Bristol, UK, Statistics Department, September 1994.

[21] J. Liang and T. W. Parks. A two-dimensional translation invariant wavelet representation and its applications. In *Proc. Int. Conf. Image Processing*, volume I, pages 66–70, Austin, TX, November 1994. IEEE.

[22] S. Mallat. Zero-crossings of a wavelet transform. *IEEE Trans. Inform. Theory*, 37(4), July.

[23] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. Technical Report 619, New York University, Courant Institute of Mathematical Sciences, 1992.

[24] P. Moulin. A wavelet regularization method for diffuse radar-target imaging and speckle-noise reduction. *Journal of Mathematical Imaging and Vision*, 3(1):123–134, January 1993.

[25] N. J. Munch. Noise reduction in tight Weyl-Heisenberg frames. *IEEE Trans. Inform. Theory*, 38(2):608–616, March 1992.

[26] G. P. Nason. Wavelet regression by cross-validation. Technical Report 447, Stanford University, Statistics Department, March 1994.

[27] L. M. Novak, M. C. Burl, and W. W. Irving. Optimal polarimetric processing for enhanced target detection. *IEEE Trans. AES*, 29:234–244, January 1993.

[28] J. E. Odegard, H. Guo, M. Lang, C. S. Burrus, R. O. Wells, L. M. Novak, and M. Hiett. Wavelet based SAR speckle reduction and image compression. In *Algorithms for Synthetic Aperture Radar Imagery II at AeroSense '95*, Orlando, FL, April 1995.

[29] N. Saito. *Local Feature Extraction and Its Applications Using a Library of Bases.* PhD thesis, Yale University, December 1994.

[30] N. Saito. Simultaneous noise suppresion and signal compression using a library of orthonormal bases and the minimum description length criterion. In *Wavelet Applications*, volume 2242, pages 224–235, Bellingham, WA, 1994. SPIE.

[31] M. J. Shensa. The discrete wavelet transform: Wedding the à trous and Mallat algorithms. *IEEE Trans. Inform. Theory*, 40:2464–2482, 1992.

[32] J.-L. Starck, F. Murtagh, and A. Bijaoui. Image restoration with noise suppresion using a wavelet transform and a multiresolution support constraint. In *Image Reconstruction and Restoration*, volume 2302, pages 132–143, Bellingham, WA, 1994. SPIE.

[33] D. Wei, H. Guo, J. E. Odegard, M. Lang, and C. S. Burrus. Simultaneous speckle reduction and data compression using best wavelet packet bases with application to SAR based ATD/R. In *SPIE conference on wavelet applications*, volume 2491, Orlando, FL, April 1995.

[34] N. Weyrich and G. T. Warhola. De-noising using wavelets and cross validation. Technical Report AFIT/EN/TR/94-01, Air Force Institute of Technology, Department of Mathematics and Statistics, June 1994.