Wavelet Thresholding for Multiple Noisy Image Copies

S. Grace Chang, Bin Yu, and Martin Vetterli

This work was supported in part by the NSF Graduate Fellowship and the University of California Dissertation Fellowship (Chang), in part by ARO Grant DAAH04-94-G-0232 and NSF Grant DMS-9322817 (Yu), and in part by NSF Grant MIP-93-213002 and Swiss NSF 20-52347.97 (Vetterli). S.G. Chang is formerly with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, and now with Hewlett-Packard Laboratories, Palo Alto, CA. B. Yu is with the Department of Statistics, University of California, Berkeley, and Bell Laboratories, Lucent Technologies, NJ. M. Vetterli is with the Laboratory of Audiovisual Communications, Swiss Federal Institute of Technology (EPFL), in Lausanne, Switzerland, and is affiliated with the Department of Electrical Engineering and Computer Sciences at U.C. Berkeley.

Part of this work was presented at the *IEEE International Conference on Image Processing (ICIP)*, Chicago, IL, October 1998.

Abstract

This work addresses the recovery of an image from its multiple noisy copies. The standard method is to compute the weighted average of these copies. Since the wavelet thresholding technique has been shown to effectively denoise a single noisy copy, we consider in this paper combining the two operations of averaging and thresholding. Because thresholding is a non-linear technique, averaging then thresholding or thresholding then averaging produce different estimators. By modeling the signal wavelet coefficients as Laplacian distributed and the noise as Gaussian, our investigation finds the optimal ordering to depend on the number of available copies and on the signal-to-noise ratio. We then propose thresholds that are nearly optimal under the assumed model for each ordering. With the optimal and near-optimal thresholds, the two methods yield similar performance, and both show considerable improvement over merely averaging.

I. Introduction

Denoising via wavelet thresholding proposed by Donoho and Johnstone [3] is a simple non-linear yet effective technique which outperforms linear techniques in theory and practice (cf. Chapter 10, [5]). Since this seminal work, there have been many extensions. Most of these works are for situations where there is only one set of observations (e.g. one time series sequence or one still image). However, in many applications there are multiple copies of the same or similar images, thus it is necessary to investigate denoising techniques which remove noise from multiple corrupted copies of the same signal. For a corrupted video sequence, suppose we choose a few consecutive frames in which the motion is not significant and that we have already taken care of the registration problem, one can view the frames as multiple noisy copies of the same image. Another example is when one scans a picture, but with unsatisfactory result, thus one does multiple scans, and then combines these copies to obtain the most noise-free copy possible. Since wavelet thresholding has worked well for one copy of corrupted image (cf. [3], [1], [5], [7]), we consider its extension to multiple copies in this paper.

The standard method for combining the multiple copies is to compute their weighted average. One can only do better by incorporating a thresholding step. The question is, which ordering is better, thresholding first or averaging first, and what is the threshold value for each ordering? The answer is not clear at all *a priori* because thresholding is a non-linear technique that reduces variance at the expense of increasing bias. We address these issues in this paper. With the coefficients of each subband modeled as samples of a Laplacian random variable and the noise as samples of a Gaussian variable, we will show that the optimal ordering (in the mean squared error sense) depends on the number of available copies and the proportion between the noise power and the signal power. Moreover, we propose near-optimal subband adaptive thresholds for both orderings. Results show that with the optimal or the proposed near-optimal thresholds, the two methods yield very similar performance, and both outperforms weighted averaging substantially.

II. Denoising Algorithm for Multiple Noisy Copies

Let $\boldsymbol{f} = \{f_{ij}\}$ denote the $M \times M$ matrix of the original image to be recovered. (Note that we use the boldfaced letters for the *matrix* of coefficients, and regular letters f_{ij} for *individual* pixels.) The signal $\boldsymbol{f}$ has been transmitted over a Gaussian additive noise channel N times, and at the receiver we have N copies of noisy observations, $\boldsymbol{g}^{(n)} = \boldsymbol{f} + \boldsymbol{\varepsilon}^{(n)}, n = 1, \ldots, N$. For the n-th copy, $\{\varepsilon_{ij}^{(n)}\}$ are *iid* Gaussian $N(0, \sigma_n^2)$, where σ_n^2 is the noise variance of the n-th copy. The noise samples between different copies are assumed independent. The goal is to find an estimator $\hat{\boldsymbol{f}}$ which minimizes the mean squared error (MSE), $MSE = \frac{1}{M^2} \sum_{i,j=1}^{M} (\hat{f}_{ij} - f_{ij})^2$.

The recovery of the image is done in the orthogonal wavelet transform domain (the readers are referred to standard wavelet literature such as [6], [8] for details of the 2D dyadic wavelet transform). Let the wavelet transform of the noisy observation $\boldsymbol{g}^{(n)} = \boldsymbol{f} + \boldsymbol{\varepsilon}^{(n)}$ be denoted by $\boldsymbol{Y}^{(n)} = \boldsymbol{X} + \boldsymbol{V}^{(n)}$. The wavelet coefficients are often grouped into *subbands* of different scale and orientation, with one lowest frequency subband, and the rest called *detail subbands*. It has been found that for a large class of images, the coefficients in each detail subband form a histogram well-described by a Generalized Gaussian distribution (cf. [9], [6]), often simplified to Laplacian distribution for tractability. In this work, we also use the Laplacian distribution. Then it follows that the MSE is well approximated by the *expected* squared error, or *risk* under the squared loss. Thus we wish to find the estimator $\hat{f}(g)$ of the coefficients which minimizes the risk, $Risk = E[(f - \hat{f})^2]$.

A. *Wavelet Thresholding and Threshold Selection*

To denoise one copy, the wavelet thresholding operation proposed by Donoho and Johnstone [3] has three steps. First, take the wavelet transform of the noisy observation $\boldsymbol{g}$ to yield $\boldsymbol{Y}$. Then each coefficient Y_{ij} (except in the lowest resolution subband) is thresholded with a chosen threshold. Finally, the thresholded coefficients are transformed back to yield the recovered signal.

There are two popular thresholding functions: the *soft-threshold* function, $\eta_\lambda(t) = \mathrm{sgn}(t) \cdot \max(0, |t| - \lambda)$, which shrinks the input towards zero by amount λ, and the *hard-threshold* function, $\psi_\lambda(t) = t \cdot \mathbf{1}\{|t| > \lambda\}$, which keeps the input only if it is above the threshold λ. Although the soft-thresholding operation tends to smooth the image slightly more than the hard-threshold function, it yields images with better visual quality especially when the noise power is significant. Furthermore, with the chosen probability distribution, optimal soft-thresholding yields a lower risk than optimal hard-thresholding, as was shown in [1]. Thus, soft-thresholding is the preferred operation in this work.

The next issue is the selection of the threshold value. The threshold selection is performed once for each subband. Thus, let us consider coefficients from one detail subband. As in [1], we model the wavelet coefficients in a detail subband of the original image $\boldsymbol{f}$ as samples from a zero-mean Laplacian random variable with an unknown parameter. That is, each X_{ij} is modeled as $X \sim p(x) = \mathrm{LAP}(\beta) \triangleq \dfrac{1}{\sqrt{2}\beta} \exp\left\{-\dfrac{\sqrt{2}}{\beta}|x|\right\}$.

Since the noise is Gaussian and the wavelet transform is orthogonal, each wavelet coefficient, Y_{ij}, of the corrupted image has distribution $Y|X \sim p(y|x) = N(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-x)^2}{2\sigma^2}\right\}$. We use the soft-threshold estimate, $\hat{X} = \eta_\lambda(Y)$, and the optimal threshold is defined to be

$$\lambda^* = \arg\min_\lambda E_X E_{Y|X}(\hat{X} - X)^2 = \arg\min_\lambda \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\eta_\lambda(y) - x)^2\, p(y|x)\, p(x)\, dy\, dx. \tag{1}$$

As far as we know, λ^* does not have a closed form solution and is thus found numerically (with minimization and numerical integration techniques). A good approximation of λ^* was found in [1] to be $\tilde{\lambda}(\beta) = \frac{\sigma^2}{\beta}$, where β is the standard deviation of X. It results in less than .8% deviation from the minimum MSE (with the optimal threshold λ^*). The threshold $\tilde{\lambda}(\beta)$ is simple and effective and has an intuitive explanation. When the noise power is much smaller than the signal power, $\sigma/\beta << 1$, the normalized threshold $\tilde{\lambda}/\sigma$ is small to preserve most of the signal features; on the other hand, when $\sigma/\beta >> 1$, $\tilde{\lambda}/\sigma$ is chosen to be large to remove the noise which has overwhelmed the signal. By allowing the unknown parameter β to be estimated, this method also allows a *data-driven* selection of the threshold which adapts to each subband.

B. Combining Thresholding and Averaging

When there are multiple copies available, the standard method is to use the (pixel-wise) weighted average as the estimate. Let $V^{(n)} \sim N(0, \sigma_n^2), n = 1, \ldots, N$, be the noise variable for the n-th copy, and Z be the weighted average of $Y^{(n)} = X + V^{(n)}$, $Z = \sum_{n=1}^{N} \alpha_n Y^{(n)} = X + \sum_{n=1}^{N} \alpha_n V^{(n)}$, where $\sum \alpha_n = 1$. It is well-known that the optimal α_n are $\alpha_n^* = \frac{1}{\sigma_n^2} / \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$, and the resulting MSE is $\sigma_{\text{total}}^2 = \text{Var}(Z - X) = \text{Var}\left(\sum_{n=1}^{N} \alpha_n^* V^{(n)}\right) = \left(\sum_{n=1}^{N} \frac{1}{\sigma_n^2}\right)^{-1}$, where $\text{Var}(\cdot)$ denotes the variance.

Now let us incorporate thresholding into averaging. Z is a new random variable with $Z|X \sim N(x, \sigma_{\text{total}}^2)$. Since this is exactly the setting for one copy thresholding, we can simply use $\sigma_{\text{total}}/\beta$ as the threshold. However, can we do better than that? More specifically, since we have two operations here, averaging and thresholding, of which one is linear and the other not, the ordering could make a difference. Thus we investigate which ordering is best in the mean squared sense.

Consider the special case when $\sigma_1 = \sigma_2 = \cdots = \sigma_N \triangleq \sigma$. Thus, $\alpha_1 = \cdots = \alpha_N = \frac{1}{N}$. To make references more convenient, let $\mathcal{A}(\cdot)$ denote the weighted average operation and $\mathcal{T}(\cdot)$ the thresholding operation, and we give the following notation to the two orderings:

$$\mathcal{A}(\mathcal{T}(Y^{(1)}, \ldots, Y^{(N)})) \quad : \qquad \hat{X}_{\mathcal{AT}}(\lambda) = \frac{1}{N}\sum_{n=1}^{N} \eta_\lambda(Y^{(n)}) \tag{2}$$

$$\mathcal{T}(\mathcal{A}(Y^{(1)}, \ldots, Y^{(N)})) \quad : \qquad \hat{X}_{\mathcal{TA}}(\lambda) = \eta_\lambda\left(\frac{1}{N}\sum_{n=1}^{N} Y^{(n)}\right). \tag{3}$$

The risk of the estimator $\hat{X}_{\mathcal{AT}}(\lambda)$ is

$$R_{\mathcal{AT}}(\lambda) = E_X E_{Y^{(1)},\ldots,Y^{(N)}|X} \left(\frac{1}{N} \sum_{n=1}^{N} \eta_\lambda(Y^{(n)}) - X \right)^2 \tag{4}$$

$$= E_X E_{Y^{(1)},\ldots,Y^{(N)}|X} \left[\frac{1}{N} \sum_{n=1}^{N} \left(\eta_\lambda(Y^{(n)}) - X \right) \right]^2 \tag{5}$$

$$= E_X E_{Y^{(1)},\ldots,Y^{(N)}|X} \left[\frac{1}{N^2} \sum_{n=1}^{N} \left(\eta_\lambda(Y^{(n)}) - X \right)^2 + \frac{1}{N^2} \sum_{n=1}^{N} \sum_{\substack{m=1 \\ m \neq n}}^{N} \left(\eta_\lambda(Y^{(n)}) - X \right) \left(\eta_\lambda(Y^{(m)}) - X \right) \right] \tag{6}$$

$$= \frac{1}{N} E_X E_{Y|X} (\eta_\lambda(Y) - X)^2 + \frac{N-1}{N} E_X \left[E_{Y|X}(\eta_\lambda(Y) - X) \right]^2, \tag{7}$$

where (7) follows from the fact that $\{Y^{(1)}, \ldots, Y^{(N)}\}$ conditioned on X are independent. The risk of $\hat{X}_{\mathcal{TA}}(\lambda)$ is

$$R_{\mathcal{TA}}(\lambda) = E_X E_{Y^{(1)},\ldots,Y^{(N)}|X} (\hat{X}_{\mathcal{TA}}(\lambda) - X)^2 = E_X E_{Z|X} (\eta_\lambda(Z) - X)^2, \tag{8}$$

where $Z = \frac{1}{N} \sum_{n=1}^{N} Y^{(n)}$, and $Z|X \sim N(x, \sigma^2/N)$. The optimal thresholds are $\lambda^*_{\mathcal{AT}} = \arg\min_\lambda R_{\mathcal{AT}}(\lambda)$ and $\lambda^*_{\mathcal{TA}} = \arg\min_\lambda R_{\mathcal{TA}}(\lambda)$. Note that $\lambda^*_{\mathcal{AT}}$ and $\lambda^*_{\mathcal{TA}}$ depend on β and σ. We do not have closed form solutions for $\lambda^*_{\mathcal{AT}}$ and $\lambda^*_{\mathcal{TA}}$, thus we resort to numerical methods. Without loss of generality, we can set $\sigma = 1$ (alternatively, one can solve for $\lambda^*_{\mathcal{AT}}/\sigma$ and $\lambda^*_{\mathcal{TA}}/\sigma$ as a function of β/σ). For a fixed value β_0, $\lambda_{\mathcal{TA}}(\beta_0)$ and $\lambda_{\mathcal{AT}}(\beta_0)$ are found by locating the zero crossing of the derivatives of the risks, $R'_{\mathcal{TA}}(\lambda)$ and $R'_{\mathcal{AT}}(\lambda)$, respectively. Standard numerical integration routines (such as the Romberg integration found in Numerical Recipes) are used to calculate the expectations.

To compare the risks of these two methods, we look at the scaled risk difference, $(R_{\mathcal{AT}}(\lambda^*_{\mathcal{AT}}) - R_{\mathcal{TA}}(\lambda^*_{\mathcal{TA}}))/\sigma^2$, as a function of N and of the ratio β/σ, illustrated in Figure 1. For each $N \leq 5$, there is a cutoff point C^*_N below which $R_{\mathcal{AT}}(\lambda^*_{\mathcal{AT}}) > R_{\mathcal{TA}}(\lambda^*_{\mathcal{TA}})$, and above which $R_{\mathcal{AT}}(\lambda^*_{\mathcal{AT}}) < R_{\mathcal{TA}}(\lambda^*_{\mathcal{TA}})$. For $N > 5$, however, the $\mathcal{T}(\mathcal{A}(\cdot))$ method is better for any value of β/σ. The cutoff points C^*_N for each N are listed in Table I. This indicates that *the best method depends on the relative power between the noise and signal, and also on the value of* N. With the optimal thresholds, the improvement of one method over the other is small, however, on the order of $10^{-3}\sigma^2$. $\mathcal{T}(\mathcal{A}(\cdot))$ requires much less computation than $\mathcal{A}(\mathcal{T}(\cdot))$ (since the former can be implemented by computing the wavelet transform once, whereas the latter computes it N times). Thus if computation is an issue, $\mathcal{T}(\mathcal{A}(\cdot))$ is preferred. However, thresholding each individual copy of the image may be advantageous. It is possible that a different noisy copy is collected and processed at each receiving station, and only this processed copy is kept. At a later time, these separately processed copies can be collected by a central receiver to yield one better copy.

By plotting the numerically found $\lambda^*_{\mathcal{TA}}/\sigma$ and $\lambda^*_{\mathcal{AT}}/\sigma$ against β/σ and N, we discover that there is a simple analytical expression which well approximates $\lambda^*_{\mathcal{TA}}$ and $\lambda^*_{\mathcal{AT}}$. For the $\mathcal{T}(\mathcal{A}(\cdot))$ estimator, the threshold is simply a modification of $\tilde{\lambda}$ for one copy denoising, but with a change in the noise variance,

$$\tilde{\lambda}_{\mathcal{T}\mathcal{A}} = \frac{\sigma^2/N}{\beta}. \tag{9}$$

For the $\mathcal{A}(\mathcal{T}(\cdot))$ method, we found the approximation

$$\tilde{\lambda}_{\mathcal{A}\mathcal{T}} = \frac{\sigma^2/N^{(3/4)}}{\beta}. \tag{10}$$

Figure 2 compares the optimal and approximate thresholds for both methods as a function of N, for $\sigma = 1$ and $\beta = 1$. The thresholds $\tilde{\lambda}_{\mathcal{T}\mathcal{A}}$ and $\tilde{\lambda}_{\mathcal{A}\mathcal{T}}$ result in less than .2% deviation from the minimum MSE (with $\lambda^*_{\mathcal{T}\mathcal{A}}$ and $\lambda^*_{\mathcal{A}\mathcal{T}}$, respectively). Figure 3 compares the optimal threshold $\lambda^*_{\mathcal{A}\mathcal{T}}$ and the approximation $\tilde{\lambda}_{\mathcal{A}\mathcal{T}}$ (scaled by $1/\sigma$) as a function of β/σ for $N = 2, \ldots, 6$. The deviation from the optimal MSE deviating from the optimal MSE by less than 3.5% for $\beta/\sigma < 1$ and less than .1% for $\beta/\sigma > 1$. Since typically the signal power is much larger than the noise power (otherwise the image features would be much destroyed), inaccurate approximations for small β/σ are acceptable. The thresholds $\tilde{\lambda}_{\mathcal{T}\mathcal{A}}$ and $\tilde{\lambda}_{\mathcal{A}\mathcal{T}}$ also yield a different set of cutoff values $\tilde{C}_N$ (tabulated in Table I for $N \leq 5$). Notably for $N > 5$, there are some $\tilde{C}_N$ less than ∞, but even above these $\tilde{C}_N$, the MSE of the two methods are so close that either one can be chosen. The scaled MSE difference $(R_{\mathcal{A}\mathcal{T}}(\tilde{\lambda}_{\mathcal{A}\mathcal{T}}) - R_{\mathcal{T}\mathcal{A}}(\tilde{\lambda}_{\mathcal{T}\mathcal{A}}))/\sigma^2$ is similar to the curves shown in Figure 2 for optimal thresholds and is of the same order of magnitude. Thus, the use of $\tilde{\lambda}_{\mathcal{T}\mathcal{A}}$ and $\tilde{\lambda}_{\mathcal{A}\mathcal{T}}$ do not change the previous conclusions.

Notice that the threshold for $\mathcal{A}(\mathcal{T}(\cdot))$ decreases as N increases, even though at the thresholding stage, each copy is thresholded independently of the other copies. To explain this dependency on N, we re-write the inner expectation of $R_{\mathcal{A}\mathcal{T}}(\lambda)$ as

$$\begin{aligned}
& E_{Y^{(1)},\ldots,Y^{(N)}|X}(\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - X)^2 && (11)\\
= \; & E_{\{Y^{(n)}\}|X}\left[\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) + E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - X\right]^2 && (12)\\
= \; & E_{\{Y^{(n)}\}|X}\left[\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda)\right]^2 + \left(E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - X\right)^2 && \\
& +2E_{\{Y^{(n)}\}|X}\left[\left(\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda)\right)\left(E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - X\right)\right] && (13)\\
= \; & E_{\{Y^{(n)}\}|X}\left[\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda)\right]^2 + \left(E_{\{Y^{(n)}\}|X}\hat{X}_{\mathcal{A}\mathcal{T}}(\lambda) - X\right)^2 && (14)\\
= \; & E_{\{Y^{(n)}\}|X}\left[\frac{1}{N}\sum_n \eta_\lambda(Y^{(n)}) - E_{Y|X}\eta_\lambda(Y)\right]^2 + \left(E_{Y|X}\eta_\lambda(Y) - X\right)^2 && (15)\\
= \; & \frac{1}{N}E_{Y|X}\left(\eta_\lambda(Y) - E_{Y|X}\eta_\lambda(Y)\right)^2 + \left(E_{Y|X}\eta_\lambda(Y) - X\right)^2, && (16)
\end{aligned}$$

where $\{Y^{(n)}\}$ has been used as a shorthand for $Y^{(1)}, \ldots, Y^{(N)}$. The first term is the variance due to thresholding, while the second term is the square of the bias. The optimal threshold is obtained from the tradeoff between the variance term (which decreases with increasing λ) and the bias term (which increases with increasing λ). As N becomes large, the variance term decreases due to the $1/N$ factor while the bias term stays the same. Thus, λ needs to be decreased as well to obtain the minimum total.

Up to now we have assumed the knowledge of the noise variance σ^2 and the standard deviation, β, of X. In practice, these two values may not be known and need to be estimated from the noisy observations. For both methods, these two parameters are estimated the same way for a fair comparison. First the noise variance σ_n^2 is estimated by the robust median estimator in the highest subband (also used in [3]), $\hat{\sigma}_n = \text{Median}(|Y_{ij}^{(n)}|)/.6745$, with all $Y_{ij}^{(n)}$ in the HH_1 subband of the n-th copy, then $\hat{\sigma}^2$ is taken to be the average of these N estimates. Since the noise is independent from the signal, $\text{Var}(Z) = \text{Var}(X) + \sigma^2/N = \beta^2 + \sigma^2/N$. Thus, for each subband of $\boldsymbol{Z} = \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \boldsymbol{Y}^{(n)}$, the sample variance estimate of $\text{Var}(Z)$, $\hat{\sigma}_Z^2 = \text{Average}(Z_{ij} - \text{Mean}(Z_{ij}))^2$, is calculated, and the estimate of the standard deviation of the Laplacian distribution is

$$\hat{\beta} = \sqrt{\max(0, \hat{\sigma}_Z^2 - \hat{\sigma}^2/N)}\ . \tag{17}$$

Heterogeneous Noise Variances Now consider the case when the noise variances σ_n^2 are different. This extension is straightforward in the $\mathcal{T}(\mathcal{A}(\cdot))$ case. The multiple copies are averaged with coefficients α_n^*, and the threshold is $\tilde{\lambda}_{\mathcal{TA}}$ in (9) but with σ^2/N replaced by σ_{total}^2.

For the $\mathcal{A}(\mathcal{T}(\cdot))$ method, one needs to find the optimal threshold λ_n for each copy and the optimal weights α_n. By minimizing the risk $E_X E_{Y^{(1)},\ldots,Y^{(N)}|X} \left(\sum_{n=1}^N \alpha_n \eta_{\lambda_n}(Y^{(n)}) - X\right)^2$ with respect to $\alpha_1, \ldots, \alpha_N$ subject to $\sum \alpha_n = 1$, and also with respect to $\lambda_1, \ldots, \lambda_N$, one can find the optimal values. The optimal α_n are found numerically to be almost identical to the values in α_n^*. The optimal thresholds can be approximated by $\tilde{\lambda}_{\mathcal{AT}}^{(n)} = \frac{\sigma_n^{1/2}}{\beta}\left(1/(\sum_{i=1}^N \frac{1}{\sigma_i^2})\right)^{3/4}$, $n = 1, \ldots, N$, which yields $\tilde{\lambda}_{\mathcal{AT}}$ in (10) when $\sigma_1 = \sigma_2 = \cdots = \sigma_N$. Figure 4 compares the optimal thresholds (—) and the thresholds $\tilde{\lambda}_{\mathcal{AT}}^{(n)}$ ($\cdots$) for $N = 3$ and $\{\sigma_n\} = \{1, 3, 5\}$, against β on the horizontal axis. The plot shows the best fit for the threshold corresponding to $\min\{\sigma_n\}$, and the approximation worsens for thresholds whose corresponding σ_n is further from $\min\{\sigma_n\}$, especially in the region of small β (relative to $\min\{\sigma_n\}$). This inaccuracy is mitigated by the fact that for the α_n^*'s for large σ_n's are small, thus the overall MSE is still close to the optimal MSE (less than .3% for $\beta > 1$ for the graph shown).

III. Experimental Results

To validate our proposed methods, we take as the test image a 256×256 block from the greyscale images *barbara* and *lena*, with $\sigma_1 = \ldots = \sigma_N = \sigma = 30$, using Daubechies' least unsymmetric wavelet with 8 vanishing moments (i.e. Symmlet8) and 4 scales of wavelet transform. The algorithm is not too sensitive to the choices of wavelet and number of levels, as long as the wavelet is smooth enough (has several vanishing moments) and at least 3 or 4 levels are used for this magnitude of σ (for example, using Symmlet4 increases the MSE by 1-2% and using 2 levels of decomposition increases the MSE by 2-4%). The parameters β and σ are estimated as in the previous discussion. We compare the MSEs of five methods for N ranging from 1 to 25: (a) averaging, (b) $\mathcal{A}(\mathcal{T}(\cdot))$, (c) $\mathcal{T}(\mathcal{A}(\cdot))$, (d) switching between the two thresholding methods (only for $N \leq 5$) with cutoff values $\tilde{C}_N$ (thus the switching method becomes

$\mathcal{T}(\mathcal{A}(\cdot))$ for $N > 5$), and (e) Wiener filtering of the averaged copy (from (a)). The Wiener filter is the "wiener2" routine from the MATLAB image processing toolkit, with the input noise power unknown and the adaptation window size set to the default (3×3). The resulting MSEs are shown in Figure 5 for *barbara* and Figure 6 for *lena*. The three thresholding methods show significant improvement over merely averaging, ranging from 70% to 30% reduction in MSE for N varying from 2 to 25, and are either comparable or up to 5-15% improvement over Wiener filtering (depending on N). The removal of noise due to thresholding is also significant visually (see Figure 7 for *barbara* at $N = 4$), especially for small N. For $N = 1$ and $\sigma = 30$, Wiener filtering does poorly because of the large noise power. For larger N, the averaging reduces the noise power substantially, and thus Wiener filtering yields images comparable with those from thresholding, though slightly more blurry. Among the thresholding methods, the $\mathcal{T}(\mathcal{A}(\cdot))$ method is the best in terms of MSE, even better than switching, suggesting that perhaps the $\mathcal{A}(\mathcal{T}(\cdot))$ method is more sensitive to model errors and threshold estimation errors. For large values of σ (e.g. 30), $\mathcal{A}(\mathcal{T}(\cdot))$ produces noticeably more noisy images than $\mathcal{T}(\mathcal{A}(\cdot))$, though for smaller values, it yields images of comparable quality as those from $\mathcal{T}(\mathcal{A}(\cdot))$. For $1 < N \leq 5$, the switching method yields MSEs that are between those of $\mathcal{A}(\mathcal{T}(\cdot))$ and $\mathcal{T}(\mathcal{A}(\cdot))$. The $\mathcal{T}(\mathcal{A}(\cdot))$ method requires the least amount of computation since it can be implemented with only one wavelet transform and seems to work well for large values of σ as well. Thus, in practice, the $\mathcal{T}(\mathcal{A}(\cdot))$ method suffices to combine multiple noisy copies.

It is curious to investigate if an additional stage of thresholding (i.e. performing $\mathcal{T}(\mathcal{A}(\mathcal{T}(\cdot)))$) can have a significant improvement. It cannot do worse, since we can always choose the latter stage threshold to be zero. To test this idea, we take the output of $\mathcal{A}(\mathcal{T}(\cdot))$ and optimally threshold it assuming that we have the original. The resulting MSE is only slightly better than the $\mathcal{T}(\mathcal{A}(\cdot))$, suggesting that thresholding of the weighted average yields a sufficiently denoised image already. Furthermore, finding the optimal thresholds of a two-stage thresholding operation is difficult.

IV. Conclusion

In this paper we addressed the issue of image recovery from its multiple noisy copies. We explored the idea of combining the wavelet thresholding technique with the more traditional averaging operation. Our investigation showed that the optimal ordering of these two operations is not so straightforward and is in fact a function of the number of available copies and of the relative energy between noise and signal. We proposed a near-optimal threshold for each ordering. With these thresholds, the performances are similar, and for computational reasons, averaging followed by thresholding is recommended. Furthermore, all of these thresholding methods show substantial improvement over mere averaging and moderate improvement over Wiener filtering, both visually and in the MSE sense.

References

[1] S.G. Chang, B. Yu, and M. Vetterli, "Image Denoising via Lossy Compression and Wavelet Thresholding," *Proc. IEEE Int. Conf. Image Processing*, Vol.1, pp. 604-607, Nov. 1997.

[2] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1992.

[3] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol 81, pp. 425-455, 1994.

[4] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, July 1989.

[5] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.

[6] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, July 1989.

[7] P. Moulin and J. Liu, "Analysis of Multiresolution Image Denoising Schemes Using Generalized-Gaussian and Complexity Priors," *IEEE Trans. Info. Theory*, to appear, 1999.

[8] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[9] P.H. Westerink, J. Biemond, and D.E. Boekee, "An optimal bit allocation algorithm for sub-band coding," *Proc. Int. Conf. on Acous., Speech and Signal Process.*, pp. 1378–1381, Dallas, TX, April 1987.

TABLE I

CUTOFF VALUES (IN UNIT σ_x/σ) FOR EACH N, WHERE C_N^* IS THE CUTOFF VALUE FOR WHEN USING THE OPTIMAL THRESHOLDS, AND $\tilde{C}_N$ (LISTED ONLY FOR $N \leq 5$) IS THE CUTOFF VALUE WHEN USING THE PROPOSED THRESHOLDS, $\tilde{T}_{\mathcal{AT}}$ AND $\tilde{T}_{\mathcal{TA}}$.

	C_N^*	$\tilde{C}_N$
$N=2$	.6367	.1379
$N=3$	.7154	.7654
$N=4$	.9601	.9466
$N=5$	1.9768	1.0884
$N>5$	∞	>1.23

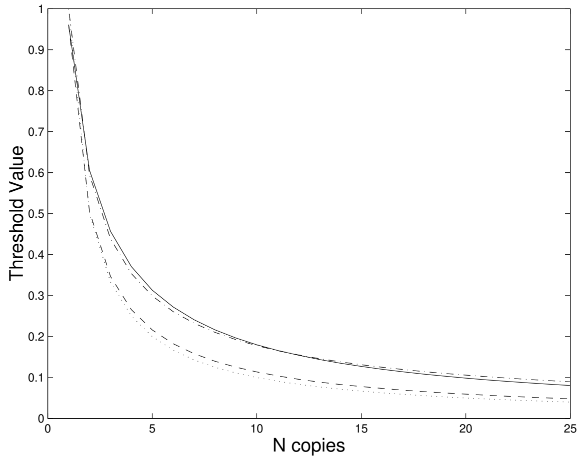

Fig. 2. Comparing $\lambda^*_{\mathcal{TA}}$ $(---)$ versus $\tilde{\lambda}_{\mathcal{TA}}$ $(\cdots)$, and $\lambda^*_{\mathcal{AT}}$ (—) versus $\tilde{\lambda}_{\mathcal{AT}}$ $(-\cdot-\cdot)$, when $\sigma = 1$ and $\beta = 1$.

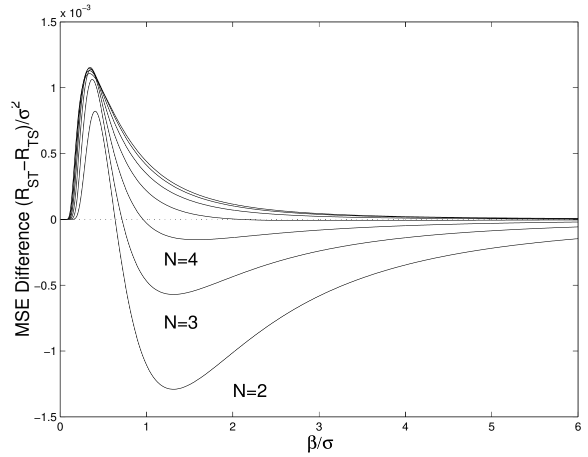

Fig. 1. Scaled MSE difference $(R_{\mathcal{AT}}(\lambda^*_{\mathcal{AT}}) - R_{\mathcal{TA}}(\lambda^*_{\mathcal{TA}}))/\sigma^2$ as a function of N and β/σ.

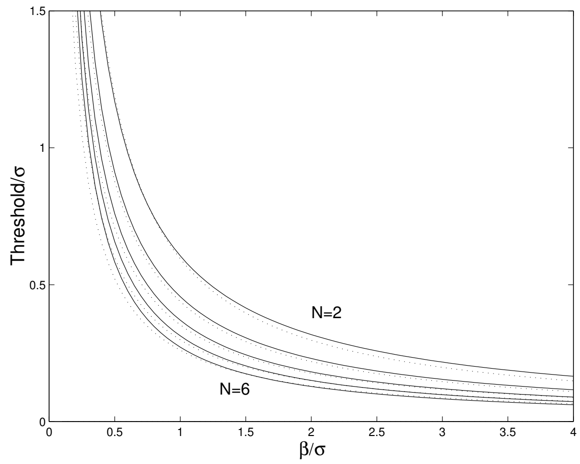

Fig. 3. Comparing $\lambda^*_{\mathcal{AT}}$ (—) and $\tilde{\lambda}_{\mathcal{AT}}$ $(\cdots)$ for $\sigma_1 = \cdots, \sigma_N \stackrel{\triangle}{=} \sigma$ as a function of β/σ and $N = 2, \ldots, 6$.

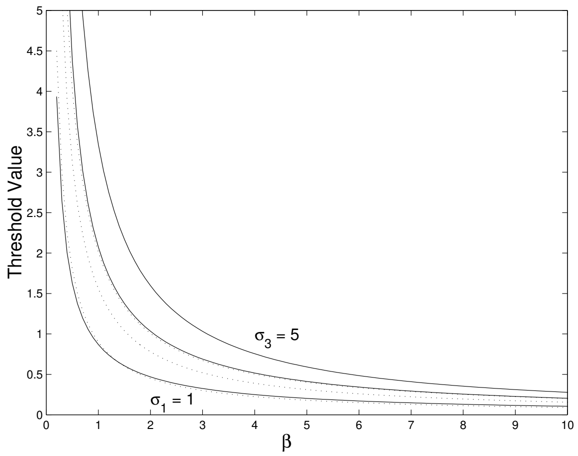

Fig. 4. For $N = 3$ and different noise σ_n ([1,3,5]), compare the optimal threshold for $\mathcal{A}(\mathcal{T}(\cdot))$ (—) and $\tilde{\lambda}^{(n)}_{\mathcal{AT}}$ $(\cdots)$.

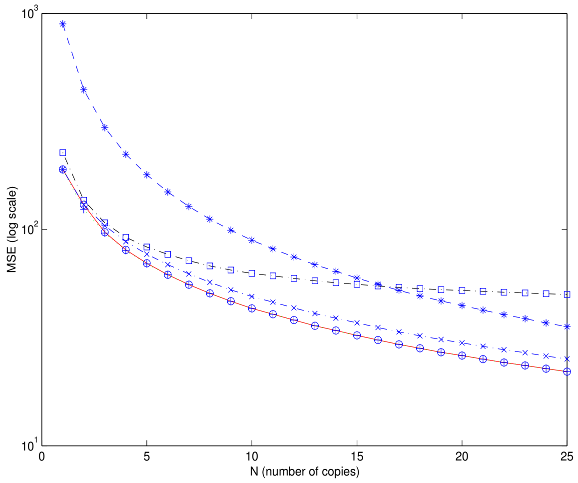

Fig. 5. For the image *barbara*, comparing as a function of N the MSE of averaging ($*$), $\mathcal{A}(\mathcal{T}(\cdot))$ (x), $\mathcal{T}(\mathcal{A}(\cdot))$ ($+$), switching (o), and Wiener filtering ($\square$), for $\sigma = 30$. Note that the ($+$) and (o) curves are overlapping.

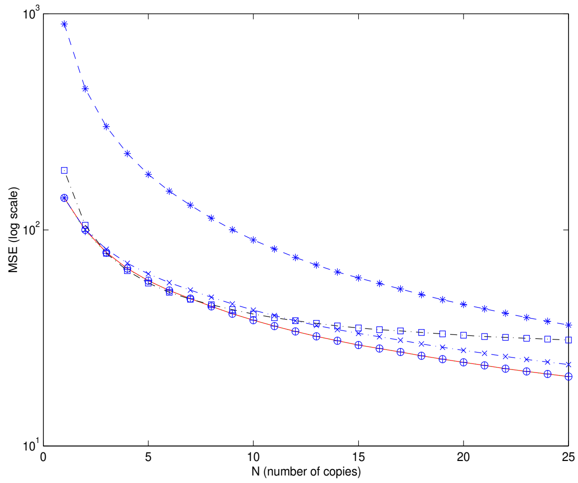

Fig. 6. For the image *lena*, comparing as a function of N the MSE of averaging ($*$), $\mathcal{A}(\mathcal{T}(\cdot))$ (x), $\mathcal{T}(\mathcal{A}(\cdot))$ ($+$), switching (o), and Wiener filtering ($\square$), for $\sigma = 30$. Note that the ($+$) and (o) curves are overlapping.

Fig. 7. Denoised images, for $N = 4$. From left to right, top to bottom: original, noisy image with $\sigma = 30$, averaging, switching, $\mathcal{A}(\mathcal{T}(\cdot))$, $\mathcal{T}(\mathcal{A}(\cdot))$, and Wiener filtering. This image (and others) can be found on the Web, at `http://www-wavelet.eecs.berkeley.edu/~grchang/multiThresh/`.