Minimax Threshold for Denoising Complex Signals with Waveshrink

Sylvain Sardy

***Abstract*—For the problem of signal extraction from noisy data, Waveshrink has proven to be a powerful tool, both from an empirical and an asymptotic point of view. Waveshrink is especially efficient at estimating spatially inhomogeneous signals. A key step of the procedure is the selection of the threshold parameter. Donoho and Johnstone propose a selection of the threshold based on a minimax principle. Their derivation is specifically for real signals and real wavelet transforms. In this paper, we propose to extend the use of Waveshrink to denoising complex signals with complex wavelet transforms. We illustrate the problem of denoising complex signals with an electronic surveillance application.**

***Index Terms*—Complex signals, complex wavelet transform, electronic surveillance, minimax, nonparametric denoising, waveshrink.**

I. Background

SUPPOSE we observe a univariate real signal $\underline{s} = (s_1, s_2, \cdots, s_N) \in \mathbb{R}^N$ at equispaced locations x_n according to the model

$$s_n = f(x_n) + z_n, \quad n = 1, \cdots, N$$

where the z_n are identically and independently distributed standard Gaussian random variables. Therefore, we assume that the variance of the noise is known and unity (i.e., $\sigma = 1$). In practice, the variance can be estimated by taking the median absolute deviation of the high level wavelet coefficients, as proposed in [1]. Our goal is to find a good estimate $\underline{\hat{f}}$ of the underlying signal $\underline{f} = (f(x_1), \cdots, f(x_N))$. The predictive performance of $\underline{\hat{f}}$ is measured by the risk (also called mean squared error), which is defined as

$$R(\underline{\hat{f}}, \underline{f}) = \frac{1}{N} E\|\underline{\hat{f}} - \underline{f}\|_2^2 \tag{1}$$

where E stands for the expectation over the observed noisy signal $\underline{s}$. Expansion-based nonparametric estimators assume that the underlying signal can be well approximated by a linear combination of P known basis functions ϕ_p (e.g., splines, Fourier trigonometric functions), namely, that $f(x) \approx \sum_{p=1}^{P} \mu_p \phi_p(x)$. Once a set of basis functions is chosen, the only quantities to estimate are the basis function coefficients $\underline{\mu} = (\mu_1, \cdots, \mu_P)$. The signal estimate is then $\underline{\hat{f}} = \Phi\underline{\hat{\mu}}$, where Φ is the matrix of discretized ϕ_p. The hat on top of a variable is the notation used throughout this paper to indicate the estimate of the corresponding variable.

Manuscript received January 14, 1999; revised October 26, 1999. This work was supported in part by NSF Phase II and Navy Phase I SBIR Award N68936-97-C-0150 to MathSoft, Inc. The associate editor coordinating the review of this paper and approving it for publication was Prof. Victor A. N. Barroso.

The author is with the Department of Mathematics, Swiss Federal Institute of Technology (EPFL-DMA), Lausanne, Switzerland (e-mail: Sylvain.Sardy@epfl.ch).

Publisher Item Identifier S 1053-587X(00)02361-8.

A. Waveshrink

Waveshrink is an expansion-based estimator proposed by Donoho and Johnstone [2]. The expansion is on wavelets: a set of $P = N$ compactly supported and orthonormal functions on the real line. Under the hypothesis that the underlying signal is periodic, the wavelets are orthonormal on the support of the signal, and the Φ matrix is orthonormal. Let us denote by Φ' the transpose of Φ. The orthonormality property has two important consequences. First, the least squares estimate of the wavelet coefficients

$$\underline{\hat{\mu}} = \Phi'\underline{s} \sim \text{Normal}(\underline{\mu}, I_N) \tag{2}$$

is unbiased, and the coefficients are independent of each other with same variance. Second, the risk in function values (1) equals the risk in coefficient values; for any estimate $\underline{\hat{\mu}}$ of $\underline{\mu}$, we have $R(\underline{\hat{f}}, \underline{f}) = R(\underline{\hat{\mu}}, \underline{\mu})$. Therefore, we can concentrate on estimating the wavelet coefficients and measure the predictive performance of an estimator in the coefficient values.

The least squares estimate is unbiased but does not denoise the original signal since $\underline{\hat{f}} = \Phi(\Phi'\underline{s}) = \underline{s}$. The corresponding risk equals the variance of the noise, namely, $R(\underline{\hat{\mu}}, \underline{\mu}) = 1$. To estimate the wavelet coefficients with a smaller risk at the cost of introducing some bias, Donoho and Johnstone [2] propose to apply, componentwise, a function $\eta_\lambda(\cdot)$ that shrinks the least squares estimate toward zero to obtain the estimate

$$\underline{\hat{\mu}}_\lambda = \eta_\lambda(\underline{\hat{\mu}}) \tag{3}$$

where λ is the threshold parameter of the shrinkage function, which is a meta parameter of the Waveshrink procedure. The coefficient estimate is defined for a given λ, and the shrinkage is performed component-wise. The risk of the estimate is the sum of the risks component-wise, namely

$$R(\lambda, \underline{\mu}) = \sum_{n=1}^{N} \rho(\lambda, \mu_n) \tag{4}$$

with $\rho(\lambda, \mu_n) = (\hat{\mu}_{\lambda,n} - \mu_n)^2$. Because a wide class of functions f can be well approximated by the linear combination of a few wavelets, Donoho and Johnstone [2] propose to enforce sparsity by using either the hard shrinkage function

$$\eta_\lambda^{(hard)}(x) = x \cdot 1(|x| > \lambda)$$

where $1(x \in A)$ is the identity function on A or the soft shrinkage function

$$\eta_\lambda^{(soft)}(x) = \text{sign}(x) \cdot (|x| - \lambda)_+ \tag{5}$$

where x_+ is x for $x > 0$ and zero otherwise. We see that any value of the least squares coefficients between $-\lambda$ and λ is set to zero, hence, enforcing sparsity. Other shrinkage methods have been proposed, e.g., non-negative Garrote [3], Firm [4], or Bayesian [5]. However, more important than the choice of a shrinkage function, the selection of λ is the crucial step of the Waveshrink procedure. As in [2], we concentrate, in this paper, on the selection of the threshold λ for the soft shrinkage function (5).

B. Threshold Selection

The threshold parameter λ controls the bias-variance tradeoff of the risk (4). Its selection is crucial for Waveshrink to give a good estimation of the underlying signal. Donoho and Johnstone [2] proposed a selection of the meta parameter based on a minimax principal. Their approach can be summarized in three steps.

- *Oracle risk for diagonal linear projection*: They considered a "diagonal linear projection" estimator that keeps or kills each least squares coefficient with a different meta parameter δ_n, namely, $\underline{\hat{\mu}}_{\underline{\delta}} = \text{diag}(\delta_1, \cdots, \delta_N)\underline{\hat{\mu}}$, where $\underline{\delta} \in \{0,1\}^N$. This estimator has a total of N meta parameters: one for each least squares coefficient. It would be difficult in practice to estimate the N meta parameters. Donoho and Johnstone invoked an oracle, i.e., the knowledge of the quantity to estimate $\underline{\mu}$, and considered selecting $\underline{\delta}$ by minimizing the risk. The optimal risk is

$$R(\text{DP}, \underline{\mu}) := \min_{\underline{\delta}} R(\underline{\hat{\mu}}_{\underline{\delta}}, \underline{\mu}) = \sum_{n=1}^{N} \min(|\mu_n|^2, 1). \tag{6}$$

  In practice, we would like to closely approach this ideal risk.
- *Universal threshold*: They considered the Waveshrink estimate (3) with the soft shrinkage function, and select the single meta parameter $\lambda_N = \sqrt{2 \log N}$, so that the estimator achieves the performance of the oracle estimator within a factor of essentially $2 \log N$ for all possible true coefficients, namely

$$R(\underline{\hat{\mu}}_{\lambda_N}, \underline{\mu}) \leq (2 \log N + 1) \left(1 + R(\text{DP}, \underline{\mu})\right) \tag{7}$$

  for all $\underline{\mu} \in \mathbb{R}^N$. The universal threshold also has the advantage that if the signal is in fact white noise, i.e., $s_n = z_n$, then the signal will be estimated as zero with high probability since

$$P(\max_n |z_n| > \sqrt{2 \log N}) \longrightarrow 0 \quad \text{as} \quad N \longrightarrow \infty.$$

  Hence, the universal threshold has the advantage of giving a signal estimate with a nice visual appearance (see, for instance, [1]).
- *Minimax threshold*: They improve the bound $\Lambda_N = 2 \log N + 1$ and select the threshold λ_N^* to achieve the minimum bound Λ_N^*, namely,

$$R(\underline{\hat{\mu}}_{\lambda_N^*}, \underline{\mu}) \leq \Lambda_N^* \left(1 + R(\text{DP}, \underline{\mu})\right)$$

  for all $\underline{\mu} \in R^N$, $\Lambda_N^* \leq 2 \log N + 1$, and $\Lambda_N^* \sim 2 \log N$ asymptotically. The minimax threshold does not usually give an estimate with a nice visual appearance. However, it has the advantage of giving good predictive performance.

For details on the properties of these two thresholds, see [1] and [2].

II. WAVESHRINK FOR COMPLEX SIGNALS

Thus far, it has been assumed that the signal is real valued. In some applications, however, the signal is complex valued (see, for instance, our application in Section III), and the wavelets have real and imaginary parts. Examples of complex wavelets include the complex Daubechies wavelets [6], chirplets [7], and brushlets [8].

To denoise the signal, both the real and imaginary parts of the least squares coefficients have to be shrunk toward zero. A simple shrinkage procedure would consist of independently shrinking the real and imaginary parts of the least squares wavelet coefficients (2). One drawback of this procedure is that the underlying signal estimate is not guaranteed to have a sparse wavelet representation (a shrunk coefficient may have its real part set to zero but not its imaginary part, and vice versa). Moreover, the phases of the least squares coefficients are changed by this kind of shrinkage. If instead the moduli of the least squares coefficients are shrunk, then both the real and imaginary parts are guaranteed to be set to zero together, and the phases remain the same. Shrinking the moduli is the natural generalization of Waveshrink to complex signals since it forces sparsity on the wavelet representation.

To be more specific, suppose we observe a univariate complex signal $\underline{s} = (s_1, s_2, \cdots, s_N) \in C^N$ at equispaced locations x_n according to the model

$$s_n = f(x_n) + z_n, \qquad n = 1, \cdots, N$$

where the $z_n = z_{1n} + iz_{2n}$ are identically and independently distributed complex random variables with $(z_{1n}, z_{2n})' \sim \text{Normal}(\underline{0}, I_2)$. In addition, let us again assume that the underlying complex function can be well approximated by a linear combination of wavelets, namely, $f(x) \approx \sum_{p=1}^{P} \mu_p \phi_p(x)$, where now, the μ's are complex coefficients, and the $\phi(\,)$'s are complex wavelets. As in the real case (3), we define the soft-Waveshrink estimate by $\underline{\hat{\mu}}_\lambda = \eta_\lambda^{(soft)}(\underline{\hat{\mu}})$, where $\underline{\hat{\mu}} = \Phi' \underline{s}$ is the (complex) least squares estimate. We generalize the soft shrinkage function to complex values as

$$\eta_\lambda^{(soft)}(\mu) = \frac{\mu}{|\mu|} \cdot (|\mu| - \lambda)_+ \tag{8}$$

where $|\mu|$ is now the modulus of the complex number μ. After applying the soft shrinkage function to the least squares coefficients, the moduli are shrunk toward zero and the phases of the least squares coefficients are unchanged. Any complex least squares coefficient which modulus is less than λ is set to zero; the sparsity in the wavelet representation of the underlying complex signal is ensured.

As in the real case, the key step is in the selection of the threshold parameter λ. The selection of the meta parameter must, however, be adapted to complex values since the distribution of the moduli of the least squares wavelet coefficients is no longer Gaussian. In the following, we will essentially follow the three steps of Section I-B for deriving the universal and minimax thresholds; in Section II-A1, we derive the oracle risk for diagonal linear projection; in Section II-A2, we achieve the oracle risk to a factor of $1+\log(N \log N)$ with the universal threshold λ_N; finally, in Section II-A3, we find the minimax threshold λ_N^* that achieves the optimal factor Λ_N^*.

A. Threshold Selection

As in [2], we restrict our attention to one component $\rho(\cdot,\cdot)$ of the risk (4). Let us call Y a least squares component and Y_1, Y_2 its real and complex parts. Their distribution is $(Y_1,Y_2)' \sim \text{Normal}(\mu, I_2)$, where $\mu=(\mu_1,\mu_2)'$. For any estimate $\hat{\mu}$ of μ, the risk of the component is

$$\rho(\lambda,\mu)=E\left[(\hat{\mu}_1-\mu_1)^2+(\hat{\mu}_2-\mu_2)^2\right]. \tag{9}$$

For clarity in the following derivations, we postpone the mathematical derivations to Appendix A and Appendix B.

1) Oracle Predictive Performance for Diagonal Linear Projection: One component of the estimate is $\hat{\mu}=\delta Y$ with $\delta \in \{0,1\}$. The corresponding risk is

$$\rho(\delta,\mu)=\begin{cases}|\mu|^2, & \text{if } \delta=0\\ 2, & \text{if } \delta=1.\end{cases}$$

Therefore, the oracle binary meta parameter is $\delta=1_{\{|\mu|>\sqrt{2}\}}$, and the corresponding oracle risk is $\rho(DP,\mu)=\min(|\mu|^2,2)$. Note that the least squares estimate has risk $\rho(1,\mu)=2$. After repeating the oracle projection to each wavelet coefficient independently, we find that the oracle risk is

$$R(DP,\underline{\mu})=\sum_{n=1}^{N}\min(|\mu_n|^2,2).$$

This is analogous to (6) for complex signals, and it constitutes our reference predictive performance.

2) Universal Threshold: In Appendix A, we rewrite the risk defined in (9) in the convenient form of (14). Note that the three terms of the second and third lines of (14) are negative. Therefore, for $\delta > 0$ small enough and letting $\lambda=\sqrt{2\log(\delta^{-1}\log\delta^{-1})}$, we have on the one hand

$$\begin{aligned}\rho(\lambda,\mu) &\le 2+\lambda^2\\ &\le (2+\lambda^2)(1+\delta)\\ &=(1+\log(\delta^{-1}\log\delta^{-1}))(2\delta+2).\end{aligned}$$

On the other hand, from Property (P1) and the second point of Property (P2), we have

$$\begin{aligned}\rho(\lambda,\mu) &\le \rho(\lambda,(0,0))+|\mu|^2\\ &=2\exp(-\lambda^2/2)-2\sqrt{2\pi}\lambda\Phi(-\lambda)+|\mu|^2\\ &\le 2\exp(-\lambda^2/2)(1+\lambda^2/2)+|\mu|^2(1+\lambda^2/2)\\ &=\left(\frac{2}{\delta^{-1}\log\delta^{-1}}+|\mu|^2\right)(1+\log(\delta^{-1}\log\delta^{-1}))\\ &\le(2\delta+|\mu|^2)(1+\log(\delta^{-1}\log\delta^{-1})).\end{aligned}$$

Putting both inequalities together, we have

$$\rho(\lambda,\mu)\le(1+\log(\delta^{-1}\log\delta^{-1}))(2\delta+\min(|\mu|^2,2))$$

for δ small enough such that $\log(\delta^{-1})\ge 1$, for instance $\delta=1/N$ with $N\ge 3$. Hence, we define the universal threshold $\lambda_N=\sqrt{2\log(N\log N)}$. With that choice of the threshold parameter, soft-Waveshrink approaches the oracle risk by a factor of essentially $\Lambda_N=(1+\log(N\log N))$, namely

$$\begin{aligned}R(\lambda_N,\underline{\mu}) &=\sum_{n=1}^{N}\rho(\lambda_N,\mu_n)\\ &\le(1+\log(N\log N))\cdot R(DP,\underline{\mu}).\end{aligned}$$

This equation is analogous to (7). As in the real case, as the number of observations N becomes large, and only the predominant features of a signal remain after denoising. Indeed, for N independent and identically distributed standard Gaussian complex random variables Z_n,, we have

$$P(\max_n|Z_n|>\sqrt{2\log(N\log N)})\longrightarrow 0 \quad \text{as} \quad N\longrightarrow\infty. \tag{10}$$

3) Minimax Threshold: Define the minimax quantities

$$\begin{aligned}\Lambda_N^* &\equiv \inf_{\lambda}\sup_{\mu\ge 0}\frac{\rho(\lambda,\mu)}{2/N+\min(|\mu|^2,2)}\\ \lambda_N^* &\equiv \text{the largest } \lambda \text{ attaining } \Lambda_N^* \text{ above.}\end{aligned}$$

Then, the overall risk for the N wavelet coefficients is

$$\begin{aligned}\sum_{n=1}^{N}\rho(\lambda_N^*,\mu_n) &\le \Lambda_N^*\left\{2+\sum_{n=1}^{N}\min(|\mu_n|^2,2)\right\}\\ &=\Lambda_N^* R(DP,\underline{\mu}).\end{aligned}$$

To find λ_N^*, consider the analogous quantity, where the supremum over $[0,\infty)\times[0,\infty)$ is replaced by the supremum over the endpoints $\{(0,0),(\infty,\infty)\}$, namely

$$\Lambda_N^0\equiv\inf_{\lambda}\sup_{\mu\in\{(0,0),(\infty,\infty)\}}\frac{\rho(\lambda,\mu)}{2/N+\min(|\mu|^2,2)} \tag{11}$$

and let λ_N^0 be the largest λ attaining Λ_N^0. In Appendix B, we show that $\Lambda_N^*=\Lambda_N^0$ and $\lambda_N^*=\lambda_N^0$. Now, $\rho(\lambda,(\infty,\infty))=2+$

TABLE I
COEFFICIENT λ_N^* AND RELATED QUANTITIES

N	λ_N^*	$\sqrt{2\log(N\log N)}$	Λ_N^*	$1+\log(N\log N)$
64	1.763	3.342	2.514	6.584
128	1.973	3.586	2.924	7.431
256	2.176	3.810	3.355	8.258
512	2.371	4.017	3.804	9.069
1024	2.560	4.211	4.271	9.868
2048	2.741	4.395	4.754	10.656
4096	2.917	4.569	5.252	11.436
8192	3.086	4.735	5.762	12.209
16384	3.251	4.894	6.285	12.977
32768	3.411	5.048	6.817	13.739
65536	3.566	5.195	7.360	14.496

λ^2 is strictly increasing in λ, and $\rho(\lambda,(0,0)) = 2\sqrt{2\pi}(\phi(\lambda) - \lambda\Phi(-\lambda))$ is strictly decreasing in λ so that at the solution of (11)

$$(N+1)\rho(\lambda_N^0,(0,0)) = \rho(\lambda_N^0,(\infty,\infty)).$$

Additionally, this last equation defines λ_N^0 uniquely.

Table I lists the universal λ_N and minimax λ_N^* thresholds and the corresponding bounds Λ_N and Λ_N^* for $N = 2^J$ with J between 6 and 16. The table is analogous to [2, Tab. II].

III. APPLICATION

In the previous section, we have derived the universal and minimax thresholds for denoising complex signals with Waveshrink. We now use the procedure for an electronic surveillance application. In this electronic surveillance application, we are interested in the problem of passive detection and fingerprinting of incoming radar signatures from an electronic surveillance platform. The observed complex signal plotted in Fig. 1 is $N = 2048$ samples from a chirped RF source at 3 dB with jamming interference.

We propose to model the underlying complex signal as a linear combination of wavelets. It is natural for this application to use chirplets [7], which are a collection of locally supported basis functions whose frequency changes linearly with time. The collection is, however, "over-complete" in the sense that its cardinality P is larger than the number of observations N. The matrix Φ has more columns than rows and is therefore not orthonormal. Chen *et al.* [9] proposed an extension of soft-Waveshrink, called basis pursuit, to estimate the coefficients in the over-complete situation. The coefficient estimate $\underline{\hat{\mu}}_\lambda$ is defined as the solution to the basis pursuit optimization problem

$$\min_{\underline{\mu}} \ \frac{1}{2}\,\|\underline{s} - \Phi\underline{\mu}\|_2^2 + \lambda\|\underline{\mu}\|_1 \qquad (12)$$

where $\|\underline{\mu}\|_1 = \sum_{p=1}^{P} |\mu_p|$. This definition of the coefficients is a generalization of Waveshrink because, interestingly, the soft-Waveshrink estimate $\underline{\hat{\mu}}_\lambda = \eta_\lambda^{(soft)}(\Phi'\underline{s})$ is the closed-form solution to the basis pursuit optimization problem when Φ is orthonormal. The nontrivial basis pursuit optimization problem (when Φ is over-complete) can be solved by an interior point algorithm [9] in the real case or by a block coordinate relaxation algorithm [10] in the real and complex cases. The latter is guaranteed to converge and has been found to be empirically more efficient.

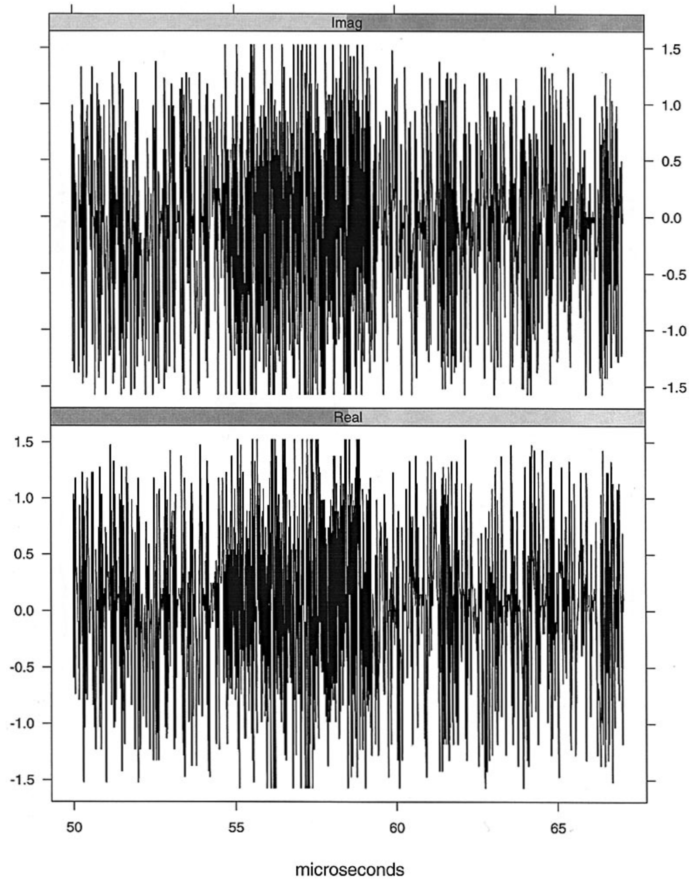

Fig. 1. Real (bottom) and imaginary (top) parts of a chirped RF source at 3 dB with jamming interference.

As for Waveshrink, the estimate of the meta parameter in (12) is a crucial point of the basis pursuit procedure. In practice, the threshold (universal or minimax) developed for Waveshrink gives a good denoising performance to basis pursuit. Because our application is concerned with feature extraction, we propose to use the universal threshold. By property (10), most interferences will be erased and the main features of the underlying signal will be revealed. Using the estimate of the standard deviation of the noise proposed by Donoho and Johnstone in [1] for Gaussian noise, we find that $\hat{\sigma} = .51$. Therefore, our selection of the meta parameter is $\lambda = \hat{\sigma}\sqrt{2\log(N\log N)} = 2.22$.

After solving the basis pursuit optimization problem, we obtain the denoised estimate $\underline{\hat{f}} = \Phi\underline{\hat{\mu}}_\lambda$. Fig. 2 shows the spectrogram of the original signal on top and the spectrogram of the denoised signal at the bottom. Although a chirp with linearly decreasing frequency is already slightly visible in the original signal, it is interesting to see how basis pursuit with the universal selection of the meta parameter has "cleaned-up" the signal from most of the jamming interference. For even better results, one can choose the meta parameter by hand. For instance Fig. 3 shows the basis pursuit denoised signal with $\lambda = 4$. An automatic procedure will more easily detect the fingerprint of an incoming radar signature on the denoised signal than on the original one.

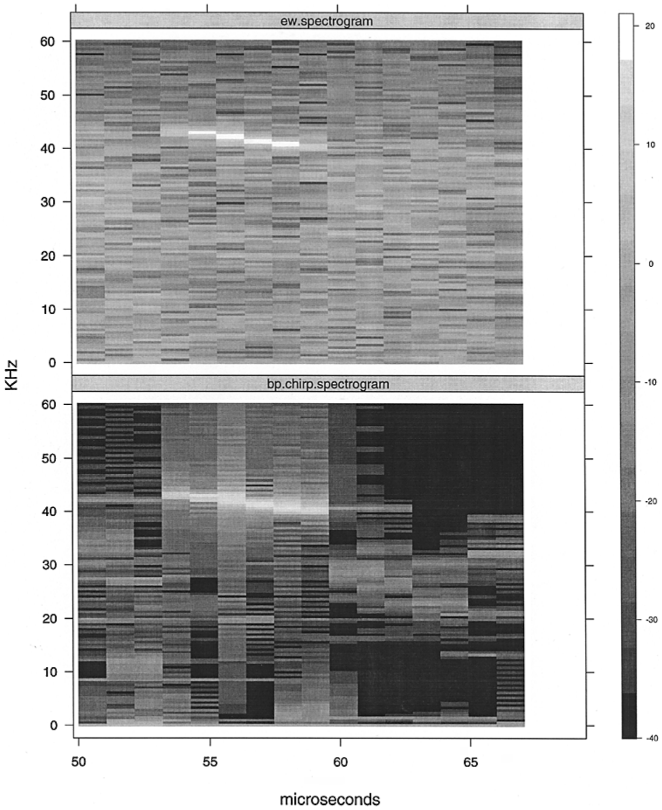

Fig. 2. Spectrogram for a chirped RF source at 3 dB with jamming interference. The top spectrogram is of the original signal, and the bottom spectrogram is of the denoised signal using the universal meta parameter $(\lambda = 2.22)$.

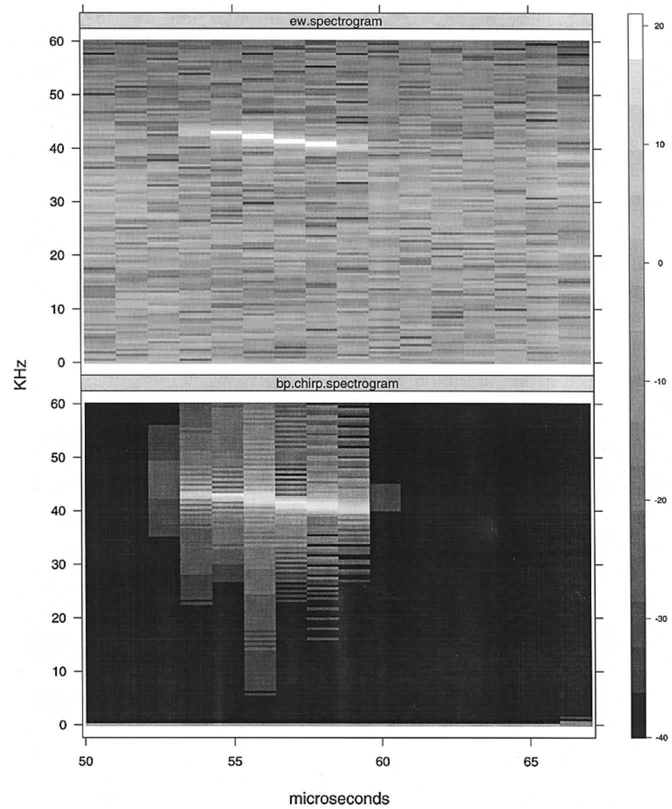

Fig. 3. Spectrogram for a chirped RF source at 3 dB with jamming interference. The top spectrogram is of the original signal, and the bottom spectrogram is of the denoised signal for a meta parameter chosen by hand $(\lambda = 4)$ to remove most of the interference.

IV. Conclusions

We have generalized Waveshrink and basis pursuit to complex signals. Both procedures need a selection of the meta parameter that controls the amount of denoising. For using the two procedures on complex signals, we have derived the universal and minimax thresholds selection of λ in a similar fashion as previously proposed for real signals by Donoho and Johnstone [2]. We then used the basis pursuit procedure with the proposed universal threshold to successfully denoise a complex signal in an electronic surveillance application.

Appendix A
Risk Properties of Soft Thresholding

The risk of the soft-thresholding estimator can be rewritten as

$$\begin{aligned}\rho(\lambda,\mu) &= E\left[\left(\frac{(|Y|-\lambda)_+}{|Y|}\,Y_1-\mu_1\right)^2 + \left(\frac{(|Y|-\lambda)_+}{|Y|}\,Y_2-\mu_2\right)^2\right] \\ &= 2+\lambda^2 \qquad (13)\\ &\quad -E\left[(\lambda^2-|Y|^2)1(|Y|<\lambda)\right]-4P(|Y|<\lambda)\\ &\quad -2\lambda\int_0^{2\pi}\int_\lambda^\infty \phi_1(r\cos\theta)\phi_2(r\sin\theta)\,dr\,d\theta \qquad (14)\end{aligned}$$

where $\phi_i(x)=\phi(x-\mu_i)$ is the Gaussian density function with mean μ_i.

We deduce the following properties of the risk:

P1) $\rho(\lambda,\mu) \le \rho(\lambda,(0,0)) + |\mu|^2$.

Proof: From (13), we easily obtain

$$\rho(\lambda,\mu) = \rho(\lambda,(0,0)) + |\mu|^2 - 2E\left[\frac{(|Y|-\lambda)_+}{|Y|}\,(\mu_1Y_1+\mu_2Y_2)\right]$$

and the third term $q = E[((|Y|-\lambda)_+/|Y|)(\mu_1Y_1+\mu_2Y_2)]$ is positive. (To see this last point, note that we can restrict out attention to $\mu_1 > 0$ and $\mu_2 > 0$. Then, define the line L_1: $\mu_1y_1+\mu_2y_2=0$, and define the line L_2 as the line parallel to L_1 that passes through (μ_1,μ_2); let P_1 be the half-plane below L_1, P_3 be the half-plane above L_2, and P_2 the remaining band. We can easily show that $q=q_1+q_2+q_3$ is positive since $q_2 \ge 0$ and $q_3 \ge -q_1 \ge 0$.)

P2) $\rho(\lambda,(\mu_1,\mu_2))$ reaches its maximum for a fixed λ at $(\mu_1,\mu_2)=(\infty,\infty)$, and $\rho(\lambda,(\infty,\infty)) = 2+\lambda^2$ is strictly increasing in λ.

Proof: The last three terms of the right side of (14) are negative. Therefore, the proof is done if we can show that their limits are zero when $\mu_1 \to \infty$ and $\mu_2 \to \infty$. It is straightforward for the first two terms whose integrals are defined on the compact circle of radius λ. The third term is also negative, but showing that its limit is zero when $\mu_1 = \mu_2 = \mu \to \infty$ demands more work. After some algebra, we can show that for $\mu_1 = \mu_2 = \mu$

$$\begin{aligned} q(\mu) &= \int_0^{2\pi} \int_\lambda^\infty \phi_1(r\cos\theta)\phi_2(r\sin\theta)\, dr\, d\theta \\ &= \frac{1}{\sqrt{2\pi}} \exp(-\mu^2) \int_0^{2\pi} \exp\left(\frac{1}{2}\left(\mu^2 + \mu^2 \sin(2\theta)\right)\right) \\ &\quad \cdot \left(\frac{1}{\sqrt{2\pi}} \int_\lambda^\infty \exp\left(-\frac{1}{2}(r - \mu(\cos\theta + \sin\theta))^2\right) dr\right) d\theta \\ &\le \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}\mu^2\right) \int_0^\pi \exp\left(\frac{\mu^2}{2}\cos\theta\right) d\theta(1) \\ &= \sqrt{2\pi} \exp\left(-\frac{1}{2}\mu^2\right) I_0\left(\frac{\mu^2}{2}\right) \end{aligned}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind with zero order. In [11], an asymptotic expansion of this function is given by $I_0(x) = (\exp(x)/\sqrt{2\pi x})(1 + (1/8x) + (1^2 \cdot 3^2/2!(8x)^2) + \cdots)$ so that

$$0 \le q(\mu) \le \frac{\sqrt{2}}{\mu} \left(1 + \frac{1}{4\mu^2} + \frac{1^2 \cdot 3^2}{2!(4\mu^2)^2} + \cdots\right).$$

Therefore, $\lim_{\mu\to\infty} q(\mu) = 0$. Hence, the maximum of $\rho(\lambda, (\cdot,\cdot))$ occurs at (∞,∞), and $\rho(\lambda, (\infty,\infty)) = 2 + \lambda^2$.

- $\rho(\lambda, (0,0)) = 2\sqrt{2\pi}(\phi(\lambda) - \lambda\Phi(-\lambda))$ is strictly decreasing in λ.

From these two points, we get that $p(N,\lambda) = (N+1)\rho(\lambda,(0,0)) - \rho(\lambda,(\infty,\infty))$ is decreasing in λ. Moreover, $p(N,0) > 0$ and $p(N,\infty) = -\infty$ so that the root λ_N^* to $p(N,\lambda) = 0$ is unique.

P3) $\lambda_N^0 > \sqrt{2/N}$ for $N \ge 4$.

Proof: $p(N,\lambda)$ is decreasing in λ. Moreover, $p(N, \sqrt{2/N})$ is increasing in N and positive for $N = 4$.Therefore, the root λ_N^0 of the equation is larger than $\sqrt{2/N}$.

Appendix B
Proof of Equivalence of $\Lambda_N^* = \Lambda_N^0, \lambda_N^* = \lambda_N^0$

It is clear that $\Lambda_N^*(\lambda) \ge \Lambda_N^0(\lambda)$ for all λ. Therefore, if we can establish that $\Lambda_N^*(\lambda_N^0) = \Lambda_N^0(\lambda_N^0)$, then λ_N^0 will also minimize Λ_N^*. We must prove that

$$L_N^*(\mu; \lambda_N^0) = \frac{\rho(\lambda_N^0, \mu)}{2/N + \min(|\mu|^2, 2)}$$

attains its maximum at either (0,0) or (∞,∞). We split the problem into two cases.

- For $|\mu|^2 > 2$, $L_N^*(\mu;\lambda_N^0) = (\rho(\lambda_N^0,\mu)/2/N)$ and the numerator [see (14)] attains its maximum at (∞,∞).
- For $|\mu|^2 < 2$, $L_N^*(\mu;\lambda_N^0) = (\rho(\lambda_N^0,\mu)/2/N + |\mu|^2) \le (\rho(\lambda_N^0,(0,0)) + |\mu|^2/2/N + |\mu|^2)$ by Property P1.

Looking now at the right-hand side of the inequality as a function of $r = |\mu|$, the sign of its derivative is the sign of $2 - N\rho(\lambda_N^0,(0,0))$, which in turn is negative because $N\rho(\lambda_N^0,(0,0)) = (2 + (\lambda_N^0)^2/1 + N^{-1}) \ge 2$ for $N \ge 4$ by Property P3. Therefore, $L_N^*(\mu;\lambda_N^0)$ reaches its maximum at the endpoint $\mu = (0,0)$.

Acknowledgment

The author would like to thank the three anonymous referees, who contributed numerous improvements to the manuscript.

References

[1] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[2] ——, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[3] H.-Y. Gao, "Wavelet shrinkage deoising using the nonnegative garrote," *J. Comput. Graph. Stat.*, vol. 4, pp. 469–488, 1998.

[4] H.-Y. Gao and A. Bruce, "WaveShrink with firm shrinkage," Stat. Sci. Div., MathSoft, Inc., Res. Rep. 39, 1996.

[5] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: Wiley, 1999.

[6] J.-M. Lina and M. Mayrand, "Complex daubechies wavelets," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 219–229, 1995.

[7] S. Mann and S. Haykin, "Adaptive "chirplet" transform: And adaptive generalization of the wavelet transform," *Opt. Eng.*, vol. 31, no. 6, pp. 1243–1256, June 1992.

[8] F. G. Meyer and R. R. Coifman, "Biorthogonal brushlet bases for directional image compression," in *Proc. 12th Int. Conf. Anal. Optimiz. Syst.*, 1996.

[9] S. Cheng, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," Stanford Univ., Stanford, CA, Tech. Rep., 1996.

[10] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *J. Comput. Graph. Stat.*, to be published.

[11] F. Bowman, *Introduction to Bessel Functions*. New York: Dover, 1958.

Sylvain Sardy received the Maitrise d'Ingénièrie mathématique degree from the University of Franche-Comté, Besancon, France, in 1989. He then received the M.S. degree in mathematics in 1991 and the M.S. degree in statistics in 1992 from Utah State University, Logan. After doing his military service for France at the Nuclear Engineering Department of the Massachusetts Institute of Technology, Cambridge, he received the Ph.D. degree in statistics from the University of Washington, Seattle.

He is now a Postdoctoral Assistant with the Swiss Federal Institute of Technology, Lausanne (EPFL), in the Mathematics Department (DMA).