

Image Denoising via Residual Kurtosis Minimization

Tristan A. Hearn* Lothar Reichel†

Abstract

A new algorithm for the removal of additive uncorrelated Gaussian noise from a digital image is presented. The algorithm is based on a data driven methodology for the adaptive thresholding of wavelet coefficients. This methodology is derived from higher order statistics of the residual image, and requires no a priori estimate of the level of noise contamination of an image.

1 Introduction

With the mass production of digital images of many kinds, the need for efficient methods to restore corrupted images has become immediate. A major source of image contamination is noise. Image noise may arise from quantization of the image data, transmission errors, electronic interference from the imaging hardware, as well as from other sources [1].

Many methods have been proposed for denoising of digital images. Some are based on finite-impulse response (FIR) filtering, where the noisy image is convolved with a smoothing kernel to reduce the visible effect of noise. This may introduce undesirable artificial blur to an image [16]. Some methods cannot reliably distinguish between noise and edge information in an image; they may require a priori knowledge of the noise level in the contaminated image, or might introduce waveform artifacts. A recent discussion of many available methods can be found in [2].

Denoising methods based on the discrete wavelet transform are popular. These methods are similar to spectral filtering methods, but can take advantage of the partial localization in space and frequency offered by wavelet transforms. This allows efficient separation of noise from impulsive signal

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.
E-mail: thearn@math.kent.edu.

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.
E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-1115385.

information, such as edges [10, 11]. Denoising is achieved by reducing the magnitude of specific wavelet coefficients.

Many denoising methods assume that the contaminating noise has a particular distribution with known parametrization. That is, some a priori estimate of the noise parameters must be available before denoising can be performed. This paper proposes a new wavelet-based method for the reduction of Gaussian noise in a digital image, and does not require an estimate of the noise parameters to be available. Instead, the method uses the kurtosis statistic of the residual image formed as the difference between the available noisy image and the computed candidate denoised image to determine how much denoising should be carried out. Our approach is based on the observation that the kurtosis statistic is minimal when the residual image is made up of Gaussian noise. We determine the amount of denoising to be carried out by minimizing the kurtosis statistic of the residual image. When the kurtosis statistic is minimal, the residual image is likely to be made up of Gaussian noise, i.e., to be the noise contamination of the available image. We remove the noise in the contaminated image by reducing the magnitude of specific wavelet coefficients in its representation (soft shrinkage). This approach makes it possible to dispense with an a priori estimate of the noise variance in the available noise-contaminated image.

A few numerical examples show our denoising method to be competitive with the BayesShrink method [3] and a Total Variation (TV) norm based denoising scheme. A careful comparison with the many available denoising methods is outside the scope of the present paper. However, we would like to point out that the kurtosis statistic also can be applied in conjunction with other denoising methods to reduce the number of parameters that have to be chosen by a user. We are currently investigating this approach.

This paper is organized as follows. Section 2 reviews the methodology of denoising through thresholding of wavelet coefficients. In Section 3, we derive our new method, which is based on the use of higher order statistics. Section 4 presents numerical experiments of the denoising of digital images (corrupted by Gaussian noise) using the proposed method. Concluding remarks can be found in Section 5.

2 Denoising in the wavelet transform domain

The use of the discrete wavelet transform for filtering additive zero-mean Gaussian noise was first proposed by Donoho and Johnson [10]. Consider the following model for the corruption of a gray-scale digital image by additive noise

$$G = F + E,$$

where $G, F, E \in \mathbb{R}^{m \times n}$. Here F is a matrix that encodes the unavailable noise-free image, the matrix E represents zero-mean, independent and identically distributed Gaussian noise with common variance σ^2 , and G is the available

observed noisy image. Let

$$\widehat{G} = \mathcal{W}_{\phi,\psi,j}G$$

represent a j -level 2D discrete wavelet transform of the observed image G with low-pass filter ϕ and high-pass filter ψ . Then \widehat{G} has a multiresolution subband structure comprised of the $3j$ detail matrices

$$LH_i, HL_i, HH_i \in \mathbb{R}^{\frac{m}{2^i} \times \frac{n}{2^i}}, \quad i = 1, \dots, j, \quad (1)$$

and the low-resolution approximation matrix

$$LL_j \in \mathbb{R}^{\frac{m}{2^j} \times \frac{n}{2^j}},$$

for a total of $3j + 1$ coefficient matrices. We assume for notational simplicity that m and n are multiples of 2^j . This restriction easily can be removed.

By the linearity of the discrete wavelet transform, it follows that the wavelet coefficients of \widehat{G} also will be contaminated by additive noise. We have

$$\widehat{G} = \mathcal{W}_{\phi,\psi,j}G = \mathcal{W}_{\phi,\psi,j}F + \mathcal{W}_{\phi,\psi,j}E.$$

Let the filter pair ϕ, ψ be chosen so that $\mathcal{W}_{\phi,\psi,j}$ is an orthonormal wavelet transform. Then the wavelet coefficients in any particular subband of

$$\widehat{E} = \mathcal{W}_{\phi,\psi,j}E$$

also will be contaminated by additive zero-mean Gaussian noise. Since the detail subbands are all zero-mean and Gaussian, this suggests that it may be possible to denoise G by treating the wavelet coefficients in each subband in a similar manner, e.g., by thresholding in each subband.

Denosing methods based on wavelet coefficient thresholding operate by choosing some parameter $\lambda > 0$ for reducing (shrinking) the magnitude of the coefficients in each detail subband. The two thresholding methods most commonly used are known as *hard thresholding* and *soft thresholding*. Given a detail coefficient matrix D , a hard thresholding filter H_λ truncates each component $D_{i,j}$ of D according to

$$H_\lambda(D_{i,j}) = D_{i,j} \chi\{|D_{i,j}| > \lambda\},$$

where χ denotes the characteristic function. In contrast, a soft thresholding filter S_λ smoothly attenuates (rather than truncates) the wavelet coefficients $D_{i,j}$ of a detail subband D according to

$$S_\lambda(D_{i,j}) = \text{sgn}(D_{i,j}) \max(|D_{i,j}| - \lambda, 0).$$

Thresholding methods only apply coefficient filtering to the detail subbands $LH_i, HL_i, HH_i, i = 1, \dots, j$, but not to the low-resolution approximation matrix LL_j . It is assumed that the subbands to which the thresholding filters are applied are sparse, i.e., only few wavelet coefficients are

nonvanishing. The low-pass approximation subband LL_j generally is not sparse. Therefore, if LL_j is contaminated by noise, then an additional wavelet transform should be applied to LL_j to produce a new decomposition level. Thresholding operations may then be applied to the new detail subbands so obtained. It is our aim to choose the decomposition level j so that LL_j primarily contains information pertaining to F and not to E .

Most wavelet denoising methods for digital image restoration apply soft thresholding, because this yields images that tend to be more visually pleasing than images obtained with hard thresholding [11]. This is due to the discontinuous nature of the hard thresholding rule, which can lead to synthetic waveform artifacts in the resulting denoised image. Moreover, it has been shown that denoising based on soft thresholding of wavelet coefficients attains asymptotically near-optimal results (in a minimax sense) in a number of different function spaces (such as suitable Besov spaces); see [11].

Consider the soft thresholding operation $S_{\{\lambda_1, \dots, \lambda_j\}}(\mathcal{W}_{\phi, \psi, j}G)$ across all detail subbands of the first j levels (1) using the threshold values $\{\lambda_1, \dots, \lambda_j\}$. An approximation of F is obtained by applying the appropriate j -level inverse 2D discrete wavelet transform to the filtered wavelet coefficients:

$$\hat{F} = \mathcal{W}_{\phi, \psi, j}^{-1} S_{\{\lambda_1, \dots, \lambda_j\}}(\mathcal{W}_{\phi, \psi, j}G).$$

The main difference between existing wavelet denoising methods and the method proposed in this paper is how the thresholding values are determined.

3 Residual driven denoising

Our denoising method applies soft thresholding to the wavelet coefficients. However, rather than searching for values of the thresholding parameters to minimize an objective function in the wavelet transform domain, our method seeks to minimize a function related to the difference \hat{E} between the noisy and denoised images,

$$\hat{E} = G - \hat{F}.$$

When \hat{F} is an accurate approximation of the exact image F , one would expect \hat{E} to be an accurate approximation of the additive noise E . For instance, an estimate $\hat{\sigma}^2$ of the noise variance σ^2 can be obtained from \hat{E} by a sample estimator, such as the maximum likelihood estimator

$$\hat{\sigma}^2 = \frac{1}{mn} \sum_{p=1}^m \sum_{q=1}^n \left(\hat{E}_{p,q} - \bar{\hat{E}} \right)^2,$$

where $\bar{\hat{E}}$ denotes the average of the entries $\hat{E}_{p,q}$ of \hat{E} .

If σ^2 is known a priori, then the wavelet coefficient thresholds should be selected so that the variance $\hat{\sigma}^2$ of the computed residual image \hat{E} is close to σ^2 . However, our method does not require σ^2 to be known. Instead, we

determine wavelet coefficient thresholds using a higher order sample statistic, derived in a manner that does not explicitly require an a priori estimate of the noise variance σ^2 . This approach differs from available popular wavelet denoising techniques, such as [3, 10], in that the latter demand an a priori estimate of the noise variance.

3.1 The residual kurtosis statistic

Let $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ be a random normal sample, where

$$x_1, \dots, x_n \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

The kurtosis κ_1 of X is a sample statistic defined as

$$\kappa_1 = \frac{m_4}{m_2^2},$$

where m_4 and m_2 are the sample moments

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4, \quad m_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

D'Agostino, Pearson, and Tietjen [7, 8, 17] have shown that in the case of normal samples, a normalized statistic for excess kurtosis is given by

$$\kappa_2 = \sqrt{\frac{n}{24}} \left(\frac{m_4}{m_2^2} - 3 \right) \stackrel{\text{distr.}}{\rightarrow} N(0, 1) \quad (2)$$

with convergence as $n \rightarrow \infty$. Denote the probability that a value of magnitude larger than or equal to κ_2 is attained in the normal sample X by Φ . When interpreted as a measure of normality of the sample X , this probability can be computed as

$$\Phi = 2P(y > |\kappa_2|) = 2 \int_{|\kappa_2|}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 - \text{erf} \left(\frac{|\kappa_2|}{\sqrt{2}} \right),$$

where y is a standard normal random variable $y \sim N(0, 1)$ and $\text{erf}(x)$ denotes the Gauss error function. This probability can be evaluated using standard tables or by numerical quadrature.

We remark that the robustness of using the kurtosis statistic as a measure of normality depends on the sample size. This has been observed in the literature in the context of other mathematical problems. For example, consider the problem of linear blind source separation of either the columns or rows of a matrix $A \in \mathbb{R}^{m \times n}$. This task can be solved within the framework of independent component analysis (ICA) by decreasing a selected numerical measure of normality of the columns or rows of A , respectively. It has been

pointed out in [13] that kurtosis is not always a robust enough measure of normality for the purpose of ICA when m and n are not sufficiently large.

We assume in this paper that the sampling distribution of κ_2 is accurately characterized by the asymptotic distribution in (2). Then kurtosis is a robust measure of normality. The justification for this assumption is that the sample sizes, i.e., the number of pixels of digital images to be denoised, typically is quite large. For instance, a fairly small digital image made up of 64×64 pixels yields a sample size of $64^2 = 4096$ entries. In comparison with the sample sizes from 30 to 200 considered in [7, 8, 17], this sample size is very large. Furthermore, a lot of images that arise in applications are represented by many more than 64×64 pixels.

3.2 Kurtosis as a denoising criterion

Consider applying the kurtosis statistics κ_1 or κ_2 to all of the components of the residual image, following a level-dependent wavelet coefficient soft-thresholding procedure. Then the resulting values must depend on the selected threshold values. Let $\kappa_1(G, \lambda_i | \lambda_{l < i})$ and $\kappa_2(G, \lambda_i | \lambda_{l < i})$ be the kurtosis statistics κ_1 and κ_2 applied to the residual image formed by denoising the image G using i -level wavelet coefficient thresholding, where λ_i is a variable threshold for the deepest wavelet level, and $\lambda_{l < i}$ are fixed thresholds for all the previous wavelet levels. We write

$$\kappa_1(G, \lambda_i | \lambda_{l < i}) = \frac{mn \sum_{p=1}^m \sum_{q=1}^n \left(\hat{E}_{p,q}(G, \lambda_i | \lambda_{l < i}) - \bar{\bar{E}}(G, \lambda_i | \lambda_{l < i}) \right)^4}{\left[\sum_{p=1}^m \sum_{q=1}^n \left(\hat{E}_{p,q}(G, \lambda_i | \lambda_{l < i}) - \bar{\bar{E}}(G, \lambda_i | \lambda_{l < i}) \right)^2 \right]^2},$$

where

$$\hat{E}(G, \lambda_i | \lambda_{l < i}) = G - \mathcal{W}_{\phi, \psi, j}^{-1} S_{\lambda_i | \lambda_{l < i}}(\mathcal{W}_{\phi, \psi, j} G)$$

is the residual image obtained by wavelet denoising. The likelihood probability Φ therefore is parametrically given by

$$\Phi(G, \lambda_i | \lambda_{l < i}) = 1 - \operatorname{erf} \left(\frac{|\kappa_2(G, \lambda_i | \lambda_{l < i})|}{\sqrt{2}} \right).$$

When a value of the subband threshold λ_i is found that maximizes Φ , this value maximizes the likelihood of normality of the residual image \hat{E} based on the wavelet thresholding at this particular level. This observation is the basis for our denoising method.

Our method first determines a single-level wavelet decomposition of the given noisy image, producing subbands LL_1 , LH_1 , HL_1 , and HH_1 . A threshold for this first level that maximizes Φ is then computed. Equivalently, this threshold minimizes the error function

$$1 - \Phi = \operatorname{erf} \left(\frac{|\kappa_2|}{\sqrt{2}} \right),$$

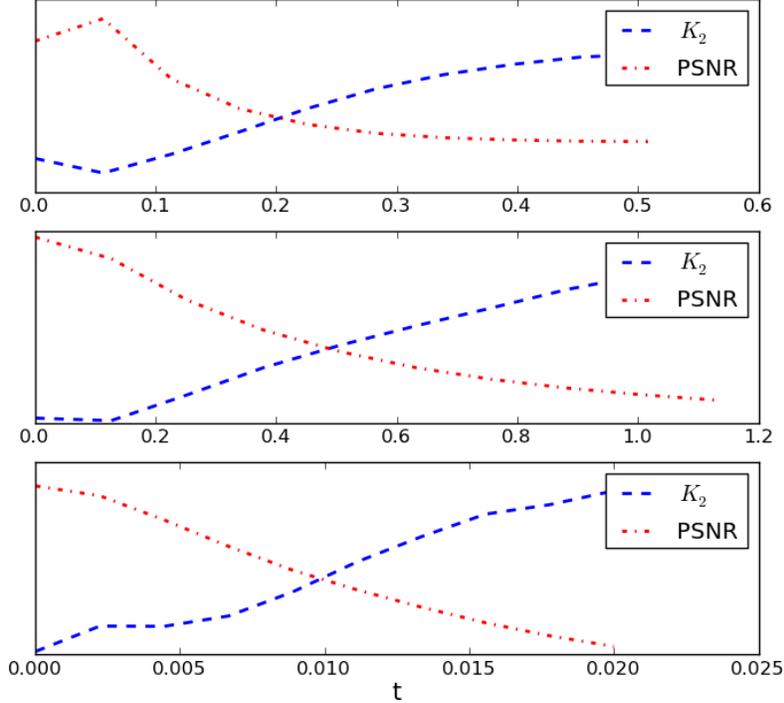


Figure 1: Residual kurtosis κ_2 and PSNR versus the wavelet threshold t for the first, second, and third wavelet levels (from top to bottom).

i.e., it minimizes $|\kappa_2|$.

The minimization generally requires repeated inversion of the wavelet transform to be able to determine the residual image and its kurtosis statistic. Once the optimal threshold has been found, the thresholding operation is applied to the detail subbands. Due to the recursive nature of the multiresolution analysis inherent in a multilevel discrete wavelet transform, the next subbands are found simply by computing another single-level wavelet transform of the current low-pass subband, say LL_j . This produces the new set of subbands LL_{j+1} , LH_{j+1} , HL_{j+1} , and HH_{j+1} , which collectively replace LL_j in the wavelet decomposition. By repeating the threshold optimization procedure and keeping track of all thresholded subbands, the entire denoising procedure can proceed by using only a single multilevel forward wavelet transform, even though the optimization via residual kurtosis minimization generally makes it necessary to compute several inverse wavelet transforms.

Figure 1 illustrates the proposed denoising method at the first three wavelet decomposition levels, and Figure 2 shows the original corrupted image used for this example, as well as the denoised result. At each of the three levels, the residual kurtosis κ_2 is plotted with the peak signal-to-noise ratio (PSNR) between the computed denoised image and the uncorrupted image, as a function of the wavelet coefficient threshold t for the level. The PSNR is a commonly used measure of the quality of a restoration for digital images. Let G denote the restoration and F the (unavailable) exact image.



(a) Noisy image



(b) Restored image

Figure 2: (a) Original noisy image, and (b) a denoised version using the three-level wavelet coefficient thresholds that minimize the residual kurtosis displayed in Figure 1.

The PSNR between F and G is defined as

$$\text{PSNR}(F, G) = 20 \log \left(\frac{255}{\text{RMSE}(F, G)} \right),$$

where $\text{RMSE}(F, G)$ is the root mean squared error given by

$$\text{RMSE}(F, G) = \frac{1}{mn} \sum_{p=1}^m \sum_{q=1}^n (F_{p,q} - G_{p,q})^2.$$

For this work, all sample images were encoded in an eight-bit gray scale format. The numerator of the PSNR, $255 = 2^8 - 1$, is the largest value an eight-bit pixel can achieve. If a different format is used, this numerator should reflect the maximum value that a single pixel may take.

The plots of Figure 1 are normalized so that the extrema of the functions may be easily compared. We notice that at each of the three levels, the value of the threshold t that minimizes κ_2 is close to the value that maximizes the PSNR. It is also notable that the threshold that minimizes the residual kurtosis on the third level is zero. Thus, the residual kurtosis is minimized by not performing any coefficient thresholding at all on the third level. This indicates that the denoising procedure may be terminated.

We now show that the error function $1 - \Phi$ is nonincreasing as the denoising method proceeds to successive wavelet levels.

Theorem 3.1. *The error function $1 - \Phi$ is non-increasing as the wavelet decomposition progresses from level j to level $j + 1$, as long as the employed minimization procedure allows the coefficient thresholds, λ_j , to take on the value zero. Furthermore, the proposed denoising method must terminate.*

Proof. The subbands at level $j + 1$ are computed by applying a one-level discrete wavelet transform to the approximation subband at the j^{th} level,

$$\{LL_{j+1}, LH_{j+1}, HL_{j+1}, HH_{j+1}\} = \mathcal{W}_{\phi, \psi, j} LL_j.$$

Equivalently, the inverse one-level discrete wavelet transform applied to the subbands at the $(j + 1)^{\text{st}}$ level yields the approximation subband LL_j for the j^{th} level,

$$\mathcal{W}_{\phi, \psi, j}^{-1} \{LL_{j+1}, LH_{j+1}, HL_{j+1}, HH_{j+1}\} \rightarrow LL_j.$$

Observe that, for both soft and hard thresholding, evaluation of a wavelet coefficient when the threshold is zero simply returns the coefficient unchanged:

$$\begin{aligned} H_0(D_{i,j}) &= D_{i,j} \chi\{|D_{i,j}| > 0\} = D_{i,j}, \\ S_0(D_{i,j}) &= \text{sgn}(D_{i,j}) \max(|D_{i,j}|, 0) = \text{sgn}(D_{i,j}) |D_{i,j}| = D_{i,j}. \end{aligned}$$

Thus, detail coefficient subbands are invariant under thresholding with the threshold $\lambda = 0$.

Let $\kappa_2(G, j)$ be the normalized kurtosis estimator evaluated at the j^{th} level wavelet decomposition. When minimizing the error function at level $j + 1$, all previously optimized thresholds (i.e., all thresholds for the previous levels) are kept fixed. Therefore,

$$\min \text{erf}(|\kappa_2(G, j + 1)|) \leq \min \text{erf}(|\kappa_2(G, j)|),$$

with equality attained when all the thresholds at the $(j + 1)^{\text{st}}$ level vanish.

Since each successive level of the discrete wavelet decomposition involves a dyadic down-sampling operation, a maximum decomposition level is well defined. It is bounded below by $\min\{\lfloor \log_2(m) \rfloor, \lfloor \log_2(n) \rfloor\}$, where $\lfloor \alpha \rfloor$ denotes the largest integer less than or equal to α . \square

Generally, one may terminate the computations at a level much smaller than the maximum decomposition level for a given image defined in the above proof. The exponential decay of the wavelet coefficients for natural images across scales is well documented; see, e.g., [10, 14]. On the other hand, wavelet coefficients for Gaussian noise do not decay with an increase in the level j . Therefore, typically the signal-to-noise ratio in decomposition level j is a decreasing function of j . This results in that the optimal intensity of the subband thresholding decreases with increasing index level j . We terminate the denoising computations as soon as the difference between the $1 - \Phi$ values for successive wavelet decomposition levels is smaller than a user-specified tolerance $\delta > 0$. This typically results in termination already when only a few wavelet decompositions have been computed. In particular, the decomposition level j should not be considered a denoising parameter.

Algorithm 1 below summarizes our residual kurtosis minimization denoising (RKMD) method. We used this algorithm to determine Figures 1 and 2, as well as in the examples of Section 4. Note that in Figure 1, we plot the kurtosis κ_2 rather than the error function computed by the algorithm.

The computationally most expensive part of the algorithm is the repeated applications of the inverse wavelet transform for the evaluation of the kurtosis of the residual. The number of computed inverse transforms depends on the optimization method used for minimizing the kurtosis of the residual at each decomposition level. It follows that the speed of execution of the RKMD algorithm depends on the choice of optimizer. An upper bound for the total number of forward and inverse wavelet transforms performed by the method can be established.

Proposition 3.2. *Let $\hat{F} \in \mathbb{R}^{m \times n}$ be a denoised image produced by Algorithm 1 with j decomposition levels. Let η be the total number of forward and inverse wavelet transforms computed. Then $\eta \leq 2 + mn(1 - 4^{-j})$. Moreover, $j \leq \min\{\lfloor \log_2(m) \rfloor, \lfloor \log_2(n) \rfloor\}$.*

Proof. The RKMD algorithm requires one forward wavelet transform to begin the denoising procedure and one at the very end to return the denoised

result. In addition, each decomposition level requires the application of a line search procedure for minimizing the kurtosis of the residual image. One inverse wavelet transform has to be applied for each of these evaluations. A discretized brute force approach to this inner-loop optimization would be to enumerate every wavelet coefficient in every detail subband, and compute the residual kurtosis with thresholds equal to each coefficient. Therefore, the number of transforms computed using this brute force technique is a function of the total number of wavelet detail coefficients. This number equals mn , the number of pixels of the original image. Therefore, the number of detail coefficients is given by the number of approximation coefficients subtracted from mn . Due to dyadic down-sampling, the number of detail coefficients is $mn - (2^{-j}m)(2^{-j}n)$. Thus, the total number of forward and inverse wavelet transforms to be computed for brute force optimization is given by

$$\eta = 2 + mn - (2^{-j}m)(2^{-j}n) = 2 + mn(1 - 4^{-j}).$$

□

The implementation of the RKMD algorithm used for the numerical examples in Section 4 applies a quasi-Newton method for residual kurtosis minimization. We observed that the minimum of the residual kurtosis objective function in all examples was well localized.

Algorithm 1 Kurtosis-driven wavelet denoising method

Given noisy image $G \in \mathbb{R}^{m \times n}$
 Select wavelet filters (ϕ, ψ) and tolerance parameter $\delta > 0$
 Let $LL = G$, $D = \{\}$, $p_{\text{last}} = 0$
for $j = 1 \rightarrow \min([\log_2(m)], [\log_2(n)])$ **do**
 Let $\{LL, LH, HL, HH\} = \mathcal{W}_{\phi, \psi, j} LL$
 Use a line search procedure to find $\lambda_j = \arg \min_{\lambda} \text{erf}(|\kappa_2(G, \lambda | \lambda_{l < j})|)$
 Let $p = \text{erf}(|\kappa_2(G, \lambda_j | \lambda_{j-1}, \dots, \lambda_1)|)$
 Append array $[LH, HL, HH]$ to array D
 if $|p_{\text{last}} - p| < \delta$ **then**
 Compute $\hat{F} = \mathcal{W}_{\phi, \psi, j}^{-1} S_{\lambda_j, \dots, \lambda_1} (\mathcal{W}_{\phi, \psi, j} G)$ using the fact that the approximation and filtered coefficients $S_{\lambda_j, \dots, \lambda_1} (\mathcal{W}_{\phi, \psi, j} G)$ are already stored in D .
 Return \hat{F}, p
 else
 Let $p_{\text{last}} = p$
 end if
end for

3.3 Two other denoising methods

The computed examples of the following section compare the RKMD method to two popular denoising methods, BayesShrink [3] and a method based on solving a Tikhonov-minimization problem with a total variation penalty term. We chose to compare with these methods because they have received considerable attention in the literature and they can be implemented largely with public-domain software.

3.3.1 The BayesShrink method

Similarly to RKMD, BayesShrink is a level-adaptive, wavelet-shrinkage denoising method [3]. For BayesShrink, the detail subband coefficients are assumed to conform to a generalized Gaussian distribution in the wavelet transform domain. The generalized Gaussian distribution has a density function of the form

$$G_{\sigma_X, \beta}(x) = C(\sigma_X, \beta) \exp \left\{ -[\alpha(\sigma_X, \beta) |x|]^\beta \right\}$$

with support $-\infty < x < \infty$ and parameters $\sigma_X > 0$ and $\beta > 0$, where σ_X denotes the standard deviation of the signal. The functions $C(\sigma_X, \beta)$ and $\alpha(\sigma_X, \beta)$ are given by

$$C(\sigma_X, \beta) = \frac{\beta \alpha(\sigma_X, \beta)}{2\Gamma(\beta-1)}, \quad \alpha(\sigma_X, \beta) = \sigma_X^{-1} \sqrt{\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}.$$

The BayesShrink method proceeds by soft thresholding each wavelet detail subband i using the threshold

$$\lambda_i^* = \frac{\hat{\sigma}^2}{\hat{\sigma}_{F,i}},$$

where the estimate of the signal standard deviation in subband i , $\hat{\sigma}_{F,i}$, is computed from

$$\hat{\sigma}_{F,i} = \sqrt{\max\{\hat{\sigma}_i^2 - \hat{\sigma}^2, 0\}}.$$

The BayesShrink algorithm requires an a priori estimate of the standard deviation of the noise, $\hat{\sigma}$. It is suggested in [3] that an estimate of $\hat{\sigma}$ based on the median of the wavelet coefficients for the first level HH_1 subband be used, i.e.,

$$\hat{\sigma} = \frac{\text{median}|HH_1|}{0.6745}. \quad (3)$$

This is a commonly used a priori noise level estimator in wavelet denoising methods; see [3, 10, 11]. The number of decomposition levels has to be specified by the user. In the numerical experiments of Section 4, the wavelet filters and decomposition level for BayesShrink were selected to match those used for RKMD.

3.3.2 Total variation denoising

Total variation (TV) image denoising methods determine a restoration \hat{F}^{TV} by minimizing a discretization of the functional

$$\hat{F}^{TV}(G; \lambda) = \arg \min_{\hat{F} \in \mathbb{R}^{m \times n}} \left\{ \|G - \hat{F}\|_F^2 + \lambda \int_{\Omega} |\nabla \hat{F}| \right\}$$

for some parameter $\lambda > 0$, where Ω is the rectangular image domain [18] and $\|\cdot\|_F$ denotes the Frobenius norm. Minimization often is carried out by applying an iterative method to the associated Euler–Lagrange equations, see [5, 9, 18], but other approaches also are possible [4]. For a discussion on the discretization of $\int_{\Omega} |\nabla \hat{F}|$, we refer to [15]. Like many wavelet denoising methods, TV denoising has been shown to be effective at smoothing noise in “flat” regions of a digital image, while preserving edge information. However, a value must be selected for the parameter λ in order to perform TV denoising. For a fair comparison with RKMD and BayesShrink, an implementation of TV denoising should be entirely data-driven. To this end, we computed a value $\lambda = \lambda^*$ that gave the smallest kurtosis of the residual image in the numerical examples. Thus, λ^* solves the minimization problem

$$\lambda^* = \arg \min_{\lambda} \left| \kappa_2 \left(G - \hat{F}^{TV}(G, \lambda) \right) \right|. \quad (4)$$

This is an extension of the RKMD methodology beyond the framework of wavelet coefficient thresholding.

4 Numerical examples

This section presents numerical examples that show the performance of the RKMD method when applied to images with different amount of noise-contamination. The use of different discrete wavelet transform filters is illustrated. The computed restorations determined by RKMD are compared to restorations obtained by BayesShrink and TV denoising. All examples have been implemented in the Python¹ programming language, making use of the Numpy², Scipy³, and Matplotlib⁴ packages. All discrete wavelet transforms were computed with the package PyWavelets⁵.

We used the *Kent* and *Barbara* images shown in Figure 3 in our experiments. They are made up of 512×512 pixels. Zero-mean Gaussian noise

¹www.python.org

²www.numpy.scipy.org

³www.scipy.org

⁴www.matplotlib.sourceforge.net

⁵www.pybytes.com/pywavelets/



Figure 3: The original noise-free *Kent* and *Barbara* images.

was added to both images for the noise levels 5%, 20%, and 35%, where the noise level is defined by

$$I_p = 100 \frac{\|F - E\|_F}{\|F\|_F}.$$

For both the RKMD and BayesShrink methods, the CDF 9/7 wavelet was used. This corresponds to a biorthogonal wavelet transform. The CDF 9/7 wavelet is composed of a symmetric wavelet and scaling filters, and is therefore able to accurately represent signals with symmetric features. This wavelet is used for one type of compression in the JPEG2000 digital image file format [6]. The tolerance parameter δ in Algorithm 1 was set to 10^{-8} .

The quality of the computed restorations is measured by the PSNR between the denoised image and the corresponding exact image. Additionally, we compute the relative error $err_\sigma(\hat{F})$ in the estimate of the exact noise variance σ^2 for each restoration \hat{F} . Thus,

$$err_\sigma(\hat{F}) = 100 \frac{|\sigma^2 - \hat{\sigma}^2(\hat{F})|}{\sigma^2}, \quad (5)$$

where $\hat{\sigma}^2(\hat{F})$ is the estimated noise variance computed from the restoration \hat{F} . This estimate was computed as

$$\hat{\sigma}^2(\hat{F}) = \frac{1}{mn} \|\hat{E} - \bar{\hat{E}}\|_F^2,$$

where $\hat{E} = G - \hat{F}$ is the residual image and $\bar{\hat{E}}$ denotes the mean of the pixel values of the residual image. Hence, the quantity (5) shows how accurately the error level was implicitly estimated by the denoising method.

Tables 1 and 2 display the PSNR and err_σ values obtained by the methods in our comparison for the *Barbara* and *Kent* images, respectively. Some of

Noise Level	PSNR(G)	TV		BayesShrink		Proposed	
		PSNR	err $_{\sigma}$	PSNR	err $_{\sigma}$	PSNR	err $_{\sigma}$
5%	31.3	32.6	89.3%	32.9	41.7%	33.0	8.9%
20%	29.4	30.6	77.5%	30.8	33.4%	31.1	5.3%
35%	27.1	28.7	51.7%	29.3	26.3%	29.1	3.0%

Table 1: Tabulation of PSNR and err $_{\sigma}$ values for denoised *Barbara* images.

Noise Level	PSNR(G)	TV		BayesShrink		Proposed	
		PSNR	err $_{\sigma}$	PSNR	err $_{\sigma}$	PSNR	err $_{\sigma}$
5%	35.7	37.3	130.2%	37.4	35.9%	37.7	7.2%
20%	32.5	33.1	64.3%	33.7	21.8%	34.2	4.6%
35%	29.3	30.5	14.6%	31.2	16.5%	31.4	2.1%

Table 2: Tabulation of PSNR and err $_{\sigma}$ values for denoised *Kent* images.

the restorations are shown in Figures 4 through 7. TV denoising consistently produced the piecewise smoothest results. Even at the highest noise level, the TV-denoised images contained no visual noise, while maintaining sharp edges. This is especially evident for the *Barbara* image at the 35% noise level (shown in Figure 7). However, in terms of PSNR and err $_{\sigma}$ values, the TV restorations generally were not the best. It may be that selection of the denoising parameter λ via residual kurtosis minimization tends to lead to over-smoothing for TV denoising. The estimated variances $\hat{\sigma}^2$ from the TV denoised images tended to be far larger than the actual noise variances σ^2 , leading to larger err $_{\sigma}$ values. However, if selecting a denoising algorithm based solely on getting a visually pleasing image, the TV denoising algorithm with λ determined by (4) might be considered the best choice.

Our method performed very similarly to the BayesShrink method in terms of visual quality, with slightly higher PSNR values for all but one image. In terms of err $_{\sigma}$ values, our method was by far the most accurate, consistently exhibiting relative errors less than 10% for the computed estimates of the noise variance. The number of wavelet and inverse wavelet transform computations required by our method was in all of our computed examples much smaller than the bound of Proposition 3.2; the number of computed transforms was less than 0.3% of this bound for every example. Nevertheless, BayesShrink is faster than our method, which requires about the same execution time as our implementation of TV denoising, typically a few seconds for each image in the computed examples. We remark that the execution time for Algorithm 1 depends on the value of the parameter δ , on the tolerance for the optimization method used, and on the choice of optimization method. It is likely that the execution time for our method can be decreased, without sacrificing the quality of the restoration significantly, by increasing δ and the tolerance used by the optimization method. The computed examples with the present implementation show our method to be the only one of the

compared methods that is competitive with regard to both PSNR and err_σ values across all images.

5 Conclusion and extension

A new technique for denoising digital images based on level-adaptive shrinkage of wavelet coefficients is presented. This method uses a novel criterion for determining wavelet coefficient thresholds by minimizing the fourth order kurtosis statistic of the computed residual image. This criterion is derived and justified within the framework of the noise contamination model, and does not require an a priori estimate of the noise variance. Numerical experiments demonstrate very good performance of the proposed RKMD technique. An application of the RKMD method to the restoration of images that have been contaminated by noise and blur is described in [12].

We remark that while wavelet denoising techniques are very powerful, the derived residual image kurtosis criterion does not explicitly require the use of wavelet coefficient shrinkage as the underlying denoising method. We illustrated this by adapting the residual kurtosis minimization methodology to TV denoising. The promising results suggest that extensions to non-wavelet-based denoising techniques are worthy of further study.

Acknowledgment

The authors would like to thank Giuseppe Rodriguez and an anonymous referee for comments that lead to improvements of the presentation.

References

- [1] C. Boncelet, Image Noise Models, In: Handbook of Image and Video Processing, ed. A. Bovik, Academic Press, New York, 2000, pp. 325–337.
- [2] A. Buades, B. Coll, and J. M. Morel, Image denosing methods. A new nonlocal principle, SIAM Review, 52 (2010), pp. 113–147.
- [3] S. G. Chang, Adaptive wavelet thresholding for image denoising and compression, IEEE Transactions on Image Processing, 9 (2000), pp. 1532–1546.
- [4] T. F. Chan, G. H. Golub, and P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, SIAM Journal on Scientific Computing, 20 (1999), pp. 1964–1977.



(a) Noisy

(b) Restored, TV



(c) Restored, BayesShrink

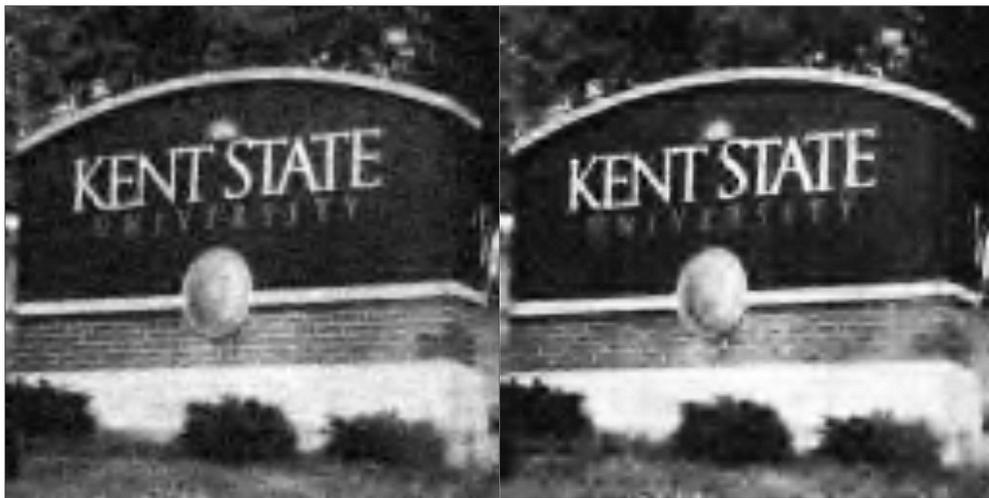
(d) Restored, Proposed method

Figure 4: The first row shows the *Kent* image contaminated with 20% noise, and a restoration using TV denoising. The second row shows restorations using BayesShrink and the proposed method.



(a) Noisy

(b) Restored, TV



(c) Restored, BayesShrink

(d) Restored, Proposed method

Figure 5: The first row shows the *Kent* image contaminated with 35% noise, and a restoration using TV denoising. The second row shows restorations using BayesShrink and the proposed method.



(a) Noisy

(b) Restored, TV



(c) Restored, BayesShrink

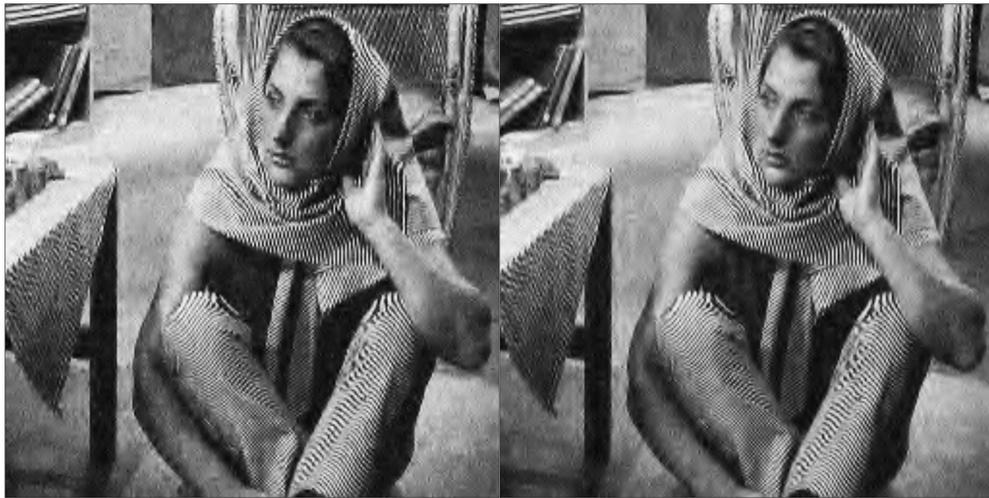
(d) Restored, Proposed method

Figure 6: The first row shows the *Barbara* image contaminated with 20% noise, and a restoration using TV denoising. The second row shows restorations using BayesShrink and the proposed method.



(a) Noisy

(b) Restored, TV



(c) Restored, BayesShrink

(d) Restored, Proposed method

Figure 7: The first row shows the *Barbara* image contaminated with 35% noise, and a restoration using TV denoising. The second row shows restorations using BayesShrink and the proposed method.

- [5] T. Chan, A. Marquina, and P. Mulet, High-order total variation-based image restoration, *SIAM Journal on Scientific Computing*, 22 (2000), pp. 503–516.
- [6] C. Christopoulos, A. Skodras, and T. Ebrahimi, The JPEG2000 still image coding system: an overview, *IEEE Transactions on Consumer Electronics*, 46 (2000), pp. 1103–1127.
- [7] R. D’Agostino and E. S. Pearson, Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, 60 (1973), pp. 613–622.
- [8] R. D’Agostino and G. L. Tietjen, Simulation probability points of b_2 for small samples, *Biometrika*, 58 (1971), pp. 669–672.
- [9] D. Dobson and C. Vogel, Convergence of an iterative method for total variation denoising, *SIAM Journal on Numerical Analysis*, 34 (1997), pp. 1779–1791.
- [10] D. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, 90 (1995), pp. 1200–1224.
- [11] D. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory*, 41 (1995), pp. 613–627.
- [12] T. A. Hearn and L. Reichel, Application of denoising methods to regularization of ill-posed problems, *Numerical Algorithms*, 66 (2014), pp. 761–777.
- [13] A. Hyvärinen, *Independent Component Analysis, Adaptive and Learning Systems for Signal Processing*, Wiley-Blackwell, 2001.
- [14] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed., Elsevier, Amsterdam, 2009.
- [15] L. Moisan, How to discretize the total variation of an image?, *Proceedings in Applied Mathematics and Mechanics*, 7 (2007), pp. 1041907–1041908.
- [16] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, Jr., Survey of image denoising techniques, *Proceedings of the Global Signal Processing Exposition and Conference*, 2004.
- [17] E. S. Pearson, Tables for percentage points of $\sqrt{b_1}$ and b_2 in normal samples: a rounding off, *Biometrika*, 52 (1965), pp. 282–287.
- [18] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, 60 (1992), pp. 259–268.