accuracy was 13 bits, with the result of each multiplication truncated to 16 bits. The analog-to-digital and digital-to-analog conversions on the system have 12-bit accuracy.

However, the state-variable cascade realization behaves considerably better for all tested approximations compared with the canonical cascade realization when a suitable scaling has been applied. Clearly, the state-variable 16-bit realization with the fixed-point two's-complement saturation arithmetic (e.g., on the DSP chip TMS320C25) offers a possibility to increase the signal-to-round-off-noise ratio at the output of the cascade realization of IIR DF's in comparison with the tested canonical cascade realization.

REFERENCES

[1] B. Pšenička and J. Zadák, "Digital filters matricial analysis," *Acta Polytechnica*, vol. III, no. 1, pp. 5–15, 1991.

[2] B. Pšenička, J. Zadák, and V. Davídek, "Design of state digital filters," *Acta Polytechnica*, vol. III, no. 3, pp. 5–12, 1991.

[3] L. M. Smith, B. W. Bomar, and R. A. Joseph, "Floating-point roundoff noise analysis of second-order state space digital filter structures," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 90–98, Feb. 1992.

[4] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed point digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 551–561, Sept. 1976.

[5] W. L. Mills, C. T. Mullis, and R. A. Roberts, "Low roundoff noise and normal realizations of fixed point IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 893–903, Aug. 1981.

[6] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[7] R. W. Schafer and L. R. Rabiner, "Digital representation of speech signal," *Proc. IEEE*, vol. 63, pp. 662–677, June 1975.

[8] L. Thiele, "On the sensitivity of linear state-space systems," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 502–510, May 1986.

[9] L. B. Jackson, A. G. Lindgren, and Y. Kim, "Optimal synthesis of second-order state-space structures for digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 149–153, Mar. 1979.

[10] C. W. Barnes, "On the design of optimal state-space realizations of second-order digital filters," *IEEE Trans. IEEE Trans. Circuits Syst.*, vol. CAS-31, pp. 602–608, July 1984.

[11] B. W. Bomar, "New second-order state-space structures for realizing low roundoff noise digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 106–110, Feb. 1985.

[12] ——, "Computationally efficient low roundoff noise second-order state-space structures," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 35–41, Jan. 1986.

[13] ——, "On the design of second-order state-space digital filter sections," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 542–552, Apr. 1989.

[14] B. W. Bomar and R. D Joseph, "Calculation of L_∞ norms for scaling second-order state-space digital filter sections," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 983–984, Aug. 1987.

[15] T. W. Parks and C. Burrus, *Digital Filter Design*. New York: Wiley, 1987.

[16] A. W. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1989.

[17] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976.

[18] L. B. Jackson, A. G. Lingren, and Y. Kim, "Optimal synthesis of second order state-space structures for digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 149–153, Mar. 1979.

[19] C. Tsai and A. T. Famm, "Roundoff noise reduction in IIR digital filters via parallel decomposition," *IEEE Trans. Circuits Syst.*, vol. 38, p. 961, Aug. 1991.

On Time-Dependent Wavelet Denoising

Brani Vidakovic and Concha Bielza Lozoya

***Abstract*—In this correspondence, we address the shrinkage of wavelet coefficients and induced denoising in the time domain by taking into consideration the "time" behavior of a noisy signal. We illustrate our time adaptation paradigm in a thresholding procedure utilizing Bayesian hypothesis tests. Both one-dimensional (1-D) and two-dimensional (2-D) signals are considered in examples to motivate and implement our method.**

***Index Terms*—Denoising, image processing, wavelet shrinkage.**

I. INTRODUCTION

Wavelet shrinkage is a simple yet powerful tool in statistical modeling of data sets and signals. It can be described as a three-step procedure.

Step 1) A discrete signal is transformed into a set of wavelet coefficients.

Step 2) A shrinkage of the coefficients is performed.

Step 3) The shrunken wavelet coefficients are transformed back to the domain of the original signal.

Wavelet shrinkage is usually done by thresholding wavelet coefficients. Thresholding is a rule by which the coefficients with absolute values smaller than a fixed threshold are replaced by zeroes. Some other thresholding policies and choices of thresholds are reviewed in [12].

Many researchers have observed that scale-dependent denoising methods in the wavelet domain can improve visual and other characteristics of signals and images. Our aim is to describe and apply one possible approach to the adaptivity of shrinkage rules. "Time" dependent shrinkage has the threshold that depends on the relative position of the coefficient within the scale level of the decomposition. This shrinkage rule is fully determined via an empirical Bayes argument in testing precise hypotheses in the wavelet domain.

We emphasize that shrinkage and denoising are two related but not identical actions. The former is performed in the wavelet domain with wavelet coefficients as arguments. The latter is a consequence in the time domain of the shrinkage action. The two are connected via Meyer's result that states that the magnitudes of the wavelet coefficients determine the smoothness space to which the signal under consideration belongs. For a formal statement of this connection, refer to [7], [8], and [10].

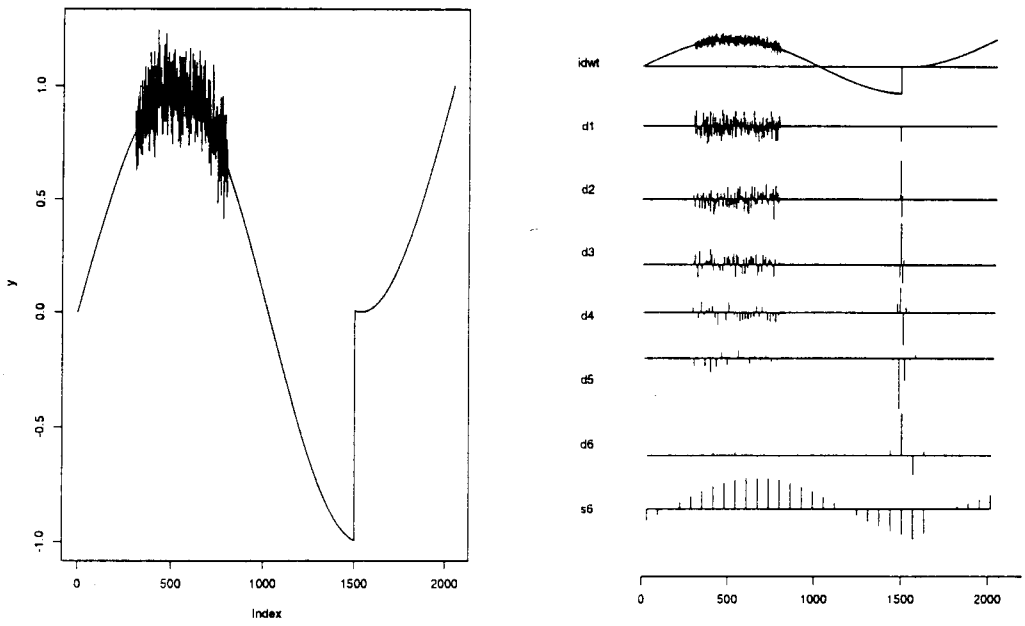
A large body of research in the statistical and engineering communities focused recently on performing shrinkage/denoising by modeling in the wavelet domain. The models are supported by the data to produce statistically optimal shrinking methods. An early reference is the work of Mallat [9], who proposed the exponential power distribution as a realistic statistical model for wavelet coefficients.

Manuscript received January 27, 1997; revised March 5, 1998. This work was supported by NSF Grant DMS-9626159. The associate editor coordinating the review of this paper and approving it for publication was Dr. Xiang-Gen Xia.

B. Vidakovic is with the Institute of Statistics and Decision Theory, Duke University, Durham, NC 27708 USA (e-mail: brani@stat.duke.edu).

C. Bielza Lozoya is with the Universidad Politecnica de Madrid, Madrid, Spain (e-mail: mcbielza@fi.upm.es).

Publisher Item Identifier S 1053-587X(98)05963-7.



(a) (b)

Fig. 1. (a) Artificial example containing a point of discontinuity and noise. (b) Discrete wavelet transformation emphasizing the difference between wavelet coefficients attributed to noise and those attributed to the discontinuity.

Why is the wavelet domain a good environment for statistical modeling? There are several reasons, including the following.

1) Low Entropy Modeling Environment: Discrete wavelet transformations tend to "unbalance" the signal. Even though the transformations preserve the ℓ_2 norm (the energy) of the signal, the energy of the transformed signal concentrates in only a few wavelet coefficients. That narrows the class of plausible models and facilitates the efficiency of thresholding. Mallat [9] gives an interesting discussion from the signal processing point of view. The unbalancing property also motivates a plethora of criteria for the best basis selection. Two standard references are [5] and [6].

2) Ockham Razor Principle: Wavelets, as building blocks of models, are well localized in both time and scale (frequency). The signals with rapid local changes (discontinuities, cusps, sharp spikes, etc.) can be well represented with just a few wavelet coefficients. Generally, this rule does not apply to other standard orthonormal bases, which may require many "compensating" coefficients to describe discontinuity artifacts or to suppress Gibbs' effects.

3) Bypassing the Curse of Heisenberg: Heisenberg's principle states that in modeling time–frequency phenomena, we cannot be precise in the time domain and in the frequency domain simultaneously. Wavelets handle the time–frequency precision automatically by their innate nature. The parsimony mentioned above can be ascribed to the ability of wavelets to cope with limitations of Heisenberg's principle in a signal-dependent manner.

4) Whitening Property: There is ample theoretical and empirical evidence that wavelet transformations tend to simplify the dependence structure in the original signal. It is even possible to construct a biorthogonal wavelet basis that decorrelates a given stationary signal at input (a wavelet counterpart of the Karhunen–Loéve transformation). For a discussion and examples, see [16].

The above arguments evince wavelet bases to be suitable tools for effective statistical modeling. Several more reasons can be provided: calculational efficiency, self-similarity in multiresolution, etc.

II. Adaptive Denoising

The first proposed methods in wavelet denoising had a universal character: A single threshold policy was applied to all wavelet coefficients. Donoho and Johnstone proposed the first adaptive denoising method, called SureShrink [7], in which wavelet coefficients at different levels are treated differently, depending on the level index and total energy in the level. Good references are [2] and [12].

From the Bayesian point of view, the level-dependent denoising results naturally from specification of a statistical model in the wavelet domain. Some recent references are [3], [4], and [11]. We refer to such adaptivity in a denoising algorithm as the "scale" adaptivity.

For time-inhomogeneous signals, it is useful to inspect the behavior of wavelet coefficients in a fixed level for different shift parameters. If the noisy signal has a singularity, the high-energy propagates locally across the coefficients to the coarsest level of detail. For a single singularity, the energetic coefficients are contained in the so called *cone of influence.*

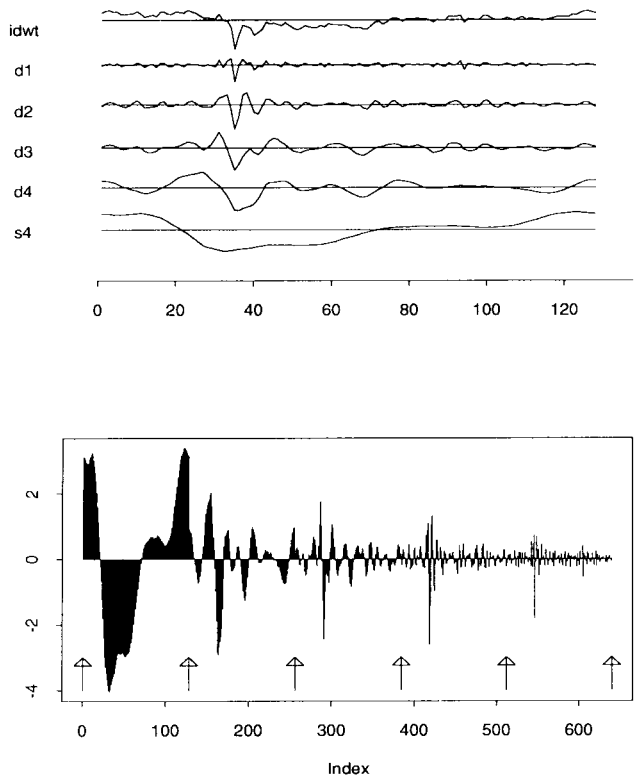
The noise is well explained by a few levels that contain fine detail coefficients (and its effect disappears at the coarser scales).

Fig. 1(a) shows an artificial but motivating example. A signal with a single discontinuity was sampled at $n = 2048$ equally spaced time points. The values between 301 and 800 have an added normal $\mathcal{N}(0, 0.1^2)$ noise.

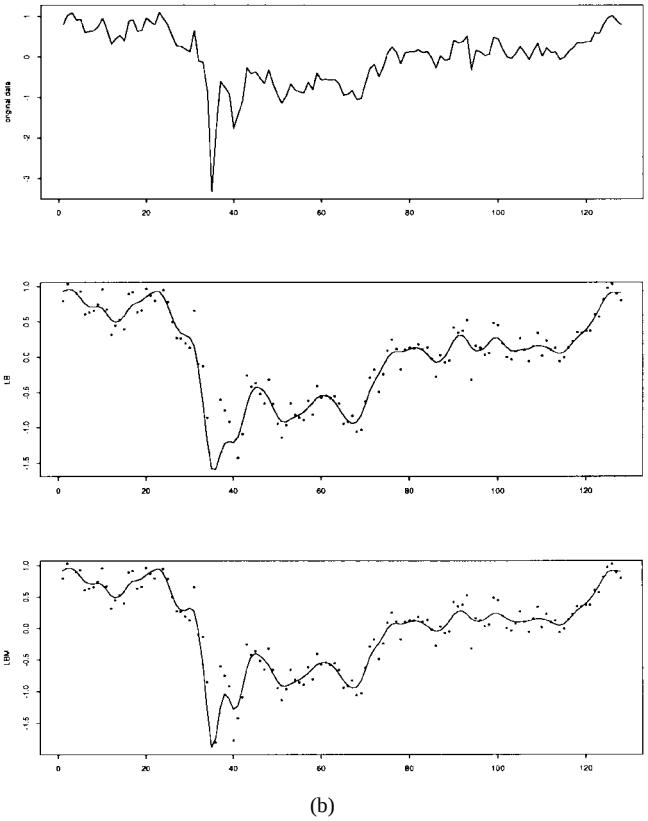
In the wavelet decomposition of the signal [see Fig. 1(b)], we notice that the levels **d1** and **d2** contain most of the energy attributed to the noise and that the noise is practically nonexistent in the detail level **d6**. However, the effect of the discontinuity artifact does not abate with coarsening of detail spaces. In this example, the least asymmetric Daubechies' wavelet with four vanishing moments (the default filter S8 in S + Wavelets) was used.

Generally, in coarser levels, the disturbance caused by an artifact of interest (sharp peak, discontinuity) dominates disturbances generated by the noise. We propose time-dependent thresholding that "sacrifices" a small proportion of the wavelet coefficients from the coarsest level of detail in order to better calibrate the underlying statistical model.

Fig. 2(a) shows time-dependent thresholding in action. The signal consists of 128 observations of Coke stock market prices. Around index 35, we can see the artifact caused by the (in)famous stock market crash in 1987. The goal is to denoise the signal while



(a)



(b)

Fig. 2. (a) Coke stock market prices signal and its nondecimated wavelet transformation. (b) Original signal, signal denoised in a traditional way, and signal denoised in a time dependent manner.

not oversmoothing the crash artifact. The upper panel of Fig. 2(a) depicts the signal (in `idwt` level) and its nondecimated wavelet

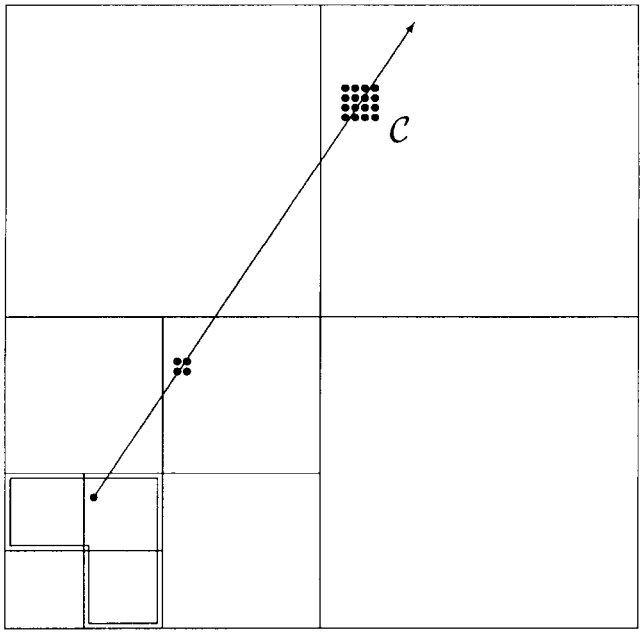


Fig. 3. Shapiro's zero-tree.

decomposition (see, for example, [2]). The bottom of the left panel contains all the coefficients $(5 \cdot 128 = 640)$ in one line. The coefficients from level **d4** have been used in performing time-dependent thresholding.

The top portion of Fig. 2(b) shows the original signal and the signal denoised by standard, level dependent, thresholding techniques (middle portion). The bottom part of the Fig. 2(b) contains "time-adaptive" denoised signal. Notice two desirable properties: In the region of no activity (indices > 50), time-dependent thresholding smooths more, whereas in the artifact area (indices 20–50), the time-dependent thresholding rule thresholds less, preserving more information and making the denoised signal more precise.

III. Formalizing "Time" Dependent Thresholding

Bayes rules with respect to squared error loss in an estimation contest are never thresholding-type rules—they are smooth. One way to obtain thresholding rules in a Bayesian framework is through statistical testing. Testing a precise hypothesis in Bayesian fashion requires a prior that has a point mass component. Otherwise, the testing is trivial since any continuous prior density will give the prior (and, hence, the posterior) probability of 0 to a precise hypothesis. For discussion on Bayesian testing of precise hypotheses, see [1].

Denote by d an observed wavelet coefficient. Assume that the distribution of d depends on a location parameter θ that represents the unknown signal of interest. In addition, assume that the d's are independent. Although the independence assumption may be an oversimplification of our model, it is justified by the whitening property of wavelet transformations (see the Introduction) and by the robustness of its resulting denoising procedures.

The model we discuss below can be found in much greater detail in [15]; however, the setup discussed there is nonadaptive. We use it here to illustrate a principle of how related denoising methods can be made time-scale dependent.

Let the wavelet coefficient d conditional on unknown location θ have a density $f(d-\theta)$. After observing d, the hypothesis H_0: $\theta = 0$ versus H_1: $\theta \neq 0$ is tested. If the hypothesis H_0 is rejected, θ is estimated by d. Let $\theta \sim \pi(\theta) = \pi_0(\mathcal{C}_j)\delta_0 + \pi_1(\mathcal{C}_j)\xi(\theta)$ be the prior

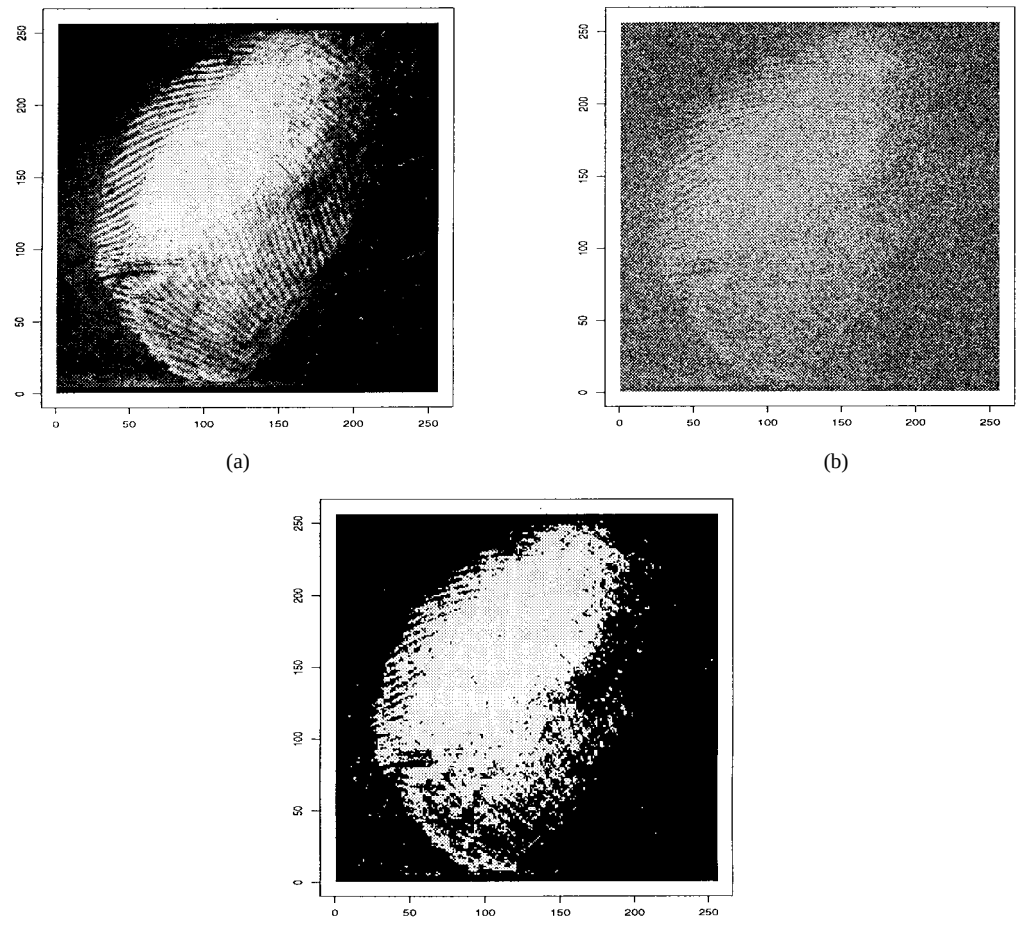


Fig. 4. (a) Original fingerprint image. (b) Fingerprint image with the normal $\mathcal{N}(0,\ 100^2)$ noise. (c) Segmented version of the denoised image.

on θ, where $\pi_0(\mathcal{C}_j)$ and $\pi_1(\mathcal{C}_j)$ are prior probabilities of hypotheses H_0 and H_1, δ_0 is a point mass at 0, and $\xi(\theta)$ is a part of the prior that describes the distribution of θ when H_0 is false. Adaptivity of the above procedure is introduced by the dependence of prior probabilities π_0 and π_1 on the energies in the coarsest level of details reflected on a clique of coefficients $\mathcal{C}_j$ from the scale j. One example of how $\mathcal{C}_j$'s are formed is described in Section IV.

In a compact form, the Bayes procedure for θ corresponding to $\mathcal{C}_j$ is

$$\hat{\theta} = d\,\mathbf{1}\left(P(H_0|\mathcal{C}_j,\ d) < \tfrac{1}{2}\right) \tag{1}$$

where $P(H_0|\mathcal{C}_j,\ d) = (1 + \pi_1(\mathcal{C}_j))/(B\pi_0(\mathcal{C}_j))^{-1}$ is the posterior probability of the H_0 hypothesis, $B = f(d)\Big/\int_{\theta\neq 0} f(d-\theta)\xi(\theta)\,d\theta$ is the Bayes factor in favor of H_0, and $\mathbf{1}(R)$ is an indicator of relation R; it is 1 if R is true and 0 if it is false. It is apparent that (1) is the shrinkage rule of hard thresholding type.

Next we specify distributions f and ξ and give an effective rule for the thresholding described in (1). The likelihood for d is a double exponential, and the continuous part of the prior on θ is the Student t.

The choice of the double exponential for the distribution wavelet coefficients can bejustified by two different arguments:

1) The First Argument is Mallat's Empirical Argument: Mallat [9] noticed that for a variety of signals and images, empirical distributions of wavelet coefficients appeared to be similar. Typically, the distributions were symmetric and had a sharp peak at zero. He proposed that a "random wavelet coefficient" D could be modeled by a distribution from the family of exponential power distributions ($\mathcal{EPD}(\alpha,\ \beta)$)

$$f(d) = C\cdot e^{-(|d|/\alpha)^{\beta}},\qquad \alpha,\ \beta > 0. \tag{2}$$

The normalizing constant C in (2) is given by $C = \beta/(2\alpha\Gamma(1/\beta))$. Because $E|D| = \alpha\Gamma(2/\beta)/\Gamma(1/\beta)$ and $ED^2 = \alpha^2\Gamma(3/\beta)/\Gamma(1/\beta)$, the parameters α and β can be estimated from the observations by the *method of moments* as $\hat{\beta} = G^{-1}(m_1^2/m_2)$ and $\hat{\alpha} = m_1\Gamma(1/\hat{\beta})/\Gamma(2/\hat{\beta})$, respectively. Here, $m_1 = 1/n\sum|d_i|$ and $m_2 = 1/n\sum d_i^2$ are empirical counterparts of $E|D|$ and ED^2, and $G(x) = \Gamma^2(2/x)/(\Gamma(1/x)\Gamma(3/x))$. Mallat's model (2) is used to perform optimal quantile wavelet thresholding (see [17] for details) and the Bayesian coring procedure by [14].

The range of estimators $\hat{\beta}$ is [0.2, 1.5] for the majority of signals encountered in practice, and the proposed double exponential model [obtained from (2) for $\beta = 1$] agrees with Mallat's empirical argument.

2) The Second Argument is Probabilistic: Conditional on an unknown parameter σ^2 (the variance of the noise) and the unknown signal θ, d is normal, i.e., $d|\theta,\ \sigma^2 \sim \mathcal{N}(\theta,\ \sigma^2)$. By integrating out σ^2 by an exponential prior,[1] the marginal likelihood becomes a double exponential. If $\sigma^2 \sim \mathcal{E}(\mu)$, where $1/\mu$ is the expectation,

[1]The exponential distribution minimizes the Fisher information in the class of all distributions supported on $\mathbb{R}^+$ with fixed first moment.

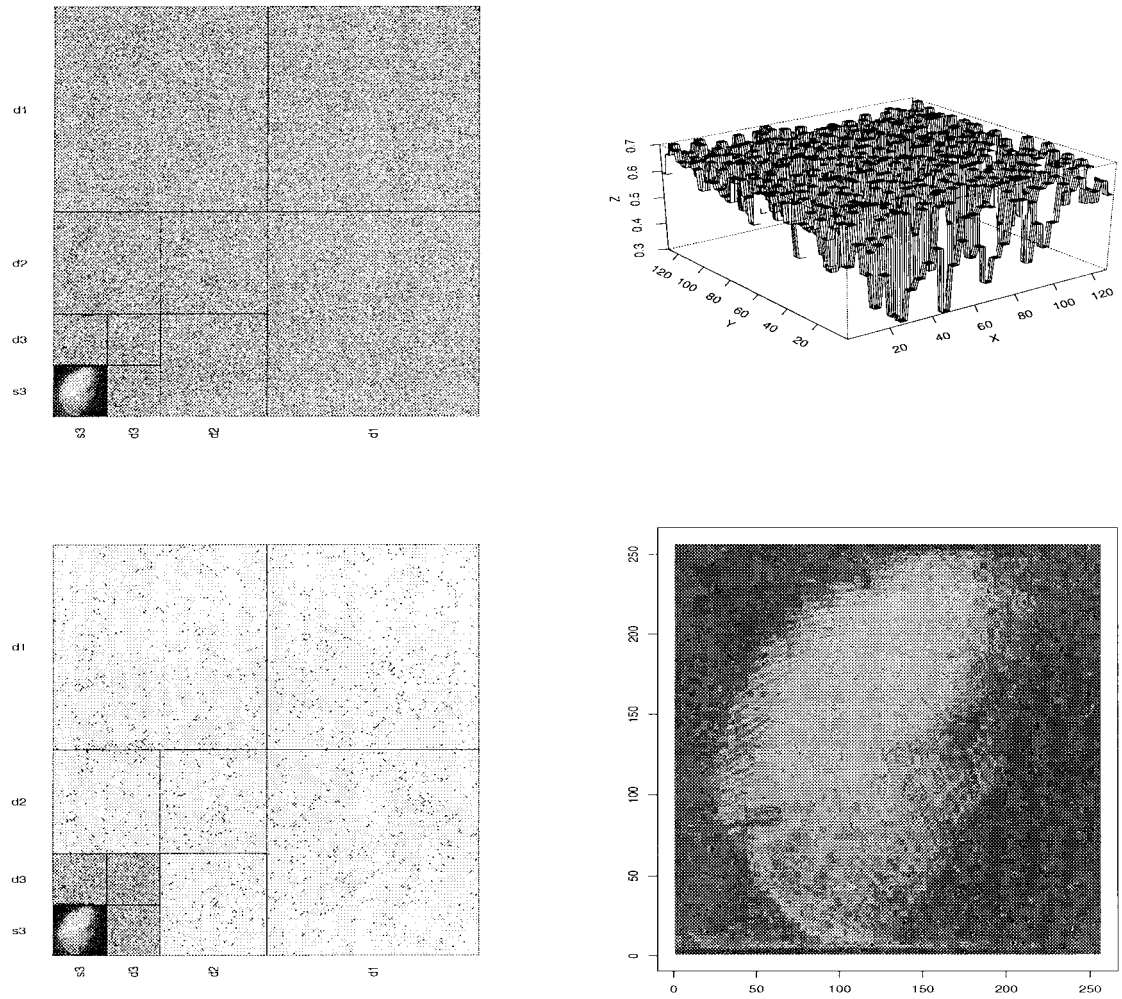


Fig. 5. (Upper left) DWT of the noised image. (Upper right) Prior probabilities generated by the intensities in the coarse matrix `s3–d3`. (Lower left) "Zero-tree" thresholded transformation. (Lower right) Denoised image with continuous pixel values.

then $d|\theta \sim \mathcal{DE}\big(\theta, 1/\sqrt{2\mu}\big)$, where $1/\sqrt{2\mu}$ is a scale parameter.[2] This calculation is a straightforward consequence of the fact that an exponential scale mixture of normals is a double exponential.

The following result gives the effective thresholding rule for general symmetric prior π.

Result 1 [15]: If an observed wavelet coefficient d in the jth level follows the model $d|\theta \sim \mathcal{DE}\big(\theta, 1/\sqrt{2\mu}\big)$ and the prior for the location (signal part) is $\pi(\theta) = \pi_0(\mathcal{C}_j)\delta_0 + \pi_1(\mathcal{C}_j)\xi(\theta)$, then d will be "thresholded" by

$$\delta(d) = d\,\mathbf{1}\left(\frac{\pi_0(\mathcal{C}_j)\,e^{-c|d|}}{\pi_0(\mathcal{C}_j)\,e^{-c|d|} + \pi_1(\mathcal{C}_j)(\Pi_1(c) + \Pi_2(c))} < \frac{1}{2}\right) \quad (3)$$

where $c = 1/\sqrt{2\mu}$, Π_1, and Π_2 are the one-sided Laplace transformations of shifted densities ξ, $\xi(\theta - d)$, and $\xi(\theta + d)$.

IV. An Application in Image Processing

The standard 2-D wavelet transformation applied to images organizes the details in a specific way. In the lower-left corner of Fig. 5, there is a "smooth" matrix, which is the result of projecting the image onto a smooth space V_{j_0} for some fixed level $j_0 \in \mathbb{Z}$. Around the "smooth" matrix, there are three matrices of the same dimension as the smooth one, corresponding to the coarsest level of detail. These three matrices contain the coefficients corresponding to the three detail spaces $W_{j_0}^1$, $W_{j_0}^2$, and $W_{j_0}^3$. The next (finer) level of detail contains three matrices surrounding the matrices grouped in the lower-left corner, and so on. The three matrices containing the finest level of detail contain 75% of all coefficients in the decomposition. The dimensions of the wavelet-transformed image are the same as the dimensions of the original image.

For natural images with decaying spectrums, it is unlikely to find an extensive amount of energy among the finest details if there is little or no energy at the same spatial location in the matrices of coarser details. This observation on relative self-similarity in detail matrices led Shapiro [13] to develop an efficient method of data compression called the *zero-tree method.* This methoid is illustrated in Fig. 3.

We propose the use of the zero-tree idea to determine the localized thresholds. The procedure scans the energies of the coefficients at some fixed coarse level of detail—usually the coarsest one. These coefficients, as in [13], are called *parents.* No shrinkage is performed on parents or on the coefficients from coarser levels, if such exist. The "smooth" level is also never thresholded. In the next finer level of detail, each parent has four *children,* and the subsequent parents will have 16, and so on. The terms parents, children, and, more generally, ancestors and descendants were proposed by Shapiro to emphasize the directions of energy influence. These directions are antithetical to the parent–children directions in wavelet packets or local cosine

[2]The density associated with $\mathcal{DE}\big(\theta, 1/\sqrt{2\mu}\big)$ is $f(d|\theta) = \frac{1}{2}\sqrt{2\mu}\,e^{-\sqrt{2\mu}|d-\theta|}$.

bases in tasks of selecting the best basis, for example.

By specifying the prior probabilities π_0 of the hypotheses H_0, we treat all descendants of a single parent $\mathcal{C}_j$ in the level j in the same way. If the parent possesses high energy, then the probability $\pi_0(\mathcal{C}_j)$ should be "small," indicating that there is a strong "local energy activity," and we should be more strict about accepting H_0 (i.e., setting $\theta = 0$). If the parents possess small or no energy, then their descendants are likely to be insignificant as well. In that case, $\pi_0(\mathcal{C}_j)$ should be large, indicating our prior opinion about the energies of the descendants in the level j. However, the inference depends on the posterior, and no-energy parents may have descendants that "survive" thresholding and vice versa.

For the sake of robustness, the distribution ξ part of the prior is chosen to be from the Student t family. In the resulting procedure, the choice of the threshold λ is driven by the imposed model or, more precisely, by the locally adaptable hyperparameters of the model.

The sensitivity issues and empirical Bayes methods in specifying the hyperparameters are not discussed here (see a related model in [15] and the discussion therein). We simply note that the procedure can be made automatic by an empirical Bayes argument: Take $\mu = \text{const}/\hat{\sigma}^2$ for some estimate of variance of d, $\hat{\sigma}^2$. We found that values of const between 0.5 and 4 work well in practice.

The prior probabilities $\pi_0(\mathcal{C}_j)$ and $\pi_1(\mathcal{C}_j)$ should be selected such that the proportion of d's from $\mathcal{C}_j$ for which $\pi_0(\mathcal{C}_j)$ is close to 1 increases when details become finer, thus contributing to the parsimony of the thresholded object.

We apply the described procedure on a real-life image. The fingerprint image (courtesy of Prof. L. Pastor) is one of the several fingerprint images with interesting criminal histories that are part of an ongoing project at Universidad Politecnica de Madrid. The image is a 256×256 matrix of pixels whose gray intensities range between 0 and 255 [Fig. 4(a)].

We note that the fingerprint image is used only to exemplify our procedure. We do not compare the proposed adaptive method with the state-of-the-art wavelet-based fingerprint compression methods developed for the FBI, for example. Before comparing methods, extensive research should be conducted on how to choose good priors, how to calibrate their hyperparameters, how to select the best decomposing basis, and how to optimally design a hybrid procedure that adds the time-adaptation (tilling).

In Fig. 4(b), we have the same image with added i.i.d. normal noise (SNR $= \sigma_{\text{image}}/\sigma_{\text{noise}} = 0.7$). The goal is to perform the time-dependent shrinkage and denoise this image.

The upper left part of Fig. 5 gives 2-D wavelet decomposition of the noised fingerprint image. There are three levels of detail. We employ the Daubechies **d8** wavelet. The third level of detail (the coarsest level) is utilized to generate the prior probabilities $\pi_0(\mathcal{C}_j)$ of the H_0 hypotheses for different levels j. The upper right part of Fig. 5 illustrates the prior probabilities constrained between $p_0 = 0.3$ and $p_1 = 0.7$. The bounds p_0 and p_1 are calibrated to control the parsimony of the thresholded wavelet decomposition. The lower left panel of Fig. 5 exhibits the thresholded version of the wavelet decomposition that used prior probabilities from the upper right panel. Finally, the lower right panel displays the denoised fingerprint image as an inverse wavelet transformation from the thresholded coefficients in the lower left panel.

V. Conclusions

In this correspondence, we considered "time" adaptation of an exemplary wavelet-based denoising procedure. This procedure is based on hard thresholding rules defined by Bayesian tests of precise hypotheses, which have been considered previously in [15] as nonadaptive. It is argued that the "time adaptation" is useful in preserving more information on discontinuities, spikes, and unusual energy activities not attributed to the noise. S+ Wavelets software was used in the implementation. The procedures are expeditious. The S+ programs can be obtained from the authors upon request.

There are several avenues for further research. Distributional and modeling issues of wavelet coefficients can be explored conditionally on coarse scales. In that case, Markov chain Monte Carlo methods should be used to perform the smoothing. "Time adaptation" of many existing procedures may be achieved by utilizing wavelet coefficients on a coarse scale via an empirical argument.

Acknowledgment

This work was initiated at MathSoft-StatSci, Seattle, WA, in the Summer of 1995 during a discussion between B. Vidakovic and A. Bruce. The comments of Associate Editor X.-G. Xia and three anonymous referees improved the exposition. Dr. L. Pastor from UPM, Boadilla del Monte, Spain, kindly provided the fingerprint image.

References

[1] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer, 1985.
[2] A. Bruce and H.-Y. Gao, *S+ Wavelets*. Seattle, WA: StatSci, 1995, user manual.
[3] H. Chipman, R. McCulloch, and E. Kolaczyk, "Adaptive Bayesian shrinkage," *J. Amer. Stat. Assoc.*, vol. 92, pp. 1413–1421, 1997.
[4] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, pp. 391–402, 1998.
[5] R. Coifman and V. Wickerhauser, "Entropy based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Apr. 1992.
[6] D. Donoho, "On minimum entropy segmentation," in *Wavelets: Theory, Algorithms, and Applications*, Chui, Montefusco, and Puccio, Eds. New York: Academic, 1994.
[7] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[8] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *J. R. Stat. Soc. B.*, vol. 57, pp. 301–369, 1995, with discussion.
[9] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
[10] Y. Meyer, *Wavelets and Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[11] P. Müller and B. Vidakovic, "Bayesian inference with wavelets: Density estimation," to be published.
[12] G. Nason, "Choice of the threshold parameter in wavelet function estimation," in *Wavelets in Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, Lecture Notes in Statistics, 1995.
[13] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
[14] E. Simoncelli and E. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. 3rd IEEE Int. Conf. Image Process.*, Lausanne, Switzerland, 1996.
[15] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *J. Amer. Stat. Assoc.*, vol. 93, pp. 173–179, 1998.
[16] G. G. Walter, *Wavelets and Others Orthogonal Systems with Applications*. Boca Raton, FL: CRC, 1994.
[17] F. Zeppenfeldt, J. Börger, and A. Koppes, "Optimal thresholding in wavelet image compressing," *Proc. SPIE*, vol. 2034, pp. 230–241, 1993.