Wavelet-Domain Filtering for Photon Imaging Systems

Robert D. Nowak

Department of Electrical Engineering
Michigan State University
East Lansing, MI 48824–1226
Fax: (517) 353–1980
Email: rnowak@egr.msu.edu
Web: http://www.egr.msu.edu/spc/

*Richard G. Baraniuk**

Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005–1892
Fax: (713) 524–5237
Email: richb@rice.edu
Web: http://www-dsp.rice.edu

Submitted to *IEEE Transactions on Image Processing*

April 1997

EDICS Numbers: IP 1.6 Multiresolution Processing, IP 1.2 Filtering

Abstract — Many imaging systems rely on photon detection as the basis of image formation. One of the major sources of error in these systems is Poisson noise due to the quantum nature of the photon detection process. Unlike additive Gaussian noise, Poisson noise is signal-dependent, and consequently separating signal from noise is a very difficult task. In this paper, we develop a novel wavelet-domain filtering procedure for noise removal in photon imaging systems. The filter adapts to both the signal *and* the noise and balances the trade-off between noise removal and excessive smoothing of image details. Designed using the statistical method of cross-validation, the filter is simultaneously optimal in a small-sample predictive sum of squares sense and asymptotically optimal in the mean square error sense. The filtering procedure has a simple interpretation as a joint edge detection/estimation process. Moreover, we derive an efficient algorithm for performing the filtering that has the same order of complexity as the fast wavelet transform itself. The performance of the new filter is assessed with simulated data experiments and tested with actual nuclear medicine imagery.

---

*This work was supported by the National Science Foundation, grant no. MIP–9457438, and the Office of Naval Research, grant no. N00014–95–1–0849.

Permission to publish abstract separately is granted.

I. Introduction

Photon detection lies at the heart of the image formation process in a host of applications, including medical imaging [1] and astronomical and low-light imaging [2]. A photon imaging system acquires images by counting photon detections at different spatial locations over an observation period of T seconds. For example, photon events are detected and located using a photomultiplier tube array. The main source of noise in photon imaging is the so-called "quantum" noise due to the discrete nature of the photon detection process [3]. Quantum noise degrades images in both qualitative and quantitative senses and hinders image analysis, interpretation, and feature extraction. Consequently, it is desirable to develop filtering methods to remove this noise.

Quantum noise obeys a Poisson law and hence is highly dependent on the underlying light intensity pattern being imaged. The basic model for photon imaging data is

$$f(k,l) \;=\; s(k,l) \;+\; \eta(k,l), \tag{1.1}$$

with $f(k,l)$ the integer number of photons detected at the k,l-th pixel location. The image formed from these "counts" is the standard image used in practice. In (1.1), we decompose this total count image into a signal and noise component. The signal $s(k,l)$ denotes the underlying light intensity at the k,l-th pixel location. The noise $\eta(k,l)$ denotes the difference between the true intensity and the number of counts detected. The probability distribution of $f(k,l)$ is Poisson with intensity $s(k,l)$. Since the mean value of this distribution is $s(k,l)$, the noise $\eta(k,l)$ represents the variability of the counts about the true mean intensity. It is well-known that the variance of a Poisson random variable is equal to the mean [4]. Therefore the variability of the noise $\eta(k,l)$ is proportional to the intensity $s(k,l)$ and hence *signal-dependent* [5].

This signal dependence has frustrated traditional attempts at noise removal in photon imagery. Not only must we contend with the usual problems associated with noise reduction, such as over-smoothing, but we also face the additional complexity of a spatially varying noise level that depends on the local intensity in the image. It is easily shown that the signal-to-noise ratio (SNR) for $f(k,l)$ is linear in $s(k,l)$, and hence any attempts at noise removal on a pixel-by-pixel basis are futile. The only way to improve the SNR is to increase the intensity (or equivalently acquire the image over a longer observation period). Clearly this could be prohibitive in many applications. Fig. 1 shows a simulated image with Poisson noise.

Rather than attempting to separate signal from noise in the spatial (pixel) domain, it is often advantageous to work in a transform domain. The motivation for transform-domain processing

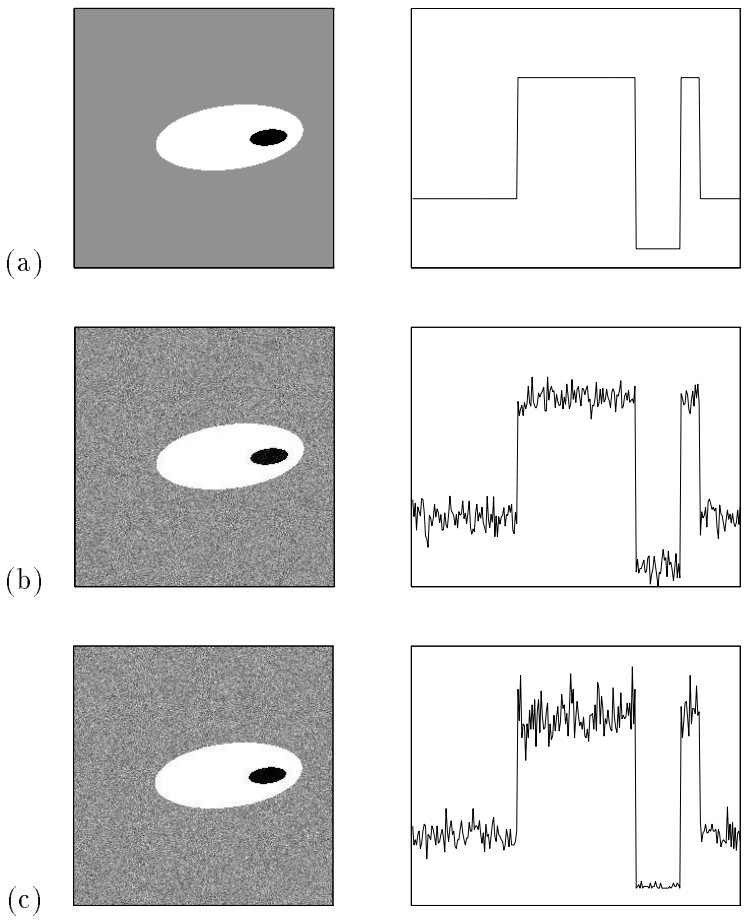

Fig. 1. *Simulated images with additive white Gaussian noise and Poisson noise. (a) Simulated, noise-free image (left) and horizontal profile through center (right). (b) Image with additive white Gaussian noise. (c) Image with Poisson noise. While the Gaussian noise level is signal-independent and hence uniform across the image, the Poisson noise level is higher in the brighter regions and lower in the darker regions. The noise levels in (b) and (c) have been adjusted so that the background noise level is comparable in both cases. The Poisson noise levels shown here are typical of many applications (nuclear medicine imaging, for example). The signal-dependent nature of Poisson noise makes the noise removal task much more difficult than in the Gaussian case.*

is that an appropriate image transform can concentrate the signal energy into a small number of coefficients. Hence, in the transform domain we will have a few coefficients with high SNR (which we keep) and a large number of coefficients with low SNR (which we remove). In this way we can effectively separate (filter) the signal from the noise. Two standard choices for the image transform are the Fourier transform and the wavelet transform.

The wavelet transform performs a local Fourier analysis by representing images in terms of a basis of functions localized in both space and frequency. The localization of the wavelet functions is particularly appropriate for imaging applications, where it is crucial to preserve fine details like edges [6]. Furthermore, empirical evidence has shown that wavelet bases generally provide more efficient representations of real-world data than spatial or frequency domain representations [7]. Because wavelets are able to concisely represent complicated signal structure, filtering techniques based in the wavelet domain perform much better at separating signals from noise than classical approaches based in the spatial or frequency domain.

Research to date in wavelet-domain filtering has focused on the removal of additive white Gaussian noise. The basic wavelet-domain filtering procedure involves (1) computing the discrete wavelet transform (DWT) of the image, (2) zeroing out the small wavelet coefficients that fall below the noise level, (3) computing the inverse DWT of the processed data [6]. If the noise is additive white Gaussian, then the noise level is uniform throughout the image and hence uniform across all wavelet coefficients, since the DWT is orthonormal. Therefore, in additive white Gaussian noise a simple global noise threshold can be determined independent of the signal.

Unfortunately, the Poisson noise that characterizes photon imagery is signal-dependent, and therefore wavelet-domain filtering based on a global threshold is inappropriate. One simple fix would be to work with the square-root of the image, since the square-root operation is a variance stabilizing transformation [8]. As the number of photon counts tends to infinity, the square-root image noise does tend to a white Gaussian noise. However, in many practical applications, such as nuclear medicine and astronomy, far too few counts are detected in order to reach such asymptotic conditions. Consequently, the square-root image is still contaminated with a signal-dependent noise. Thus, there is a need for an optimal filtering procedure for Poisson noise that adapts to the local SNR in each region of the image.

The purpose of this paper is to develop and analyze a new method for designing optimal wavelet-domain filters for noise removal in photon imagery. Using the results of [9], we derive a novel wavelet-domain Wiener filter based on the method of cross-validation that not only preserves key

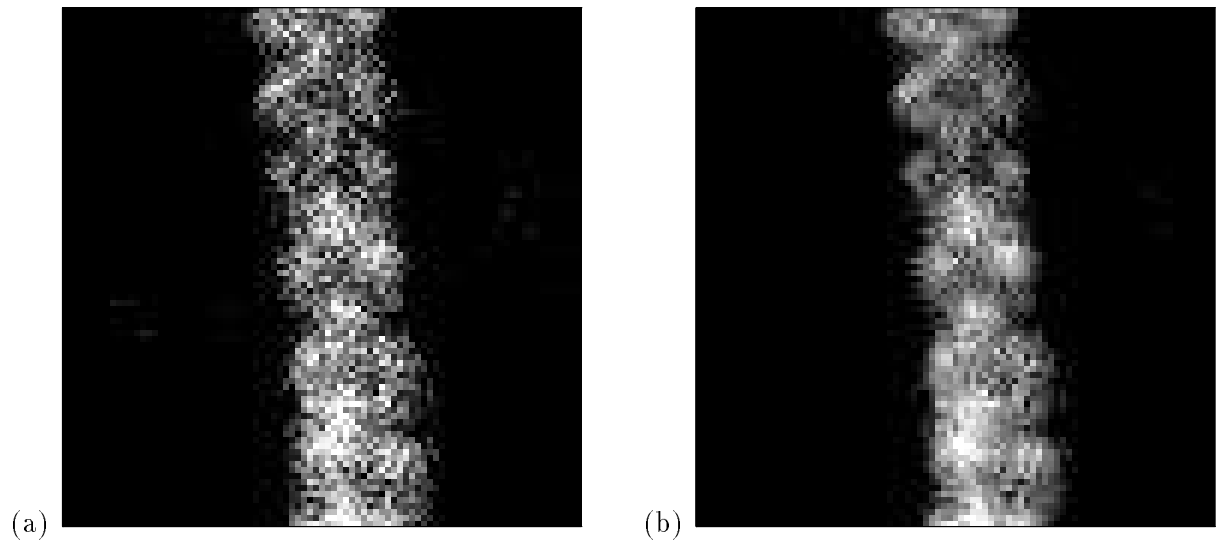

Fig. 2. *Comparison of nuclear medicine spine images. (a) Raw-count image (b) wavelet-domain filtered image using Daubechies-6 wavelet basis. The wavelet-domain filtering dramatically reduces the noise, but does not excessively smooth the fine structure of spinal column.*

image details and features, but also adapts to the local noise level of the spatially varying Poisson process underlying the image.

Figure 2 demonstrates the new filter's ability to reduce Poisson noise in nuclear medicine imaging, a widely used clinical imaging modality. Unlike many other medical imaging techniques, nuclear medicine imaging can provide both anatomical *and* functional information. However, nuclear medicine imaging has a much lower SNR compared with other techniques. Hence, improvements in image quality via optimized signal processing represent a significant opportunity to advance the state-of-the-art in nuclear medicine. The raw nuclear medicine data is an image of photon detections or counts (see Figure 2(a)). Because of the relatively small number of counts in typical nuclear medicine studies, the raw count image appears noisy and in fact has a large variance about its true mean (the underlying intensity of the Poisson process). The new wavelet-domain filtering method developed in this paper adapts to both the signal *and* noise and provides a much better estimate of the underlying process (see Figure 2(b)).

After reviewing the DWT in Section II and wavelet-domain filtering in Section III, we develop the new wavelet-domain filter in Section IV. In Section V we study the optimal filter in further detail and show that it has a simple interpretation in the spatial domain as a joint detection/estimation procedure. We derive a fast algorithm for performing the entire optimal wavelet-domain filtering procedure whose order of complexity is equivalent to that of the fast wavelet transform in Section

VI. In Section VII we study the performance of the new filter through simulations and and actual data from nuclear medicine. Finally, we offer some conclusions and indicate directions of ongoing and future work in Section VIII.

II. The Wavelet Transform

The wavelet transform represents a 1-d real-valued continuous-time signal $f(t)$, $t \in \mathbb{R}$ in terms of shifts and dilations of a lowpass *scaling function* $\phi(t)$ and bandpass *wavelet* $\psi(t)$. For special choices of these functions, the shifts and dilations form an orthonormal basis for $L_2(\mathbb{R})$ and we have the signal representation [10, 11, 12]

$$f(t) \;=\; \sum_n c_n^J \, 2^{-J/2} \, \phi(2^{-J}t - n) \;+\; \sum_{j=-\infty}^{J} \sum_n d_n^j \, 2^{-j/2} \, \psi(2^{-j}t - n) \tag{2.1}$$

with *scaling coefficients*

$$c_n^j \;=\; 2^{-j/2} \int f(t) \, \phi(2^{-j}t - n) \, dt \tag{2.2}$$

and *wavelet coefficients*

$$d_n^j \;=\; 2^{-j/2} \int f(t) \, \psi(2^{-j}t - n) \, dt. \tag{2.3}$$

A key feature of the wavelet representation is multiresolution: the first term in (2.1) is an approximation to $f(t)$ at scale (resolution) J, while the second term consists of refinements at finer and finer scales $j \leq J$.

The wavelet transforms of real-world signals and images tend to be very sparse, with a few large scaling and wavelet coefficients dominating the representation. From a theoretical perspective, it has been shown that this property stems from the fact that wavelets form *unconditional bases* for wide classes of signals [13]. Wavelet bases are unconditional for many important and practical signal spaces including Besov, Sobolev, and Treibel spaces. Roughly speaking, Besov spaces include signals that are generally smooth except for some possible points of discontinuity. The Besov norm measures signal "smoothness" quite differently from conventional frequency domain notions of smoothness, by allowing local signal discontinuities. Besov spaces are therefore very appropriate for dealing with real-world signals like images.

The scaling and wavelet coefficients can be computed using an extremely efficient recursion. Using the fact that the scaling function and wavelet are related by the two-scale difference equations [10, 11, 12]

$$\phi(x/2) \;=\; \sqrt{2} \sum_n h_n \, \phi(x - n) \tag{2.4}$$

$$\psi(x/2) \;=\; \sqrt{2}\sum_n g_n\,\phi(x-n), \tag{2.5}$$

with h_n and g_n discrete-time lowpass and highpass filters, respectively, the scaling and wavelet coefficients can be recursed as

$$c_m^{j+1} \;=\; \sum_n h_{2m-n}\,c_n^j \tag{2.6}$$

$$d_m^{j+1} \;=\; \sum_n g_{2m-n}\,c_n^j. \tag{2.7}$$

These computations can be succinctly organized into a discrete-time filter bank comprising lowpass and highpass filters and decimators [10, 11, 12].

For a discrete-time signal $f(k)$, $k \in \mathbb{Z}$, the recursion (2.6)–(2.7) with $c_k^0 = f(k)$ defines a discrete-time wavelet decomposition. The corresponding discrete-time wavelet basis elements $w_{j,n}(k)$ can be obtained in terms of products of matrices involving h_n and g_n [11, pp. 143–145]. By organizing N_f signal samples into the vector $\mathbf{f} = \Big[f(1), f(2), \ldots, f(n)\Big]^T$ and the basis elements into an $N_f \times N_f$ matrix $\mathbf{W}$, we can write the discrete-time wavelet transform as $\boldsymbol{\theta} = \mathbf{W}\mathbf{f}$. By analogy to (2.1), we will refer to the elements of $\boldsymbol{\theta}$ as the wavelet coefficients. Because of the special filter bank structure of the wavelet transform, the matrix multiplication $\mathbf{W}\mathbf{f}$ can be implemented in $O(N_h N_f)$ computations for N_h-point filters h_n and g_n. Remarkably, this cost is linear in the number of signal samples processed.

Images can be decomposed using separable 2-d wavelet bases, which are typically implemented using a 2-d filter bank where we alternate the application of the h and g filters on rows and columns of the image $f(k,l)$ [11, 12]. The result is a basis of 2-d wavelets $w_{i,j,m,n}(k,l)$ having four indices. In order to keep notation to a minimum in the sequel, we will employ the abstract index I for these four indices; hence $I \triangleq (i,j,m,n)$ and

$$\theta_I \;=\; \sum_{k,l} w_I(k,l)\,f(k,l). \tag{2.8}$$

By vectorizing the image $f(k,l)$, we can continue to use the notation $\boldsymbol{\theta} = \mathbf{W}\mathbf{f}$ for the wavelet transform, with $\boldsymbol{\theta} = \{\theta_I\}$ and $\mathbf{W} = \{w_I(k,l)\}$. Using a 2-d filter bank, the computational cost of this transform is $O(N_h N_f^2)$ for an $N_f \times N_f$ image.

III. Wavelet-Domain Filtering

Recall our goal of suppressing the signal-dependent noise term $\eta(k,l)$ in the photon imaging model of (1.1). We choose to work in the wavelet domain, because the wavelet transform will tend concentrate the energy of the desired signal $s(k,l)$ into a small number of (necessarily large) coefficients. Hence, the wavelet transform of the noisy image will consist of a small number of

coefficients with high SNR (which we will keep) and a large number of coefficients with low SNR (which we will remove). In this way we can filter the signal from the noise. Of course, this same procedure could be carried out using any orthonormal signal representation, including the Fourier transform. However, the wavelet transform enables the filtering procedure to adapt to the local variations in the signal frequency content and thereby balances the trade-off between noise removal and excessive smoothing. Fourier-domain filtering is a global operation that cannot adjust to such local variations and hence leads to excessive smoothing in regions where the image has high frequency content (edges, for example).

A. Filtering setup

In general, wavelet-domain filtering involves by the following procedure. First, we compute the 2-d wavelet transform of the photon image: $\boldsymbol{\theta} = \mathbf{W}\mathbf{f}$. With $\boldsymbol{\theta}$ in hand, we can filter (or attenuate) the contribution of a particular wavelet basis function w_I to the signal expansion by weighting the corresponding coefficient θ_I by a number $0 \leq h_I \leq 1$. That is, we modify $\boldsymbol{\theta}$ component-wise according to

$$\widetilde{\theta}_I \;=\; h_I \theta_I. \tag{3.1}$$

Setting $h_I = 0$ completely removes the contribution of the basis function w_I; setting $h_I = 1$ leaves it unaltered. Choosing $0 < h_I < 1$ attenuates the contribution of the I-th wavelet basis accordingly. Finally, we invert the modified wavelet transform to obtain our signal estimate

$$\widetilde{\mathbf{f}} \;=\; \mathbf{W}^{-1}\widetilde{\boldsymbol{\theta}}. \tag{3.2}$$

The new image $\widetilde{\mathbf{f}}$ is called a *wavelet-domain filtered* version of $\mathbf{f}$. The collection of weights $\mathbf{h} = \{h_I\}$ is the *wavelet-domain filter*.

The ability of the wavelet transform to represent signal features that are localized in both space and frequency enables even a simple filter $\mathbf{h}$ to perform adaptive smoothing that adjusts to the local image structure. Moreover, the unconditional nature of the wavelet basis is crucial to wavelet-domain filtering, because it guarantees that filtering always tends to smooth a signal, in the sense of the Besov norm [6]. This guarantees that excessive artifacts are not introduced by the filtering procedure. In contrast, filtering in the frequency domain can produce dangerous artifacts such as Gibbs phenomena.

Clearly, the crucial issue in wavelet-domain filtering is the design of the filter $\mathbf{h}$. Before considering the Poisson noise case, let us consider the simpler additive white Gaussian noise case.

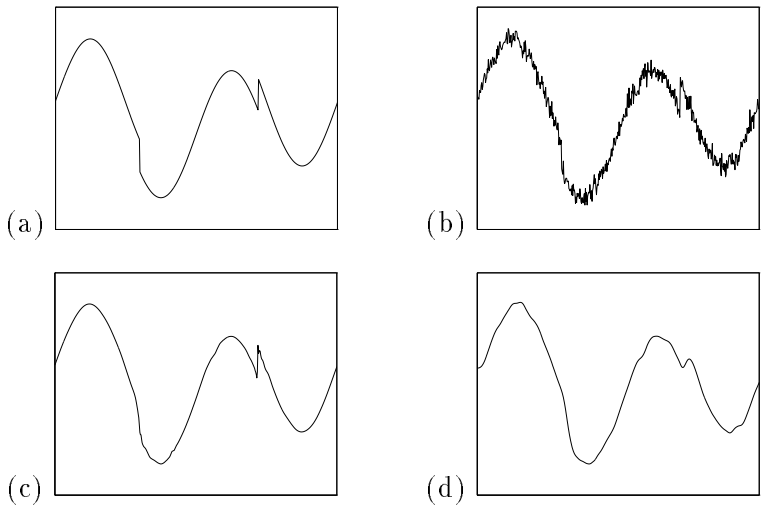

Fig. 3. *Wavelet Shrinkage of a noisy 1-d "HeaviSine" signal from [6]. (a) Noise-free signal. (b) Signal plus white Gaussian noise. (c) Denoised signal obtained using Wavelet Shrinkage. (d) For comparison, linear filtering with a Butterworth lowpass filter (frequency-domain shrinkage). Note that the nonlinear wavelet soft-thresholding both removes noise and preserves the sharp edges in the signal. Classical linear filtering also removes noise but smears the edges.*

B. Filtering additive white Gaussian noise: Wavelet Shrinkage

The most popular form of wavelet-based filtering is commonly known as *Wavelet Shrinkage* [6]. In the Wavelet Shrinkage programme, the h_I are determined by the wavelet coefficients themselves. The Wavelet Shrinkage filter corresponds to the "soft threshold" nonlinearity

$$h_I = \begin{cases} 1 - \frac{\tau \, \mathrm{sgn}(w_I)}{w_I}, & \text{if } |w_I| \geq \tau \\ 0, & \text{if } |w_I| < \tau, \end{cases} \tag{3.3}$$

with τ a user-specified threshold (usually related to the known or estimated global noise variance). The soft-threshold sets very small coefficients to zero and reduces all other coefficients by a fixed amount τ.

The important thing to note about Wavelet Shrinkage is that the filter performs an identical nonlinear transformation on each wavelet coefficient. If the noise in the image is white Gaussian, then the noise power is the same in each wavelet coefficient, and this procedure is asymptotically minimax optimal [6]. A comparison between Wavelet Shrinkage and Fourier-domain filtering is given in Fig. 3.

C. Filtering Poisson noise

In the case of Poisson noise, the noise power will differ between wavelet coefficients according to the image intensity under the support of the associated wavelet basis function. This spatial variation of the noise must be accounted for in the wavelet-domain filter design. Wavelet Shrinkage does not adjust to these differences. Given the signal and noise powers, a natural choice for a wavelet-domain filter is

$$h_I \approx \begin{cases} 1, & \text{if SNR in } w_I \text{ is high} \\ 0, & \text{if SNR in } w_I \text{ is low.} \end{cases} \tag{3.4}$$

Next, we describe a design procedure that results in an optimal filter of this form.

IV. Filter Design via Cross-Validation

Recall that a photon imaging system acquires images by counting photon detections at different spatial locations (pixels) over an observation period of T seconds. The most common estimate of the underlying intensity is the total-count image — the sum of photon detections at each pixel.

A. Filter derivation

Rather than working with the total-count image, we subdivide the observation period into N subintervals of T/N seconds each. We then acquire N low-count images in each subinterval, denoted by $\mathbf{f}^{(1)}, \ldots, \mathbf{f}^{(N)}$. The low-count images are acquired by counting the photon detections in each disjoint time interval of T/N seconds. We assume that the underlying intensity does not vary over the total observation period T, which implies that the low-count images are independent and identically distributed (iid). The total-count image is related to the low-count images by

$$\mathbf{f} = \sum_{i=1}^{N} \mathbf{f}^{(i)}. \tag{4.1}$$

We will use the low-count images to assess the consistency of the data in the wavelet domain, which in turn will lead to an optimal filtering procedure.

We use the method of cross-validation on the N low-count images to design an optimal wavelet-domain filter. First, compute an image using all but the k-th low-count images:

$$\mathbf{f}^{(/k)} \triangleq \sum_{i \neq k} \mathbf{f}^{(i)}. \tag{4.2}$$

The image $\mathbf{f}^{(/k)}$ is called a "leaving-one-out" image. Next, compute the wavelet-domain filtered version of $\mathbf{f}^{(/k)}$ and denote the result by $\widetilde{\mathbf{f}}^{(/k)}$. Note that the image $\widetilde{\mathbf{f}}^{(/k)}$ depends on the wavelet-domain filter coefficients $\mathbf{h}$.

The goal of filtering is to estimate the underlying intensity distribution. If the filter is good, then $\widetilde{\mathbf{f}}^{(/k)}$ should be a good estimate of the intensity and likewise should be a good predictor of the counts we have detected. The method of cross-validation assesses how well $\widetilde{\mathbf{f}}^{(/k)}$ predicts the low-count image that we have not used, namely $\mathbf{f}^{(k)}$. We measure the prediction error by

$$e^{(k)}(\mathbf{h}) \triangleq \left\| \frac{1}{N-1}\widetilde{\mathbf{f}}^{(/k)} - \mathbf{f}^{(k)} \right\|_F^2, \tag{4.3}$$

with $\|\cdot\|_F^2$ the sum of the squared errors, pixel by pixel. The error is a function of the wavelet-domain filter coefficients $\mathbf{h}$. We repeat this process for $k = 1, 2, \ldots, N$ and consider the *predictive sum of squares* (PRESS)

$$V(\mathbf{h}) \triangleq \sum_{k=1}^{N} e^{(k)}(\mathbf{h}). \tag{4.4}$$

The PRESS may be viewed as a small-sample optimality criterion measuring the quality of the estimate obtained using the wavelet-domain filter coefficients $\mathbf{h}$. Our objective is to choose $\mathbf{h}$ to minimize $V(\mathbf{h})$.

In [9], it is shown that the optimal $\mathbf{h}$, minimizing the PRESS, is given by[1]

$$h_I^{(\mathrm{PRESS})} = \left(\frac{\theta_I^2 - \widehat{\sigma_I^2}}{\theta_I^2 + \frac{1}{N-1}\widehat{\sigma_I^2}} \right)_+, \tag{4.5}$$

with θ_I^2 the square of the I-th wavelet coefficient of the total-count image, $\widehat{\sigma_I^2}$ an estimate of the variance in the I-th wavelet coefficient, and $(\cdot)_+$ denoting the positive part (negative values set to zero). The variance is estimated from the N low-count images according to

$$\widehat{\sigma_I^2} = \frac{N}{N-1}\sum_{k=1}^{N}\left(\theta_I^{(k)} - \frac{1}{N}\theta_I\right)^2, \tag{4.6}$$

with $\theta_I^{(k)}$ the I-th wavelet coefficient of the low count image $\mathbf{f}^{(k)}$.

We call $\mathbf{h}^{(\mathrm{PRESS})}$ the *PRESS-optimal wavelet domain filter*, and the call the resulting filtered image, denoted $\widetilde{\mathbf{f}}^{(\mathrm{PRESS})}$, the *PRESS-optimal estimator*.

Several comments are in order. First, in addition to minimizing the PRESS, the wavelet-domain filter $h_I^{(\mathrm{PRESS})}$ is asymptotically optimal in the mean-square error (MSE) sense. This result follows from a straightforward application of the Strong Law of Large Numbers [9]. This implies that as the total number of counts increases, $h_I^{(\mathrm{PRESS})}$ tends to the wavelet-domain filter that minimizes the

[1] Note that the notation in [9] differs slightly from the notation here. Here $h_I^{(\mathrm{PRESS})}$ is related to the regularization parameter γ_I in [9] by $h_I^{(\mathrm{PRESS})} = 1 - \gamma_I^{(\mathrm{PRESS})}$. Also note that θ_I^2 and $\widehat{\sigma_I^2}$ are scaled by N^{-1} in [9].

MSE, a wavelet-domain analog of the classical Wiener filter. Secondly, inspection of (4.5) shows that $h_I^{(\mathrm{PRESS})}$ is zero when the estimated SNR

$$\widehat{SNR} \;=\; \frac{(N-1)^2\,\theta_I^2}{N^2\,\widehat{\sigma_I^2}} \tag{4.7}$$

is less than 1. Furthermore, $h_I^{(\mathrm{PRESS})}$ tends to unity as the estimated SNR increases. Thirdly, the complexity of the optimal filter design is roughly N times that of Wavelet Shrinkage, because we must compute the wavelet transform of each low-count image.

Finally, note that we require $N > 1$ low-count images to compute the optimal wavelet-domain filter. The PRESS criterion measures how well the filtered image predicts the observed counts. That is, the PRESS-optimal estimator supports data that is consistent across all low count images, and rejects data that is inconsistent. Intuitively, we could expect that better results would be obtained by refining the subdivision of the low-count images, since this would better enable us to assess the consistency of the data. This suggests that best results will be obtained by subdividing T as finely as possible. This clearly becomes computationally infeasible. However, in the next section it is shown that the limiting PRESS-optimal filter (as $N \to \infty$) can be computed directly from the total-count image — eliminating the need for low-count images altogether!

B. Limiting PRESS-optimal filter as $N \to \infty$

Based on the results in [9], algebraic machinations show that the value of the PRESS is given by

$$V(\mathbf{h}) \;=\; \left(\frac{N}{N-1}\right) \sum_I \left[\, \widehat{\sigma_I^2} - 2(1-h_I)N^{-1}\widehat{\sigma_I^2} + (1-h_I)^2 N^{-2}\left(\widehat{\sigma_I^2} + (N-1)\theta_I^2\right)\right]. \tag{4.8}$$

Set $m_I = \mathrm{E}[\theta_I]$ and $\sigma_I^2 = \mathrm{E}[\widehat{\sigma_I^2}]$. Using (4.8) plus the fact $\mathrm{E}[\theta_I^2] = m_I^2 + \sigma_I^2$, we find the expected value of the PRESS is

$$\mathrm{E}\,[V(\mathbf{h})] \;=\; \sum_I (1-h_I)^2 \frac{m_I^2}{N} + h_I^2 \frac{\sigma_I^2}{N-1} + \sigma_I^2. \tag{4.9}$$

To find the PRESS-optimal filter coefficient h_I, we take the partial derivative of the PRESS with respect to h_I. This partial derivative equals, on average,

$$\frac{\partial \mathrm{E}\,[V(\mathbf{h})]}{\partial h_I} \;=\; 2(1-h_I)\frac{m_I^2}{N} + 2h_I \frac{\sigma_I^2}{N-1}. \tag{4.10}$$

Now consider the partial derivative of the MSE $= \mathrm{E}\,[\sum_I (h_I\theta_I - m_I)^2]$ with respect to h_I

$$\frac{\partial \mathrm{E}\,[\sum_I (h_I\theta_I - m_I)^2]}{\partial h_I} \;=\; 2(1-h_I)m_I^2 + 2h_I\sigma_I^2. \tag{4.11}$$

Note that (4.10) and (4.11) are nearly identical; after scaling (4.10) by N we have

$$\frac{\partial N \mathrm{E}\left[V(\mathbf{h})\right]}{\partial h_I} \;=\; 2(1-h_I)m_I^2 + 2h_I\left(\frac{N}{N-1}\right)\sigma_I^2. \tag{4.12}$$

Furthermore, (4.12) converges to (4.11) as $N \to \infty$. This shows that on average the minimum of the PRESS tends to the minimum of the MSE. Hence, this suggests that the PRESS-optimal estimator has the best possible MSE performance in the limiting case.

The following theorem studies the limiting case ($N \to \infty$) of the PRESS-optimal filter. For the proof, see the Appendix.

Theorem 4.1: In the limit as $N \to \infty$, the PRESS-optimal filter takes the form

$$\lim_{N\to\infty} h_I^{(\mathrm{PRESS})} \;=\; \left(\frac{\widehat{\theta_I^2}}{\widehat{\theta_I^2}+\widehat{\sigma_I^2}}\right)_+, \tag{4.13}$$

with

$$\widehat{\sigma_I^2} \;=\; \sum_{k,l} w_I^2(k,l) f(k,l) \tag{4.14}$$

an unbiased estimate of the noise power in the I-th wavelet coefficient, and

$$\widehat{\theta_I^2} \;=\; \theta_I^2 - \widehat{\sigma_I^2}, \tag{4.15}$$

an unbiased estimate of the signal power in the I-th wavelet coefficient.

Note that if $\lambda(k,l)$ denotes the intensity of the Poisson process at the k,l-th pixel, then

$$\mathrm{E}\left[\widehat{\theta_I^2}\right] \;=\; \left(\sum_{k,l} w_I(k,l)\lambda(k,l)\right)^2 \tag{4.16}$$

and

$$\mathrm{E}\left[\widehat{\sigma_I^2}\right] \;=\; \sum_{k,l} w_I^2(k,l)\lambda(k,l) \tag{4.17}$$

do indeed furnish unbiased estimates of the signal and noise power in I-th wavelet coefficient.

V. Analysis and Interpretation of the PRESS-Optimal Filter

A. Wiener filters and shrinkage estimators

The limiting PRESS-optimal filter

$$h_I^{(\mathrm{PRESS})} = \left(\frac{\widehat{\theta_I^2}}{\widehat{\theta_I^2}+\widehat{\sigma_I^2}}\right)_+ \tag{5.1}$$

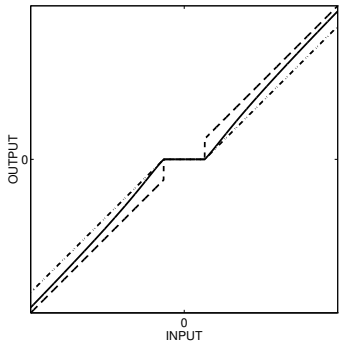

Fig. 4. *Comparison of threshold nonlinearities for nonlinear wavelet-domain filtering. Soft-threshold of Wavelet Shrinkage* (3.3) *(dot-dash). Hard-threshold nonlinearity (dash-dash). PRESS-optimal nonlinearity (solid). The soft and hard thresholds are commonly used for wavelet-domain filtering. The threshold levels are usually set based on the known or estimated noise level, which is assumed to be identical in each coefficient (additive, Gaussian white noise contamination, for example) [6]. In contrast, the threshold of the PRESS-optimal nonlinearity depends on the local noise level in each individual wavelet coefficient. Hence, the PRESS-optimal filter can adjust to variations in the Poisson noise level due to variations in the local intensity of the image. Furthermore, note that the PRESS-optimal nonlinearity is in a sense a compromise between the hard and soft thresholds — it preserves the continuity of the soft-threshold, yet provides an unbiased estimate at high SNR like the hard-threshold.*

simply weights each noisy wavelet coefficient θ_I by a factor equal to the estimated signal power divided by the estimated signal-plus-noise power. The $(\cdot)_+$ operation thresholds the weight to zero if the estimated signal to signal-plus-noise ratio is negative.[2] Hence, the PRESS-optimal filter has a very simple interpretation as a data-adaptive wavelet-domain Wiener filter.

Also note that the filter weight $0 \leq h_I^{(\mathrm{PRESS})} \leq 1$. Hence, the PRESS-optimal filtering operation is a form of shrinkage estimator [14]: the PRESS-optimal filter "shrinks" the noisy wavelet coefficient towards the origin according to the estimated signal to signal-plus-noise ratio. In fact, we can interpret the PRESS-optimal filtering operation as applying an SNR-adaptive nonlinearity to the noisy wavelet coefficients. The nonlinearity is depicted in Fig. 4.

[2]Since the signal is being estimated from the data, it is possible for the estimate to take on negative values. The thresholding operation arises naturally in the minimization of the PRESS criterion and guards against this unreasonable situation. Moreover, it can be shown that if we weight each noisy wavelet coefficient by the estimated signal to signal-plus-noise ratio without thresholding, then the resulting wavelet coefficient estimate has a worse MSE performance than the PRESS-optimal estimate. This follows from Theorem 6.2 in [14].

B. Joint edge detection/estimation

PRESS-optimal filtering can also be interpreted as a joint edge detection/estimation procedure. To begin with, consider the PRESS-optimal filter for the Haar wavelet basis [11, 12]. The Haar wavelet coefficient θ_I is the value of the inner product of the image and the wavelet $w_I(k,l)$, which is a localized edge (see Fig. 5(a)). The noise power estimate $\widehat{\sigma_I^2}$, equals the projection onto the *square* of the wavelet $w_I^2(k,l)$, which yields a local average. In other words, $\widehat{\sigma_I^2}$ is proportional to an estimate of the local intensity of the image in the region falling under the support of the Haar wavelet w_I. Expressing the PRESS-optimal filter directly (5.1) in terms of $\theta_I^2(k,l)$ and $\widehat{\sigma_I^2}$, we have

$$h_I^{(\mathrm{PRESS})} = \left(\frac{\theta_I^2 - \widehat{\sigma_I^2}}{\theta_I^2} \right)_+ \tag{5.2}$$

The action of the filter coefficient $h_I^{(\mathrm{PRESS})}$ can thus be interpreted as follows: If the strength of the local edge relative to the local intensity level falls below the threshold, then the filter does not detect the edge and sets the corresponding wavelet coefficient to zero. If the strength of the local edge relative to the local intensity level is above the threshold, then the filter detects the edge and proceeds with the estimation of the wavelet coefficient.

A similar interpretation holds for wavelets in general. The wavelet coefficients are the projections of the image onto localized "details" (see Fig. 5(b)). For the noise power estimate, we project the image onto the square of the corresponding wavelet basis function, which effectively computes a weighted average of the local intensity in the image. If the strength of the local detail relative to the weighted local intensity lies about the threshold, then the detail is detected and the filter estimates the coefficient's value. PRESS-optimal wavelet-domain filters are thus *spatially adaptive*.

Similar interpretations hold for PRESS-optimal filters derived in terms of arbitrary signal expansions (wavelet packets, local cosine bases, Karhunen-Loeve expansion, and even overcomplete expansions). In particular, consider the PRESS-optimal Fourier-domain filter. The Fourier coefficients are the projections of the image onto *global* complex sinusoids — global referring to the fact that the Fourier basis functions have support over the entire image (see Fig. 5(c)). The noise power estimates result from projecting the image onto the squared moduli of the Fourier basis functions. Since the modulus of any complex sinusoid is unity at all points in space, the noise power estimates in the Fourier domain are proportional to the mean or DC value of the image, and hence identical for every Fourier coefficient. Therefore, the Fourier domain PRESS-optimal filter detects the presence of a *global* complex sinusoid relative to the *global* average intensity in the image. It is clear that Fourier-based filtering cannot adapt to local variations in the signal or noise. This observation

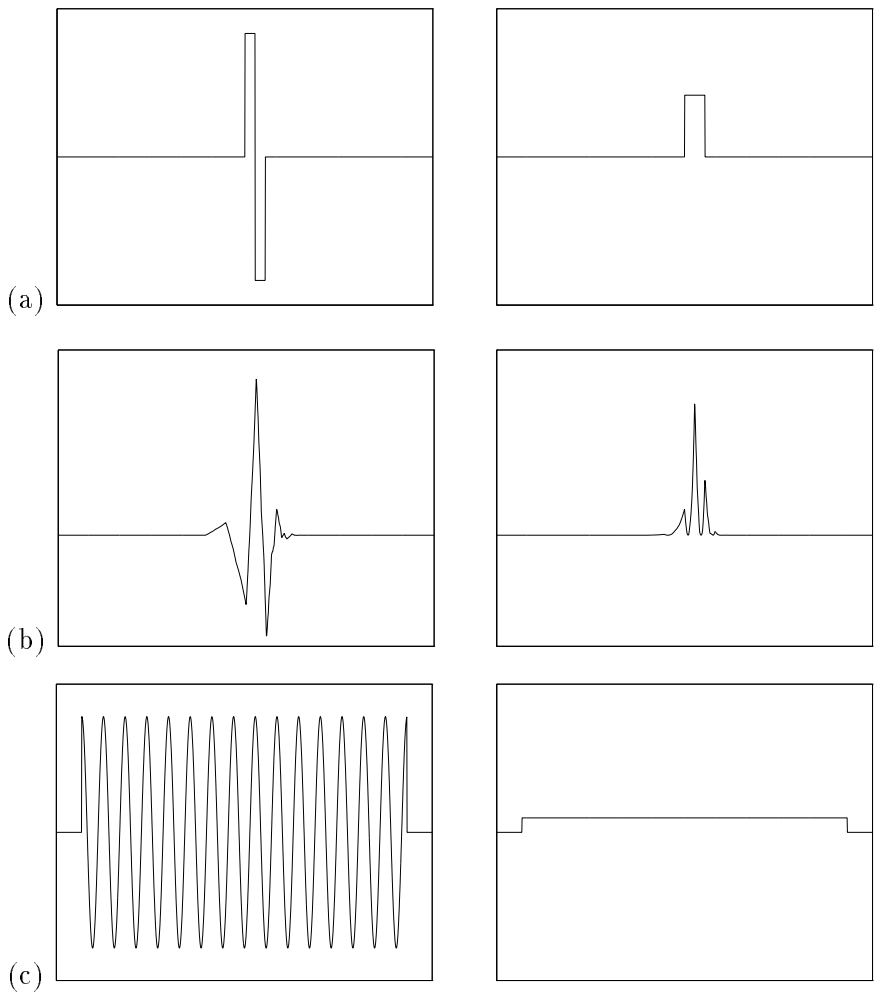

Fig. 5. *Comparison of the basis functions employed in PRESS-optimal filtering. (a) Haar wavelet w_I (edge detector, for estimating signal power) and squared wavelet w_I^2 (local averager, for estimating noise power). (b) Daubechies-6 wavelet w_I (edge detector) and squared wavelet w_I^2 (local averager). (c) Fourier-domain basis function (real part) and its modulus squared (global averager). Since the wavelet analysis is local, wavelet-domain filtering can adjust to local variations in the intensity. Fourier-domain filtering cannot.*

again demonstrates the fundamental advantage that wavelet-domain filtering offers over classical Fourier domain filtering.

C. PRESS-Optimal Filtering with Overcomplete Representations

Thus far, we have formulated our estimation problem in terms of an orthogonal wavelet expansion of the observations. However, as mentioned previously, we can design a PRESS-optimal filter for any orthogonal signal decomposition. The choice of basis can dramatically affect estimator performance (as seen in our analysis of the Fourier domain PRESS-optimal filter above). In this Section we extend our results to a more general filtering procedure based on overcomplete expansions (or frames [10]).

Overcomplete signal expansions can offer desirable invariance properties that are not available with an orthogonal expansion. For example, the orthogonal wavelet transform is not shift invariant: two different sets of wavelet coefficients (and hence PRESS-optimal filters) will result from transforming the images $f(k,l)$ and $f(k+1,l+1)$. By adding extra functions to the wavelet basis (essentially all possible shifts of the scaled wavelets and scaling functions), an overcomplete, shift-invariant wavelet decomposition results [15]. This "undecimated wavelet transform" can be computed using the same filter bank structure as the orthogonal wavelet transform, at a cost of $O(N_h N_f^2 \log N_f)$ computations for an $N_f \times N_f$ image. The undecimated wavelet transform can be inverted at the same cost using a "pseudo inverse" filter bank. Overcomplete wavelet expansions with other invariances, such as to rotations, are also easily devised.

The formulation of the PRESS-optimal filter for the overcomplete case is almost identical to the orthogonal case, except that we substitute the overcomplete basis functions into (4.13)–(4.15) and employ the pseudo-inverse to take us back to the original signal space. Since the pseudo-inverse effectively averages several correlated estimates together, it can often produce a better estimate than the orthogonal estimators previously considered.

VI. Fast Algorithm for PRESS-Optimal Wavelet-Domain Filtering

As we saw in (4.14), a key element of PRESS-optimal filtering involves projecting the image onto the pointwise *square* of the wavelet basis functions. In this Section, we will derive an efficient filter bank algorithm for computing this projection whose cost is linear in the number of image pixels. To our knowledge, this is a new (if somewhat obscure) contribution to the wavelet theory.[3]

Following the development of Section II, we begin with a 1-d continuous-time signal $f(t)$. Our

[3]The square of the wavelet has come up in other contexts; see [16] for example.

goal is to recursively compute the projections (2.2) and (2.3) but with ϕ and ψ replaced by ϕ^2 and ψ^2. We do this by developing analogous expressions to (2.6)–(2.7) using (2.4)–(2.5). Since the scaling function and wavelet are defined as a linear combination of scaling functions at the next coursest scale, the squared analogue to (2.4)–(2.5) must contain cross-products of scaling functions. With this in mind, define

$$u^j_{m,n} \triangleq 2^{-j} \int f(t)\, \phi(2^{-j}t - m)\, \phi(2^{-j}t - n)\, dt \tag{6.1}$$

$$v^j_{m,n} \triangleq 2^{-j} \int f(t)\, \psi(2^{-j}t - m)\, \psi(2^{-j}t - n)\, dt. \tag{6.2}$$

For a compactly supported scaling function and wavelet, the coefficients $u^j_{m,n}$ and $v^j_{m,n}$ can be interpreted as banded (near diagonal) square matrices of bandwidth $2N_h - 3$, with N_h the length of the associated wavelet filters h and g. Clearly we are interested in the values of these coefficients only on the diagonal ($m = n$).

Multiplying out (2.4) and (2.5) gives us

$$\phi(x/2 - i)\, \phi(x/2 - j) = 2 \sum_{m,n} h_m\, h_n\, \phi(x - 2i - m)\, \phi(x - 2j - n) \tag{6.3}$$

$$\psi(x/2 - i)\, \psi(x/2 - j) = 2 \sum_{m,n} g_m\, g_n\, \phi(x - 2i - m)\, \phi(x - 2j - n), \tag{6.4}$$

which lead directly to the recursions

$$u^{j+1}_{r,s} = \sum_{m,n} h_{2r-m}\, h_{2s-n}\, u^j_{m,n} \tag{6.5}$$

$$v^{j+1}_{r,s} = \sum_{m,n} g_{2r-m}\, g_{2s-n}\, u^j_{m,n}. \tag{6.6}$$

The recursion (6.5)–(6.6) forms the basis for a special 2-d filter bank: each iteration, the separable 2-d filters $h_m\, h_n$ and $g_m\, g_n$ convolve with the banded matrix $u^j_{m,n}$, which is then decimated in both dimensions. Note that even though these computations require that we increase the dimensionality of our problem from 1-d to 2-d, the 2-d problem is very "sparse." (Since $u^j_{m,n}$ is banded, most of its entries will be zero).

Given a discrete-time signal $f(k)$, we initialize the u matrix at scale zero by placing the signal onto the diagonal: $u^0_{m,n} = f(m)\, \delta(m - n)$.[4] By construction, the recursion (6.5)–(6.6) projects $f(k)$ onto the element-wise square $w^2_{j,n}(k)$ of each wavelet basis element $w_{j,n}$. Due to the sparsity

[4]Note that feeding the samples $f(kT)$ of a continuous-time signal $f(t)$ into a filter bank will not result in wavelet coefficients $\boldsymbol{\theta}$ that correspond to the values (2.2)–(2.3). (This would hold true only if $\phi(t-k)$ was a delta function.) In order for θ to equal the values (2.2)–(2.3), we must preprocess the signal samples by projecting them onto the space

of $u^j_{m,n}$, full 2-d convolutions are not required to implement (6.5)–(6.6). In fact, the complete squared wavelet transform of an N_f-point signal requires only $O(N_h^2 N_f)$ computations, only a mild (constant multiplier) increase over the fast wavelet transform.

The setup for projecting images onto the pointwise square of the 2-d wavelet basis is a simple extension of the above. The scaling wavelet coefficient array becomes a (very sparse) banded square 4-d array $u^j_{r,s,m,n}$, with all of the nonzero values clustered along the "diagonal" $u^j_{m,n,m,n}$ (similar for the wavelet coefficient array $v^j_{r,s,m,n}$). The filters h and g work alternately on each of the 4 dimensions as the 4-d convolutional kernels $h_r h_s h_m h_n$, $h_r h_s g_m g_n$, $g_r g_s h_m h_n$, and $g_r g_s g_m g_n$. Decimation takes place in all 4 dimensions. Due to the sparsity of the u and v arrays, the complete squared transform (4.14) requires only $O(N_h^3 N_f^2)$ computations for an $N_f \times N_f$ image.

Similar methods can be used to easily compute signal projections onto the squares of other bases associated with filter banks, such as wavelet packets, tree-structured wavelets, undecimated wavelet transforms etc.

VII. Examples and Applications

A. Simulated Data Experiment

In this Section, we present a simulated imaging experiment that compares the performance of the PRESS-optimal wavelet-domain filter with several competing methods. The simulated intensity image is shown in Fig. 6. The image has a background intensity of 5, dark regions (the circle) have intensity 0, and bright regions (the triangle) have intensity 200. The simulated image was designed to have a intensity range similar to that commonly encountered in applications (nuclear medicine imaging, for example).

Fig. 7(a) shows the raw count image obtained from the intensity image of Fig. 6 as well as the error between the raw counts and the underlying intensity. The raw counts form the Maximum Likelihood estimate (MLE) for the intensity. While unbiased, this estimate suffers from a very

spanned by the scaling functions $\phi(t-k)$ [17], [12, p. 232]. Abry and Flandrin suggested a simple approximation to this projection in [17]; they initialize the filter bank (2.6)–(2.7) with

$$c^0_n \;\approx\; \sum_r f(rT)\,\phi(r-n). \tag{6.7}$$

An analogous prefiltering step for initializing the squared wavelet transform filter bank (6.5)–(6.6) is

$$u^0_{m,n} \;\approx\; \sum_r f(rT)\,\phi(r-m)\,\phi(r-n). \tag{6.8}$$

Note that this initializes both the diagonal and off-diagonal entries.

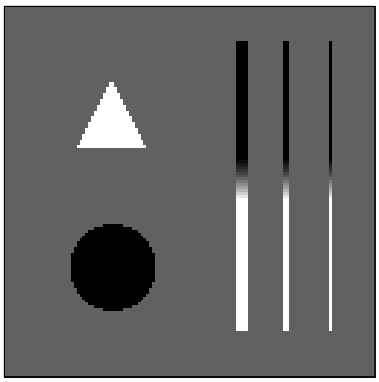

Fig. 6. *Phantom intensity image for the simulation experiments. Maximum intensity* 200 *(triangle), minimum intensity* 0 *(circle).*

large variance. The MLE might be improved by incorporating a Bayesian prior on the underlying intensity process, but this could introduce a bias. In this paper, we do not consider the use of such priors.

Fig. 7(b) depicts the estimate obtained by convolving the raw count image with the nine-point smoothing filter

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{7.1}$$

Note that the noise is significantly reduced, but the smoothing introduces large artifacts near the edges in the image.

In Fig. 8 we compare the estimates obtained from three different wavelet-domain filters, all based on the Haar wavelet. First consider Wavelet Shrinkage using a soft threshold (see (3.3) and Fig. 4). Wavelet Shrinkage assumes an additive white Gaussian noise, and the threshold is set based on the global variance of the noise. Since the Gaussian assumption clearly violated in the Poisson case, we work with the square-root of the counts (for variance stabilization). As is standard practice, we estimate the variance of the noise using the median absolute value of the finest scale wavelet coefficients [18]. We then apply the standard soft-threshold wavelet-domain filter to the wavelet transform of the square-root image, inverse wavelet transform and square the result to obtain an estimate of the intensity.

Fig. 8(a) shows the estimate obtained by the soft-threshold wavelet-domain filter. Note the large error in the high intensity region (the triangle) of the estimate. This error is due to the inherent

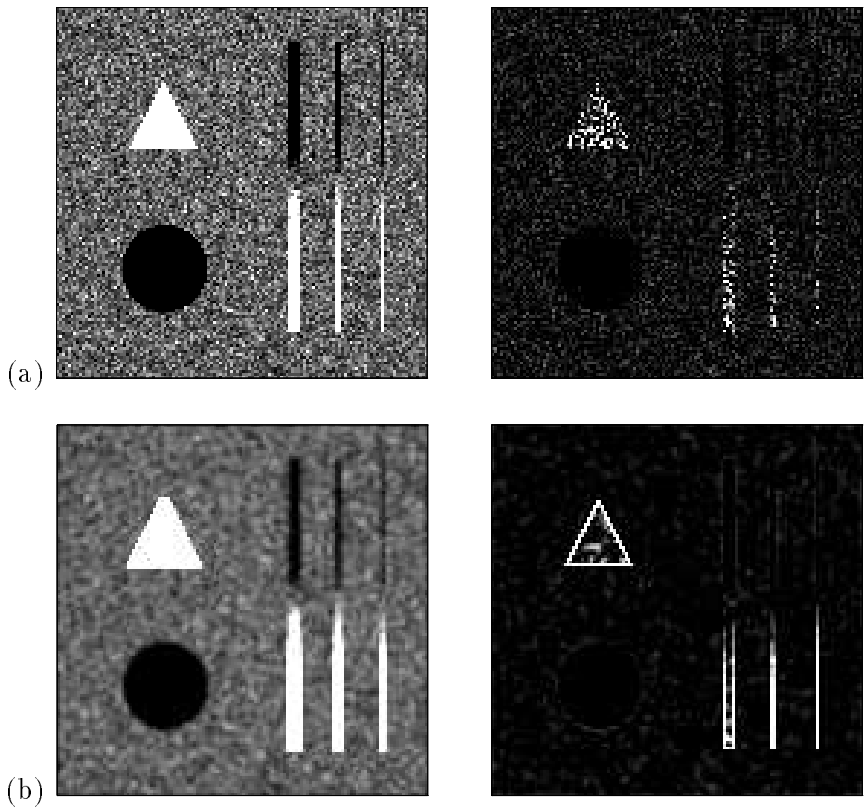

Fig. 7. *Comparison of various noise removal methods I. (a) Raw counts (MLE) estimate. (b) Nine-point convolutional smoothing filter estimate. At left we show the estimates; at right the errors between the estimates and the true simulated intensity of Fig. 6. The nine-point smooth estimate does an excellent job of reducing the noise in the background, but introduces a severe bias near edges.*

bias of the soft-threshold operation and the spatially varying noise level that is not sufficiently stabilized via the square root operation due to the low-count nature of the data. Unfortunately, at the count levels (intensities) often encountered in practice, we are far from the asymptotic condition needed to stabilize the variances by a simple square-root transformation. While working with the square-root of the counts does tend to equalize the disparity between the noise powers somewhat, it does not eliminate the problem of signal-dependent noise (noise powers remain higher in regions of higher intensity).

An alternative to the soft-threshold is the hard-threshold operation, which simply discards wavelet coefficients falling below the noise threshold and leaves those coefficients above the threshold unaltered (see Fig. 4). The hard-threshold estimate (also variance stabilized using a square root) is shown in Fig. 8(b). It is unbiased at high SNR, and therefore we see slightly less error in the high intensity region compared to the soft-threshold estimate. However, as is evident in the filtered image and error image, the hard-threshold does not reduce the noise as significantly as the soft-threshold.

The PRESS-optimal estimate, shown in Fig. 8(c), automatically adjusts to the spatially varying noise power, so we work directly with the raw count image. This estimate shows significantly less error in the high intensity regions of the image in comparison to the estimate obtained via soft-thresholding. This improvement may be attributed to the PRESS-optimal filter's automatic adaptation to the local noise level and the unbiased nature of the PRESS-optimal estimator at high SNR (see Fig. 4). The PRESS-optimal estimate far outperforms the hard-threshold estimate. Hence, in terms of balancing the trade-off between bias and variance, the PRESS-optimal estimator provides an excellent alternative to the competing methods. Furthermore, while the PRESS-optimal estimate does minimize the PRESS criterion, in small-sample situations the soft and hard-thresholding schemes are not necessarily optimal in any sense.[5]

The improvement of the PRESS-optimal estimator is not without a cost, however. Due to the added adaptability of the PRESS-optimal filter, the estimate is slightly more sensitive to the noise. As a result, the variability of the PRESS-optimal estimator is generally higher than that of the soft-threshold estimator (note the slightly higher noise level in the background of the PRESS-optimal estimate in comparison to the soft-threshold estimate).

The squared bias, variance, and overall mean-squared error (MSE) of each estimator, normalized by the total sum of pixel intensities, is given in Table I. The nine-point smooth, soft-threshold, and

[5]The soft-threshold procedure is *asymptotically* minimax optimal for a broad classes of signals [6].

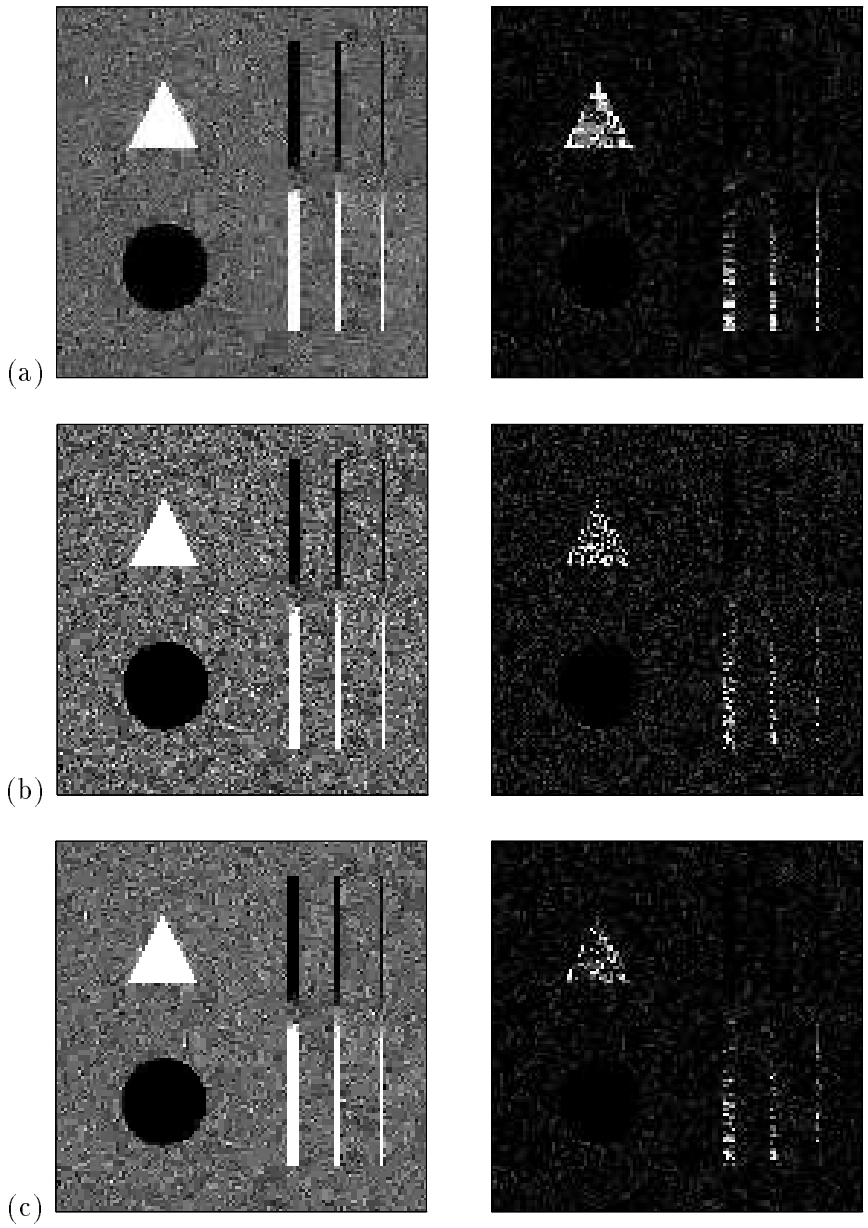

Fig. 8. *Comparison of various noise removal methods II. (a) Wavelet Shrinkage estimate using soft threshold. (b) Wavelet Shrinkage estimate using hard threshold. (c) PRESS-optimal estimate. All estimates were computing using the Haar wavelet basis. At left we show the estimates; at right the errors between the estimates and the true simulated intensity of Fig. 6. Unlike the soft-threshold estimate, the hard-threshold estimate does not introduce a large bias in the high intensity regions of the image. However, the hard-threshold operation does a relatively poor job at noise reduction. The PRESS-optimal estimate preserves edges well, significantly reduces the noise, and has a lower error in high intensity regions compared to the soft-threshold estimate.*

Table I. *Comparison of estimators of Poisson intensity.*

Estimator	Bias Squared	Variance	MSE
Raw counts	0.0010	0.9995	1.0004
Nine-point smooth	5.2218	0.1402	5.3620
Wavelet hard-threshold	0.0017	0.8497	0.8513
Wavelet soft-threshold	0.2006	0.2501	0.4507
Wavelet PRESS-optimal	0.0087	0.4095	0.4182

PRESS-optimal estimators provide excellent variance reduction. The nine-point smooth has the smallest variance but extremely large bias. While almost unbiased, the hard-threshold estimator has a variance that is only modestly lower than that of the original raw counts. In contrast, the PRESS-optimal and soft-threshold estimators provide a better balance of bias and variance. Note that the PRESS-optimal estimator has a much smaller bias than the soft-threshold estimator, but larger variance. Overall, the PRESS-optimal estimator provides the lowest MSE. The excessive bias introduced by the soft-threshold is an important concern, particularly if quantitative image analysis is to be performed on the filtered data. Comparing the PRESS-optimal estimator's performance to that of the raw count image, we see that we sacrifice a negligible bias in return for a large reduction in variance.

B. Nuclear Medicine Image Estimation

Nuclear medicine imaging is a widely used commercial imaging modality [1]. Unlike many other medical imaging techniques, nuclear medicine imaging can provide both anatomical *and* functional information. However, nuclear medicine imaging has a much lower SNR compared to other imaging techniques. Hence, improvements in image quality via optimized signal processing represent a significant opportunity to advance the state-of-the-art in nuclear medicine.

Nuclear medicine images are acquired by the following procedure [1]. Radioactive pharmaceuticals that are targeted for uptake in specific regions of the body are injected into the patient's bloodstream. As the radioactive pharmaceuticals decay, gamma rays are emitted from within the patient. Imaging the gamma ray emissions provides a mapping of the distribution of the pharmaceutical, and hence a mapping of the anatomy or physiologic function of the patient. Gamma rays are detected and spatially located using a gamma camera, which converts gamma rays into light. Photomultipier tubes then detect and locate the emissions. The raw nuclear medicine data is an

image of photon detections (counts). The raw data may be viewed directly or used for tomographic reconstruction. The major limitation of nuclear medicine imaging is the low-count levels acquired in typical studies. The count levels are low due to the limited level of radioactive dosage required to insure patient safety.

Because of the relatively small number of counts in a typical nuclear medicine study, the raw-count image appears noisy and in fact has a large variance about its true mean (the underlying intensity of the Poisson process). The variance is proportional to the signal itself and hence linear filtering of the raw image may result in unsatisfactory signal degradation. The raw image can be improved using the PRESS-optimal estimator. For this application, we have chosen the Daubechies-6 wavelet basis [10, 11, 12] for the PRESS-optimal estimator. This basis was selected for its good localization and smoothness properties [19]. Empirical studies have shown that this basis provides excellent filtering results for a wide variety of nuclear medicine images [20].

Bone Imaging: Fig. 2(a) depicts an image of the spine obtained from a nuclear medicine study. The radiopharmaceutical used here is Technetium-99m labeled diphosphonate. In bone studies such as this, brighter areas indicate increased uptake of blood in areas where bone growth is occurring. This may reflect areas where bone damage has occurred. Functional changes in bone can be detected using nuclear medicine image before they will show up in x-ray images. Fig. 2(b) shows the filtered spine image using the PRESS-optimal wavelet-domain filter using the Daubechies-6 wavelet basis.

Heart Imaging: Fig. 9(a) depicts an image of a heart obtained from a nuclear medicine study. The image was obtained using the radiopharmaceutical is Thallium-201. In this type of study, the radiopharmaceutical is injected into the bloodstream of the patient and moves into the heart wall in proportion to the local degree of blood perfusion. The purpose of this procedure is to determine if there is decreased blood flow to the heart muscle. Fig. 9(b) shows the filtered heart image using the PRESS-optimal wavelet-domain filter using the Daubechies-6 wavelet basis.

VIII. Conclusions

The PRESS-optimal wavelet-domain filter is an optimal, data-adaptive filter designed using the method of cross-validation. When the data is Poisson in nature, the filter can be computed directly from the raw data. Moreover, the filter has a simple and elegant interpretation in the spatial domain as a joint edge detection/estimation procedure. We have derived a fast algorithm for the optimal filtering procedure whose order of complexity is equivalent to that of the fast wavelet transform.

We have also explored the bias and variance trade-offs associated with various noise filtering

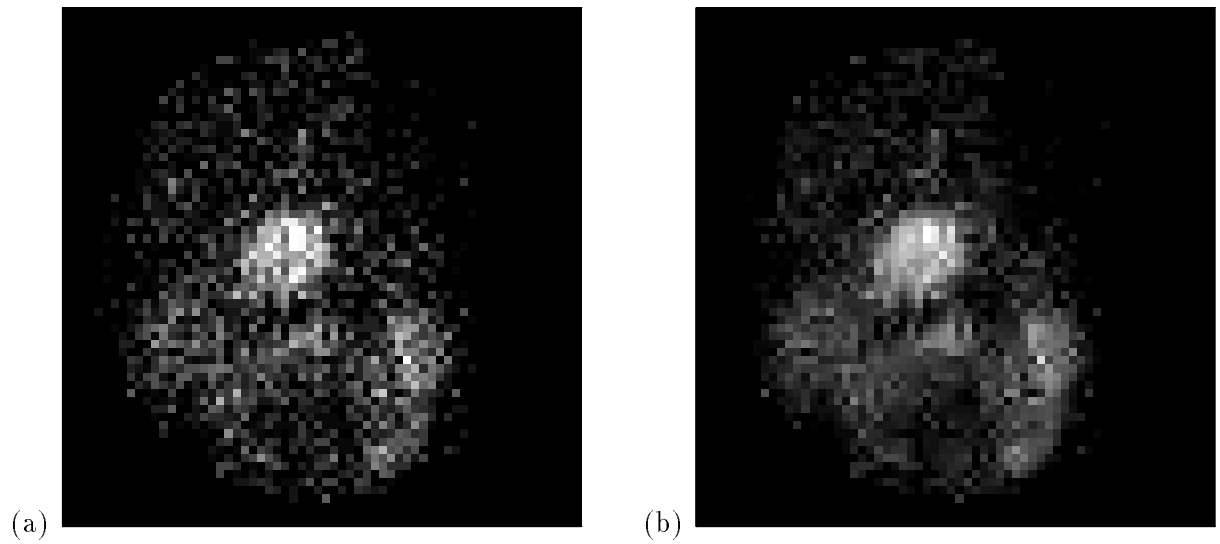

Fig. 9. *Comparison of nuclear medicine heart images. (a) Raw-count image (b) PRESS-optimal filtered image using Daubechies-6 wavelet basis. The wavelet-domain filtering dramatically reduces the noise, but does not excessively smooth the edges.*

methods for Poisson data in a simulated experiment. Linear filtering significantly reduces the variance, but can lead to large bias errors. Conventional soft-threshold Wavelet Shrinkage applied to the square-root of the raw data provides a reasonably balance between bias and variance. The PRESS-optimal wavelet-domain filter favors bias over variance reduction, and in the experiment we considered it produced a lower MSE than all other methods. Of course, it is incorrect to infer that the PRESS-optimal filter will provide the best MSE performance in general. However, the PRESS-optimal filter does minimize a well-defined data-based criterion and our experience has shown that it generally leads to an estimator that significantly reduces the variance while introducing a negligible bias. The excellent performance of the PRESS-optimal filter in this respect makes it highly desirable for a wide variety of applications involving Poisson data.

We have shown that the PRESS-optimal filter can be applied to nuclear medicine imaging, resulting in much better estimates of the underlying intensity. There are several avenues for future work in this application. First, the new filtering method could be extended to filtering for tomographic reconstruction. Second, the PRESS-optimal filter can be used in conjunction with "resolution recovery" schemes to reduce the effects of blurring inherent in the imaging process, while still reducing noise. A more comprehensive study and analysis of the PRESS-optimal wavelet-domain filter for nuclear medicine is given in [20].

Finally, we mention that the PRESS-optimal wavelet-domain filter is also applicable to a wide

variety of other imaging modalities including low-light imagery, astronomical imaging, x-ray tomography, and positron emission tomography. The filter is also appropriate for smoothing histogram estimates. Furthermore, if multiple iid observations of a noisy signal are observed, then the cross-validation framework can even handle correlated, non-Gaussian noises (see [9] for more details). We are currently investigating these topics.

Appendix A

Proof of Theorem 4.1

The key to the proof is noting that by increasing N, which subdivides the observation period T into smaller and smaller time intervals, we are eventually left with N extremely low-count images. Each low-count image is either all zero or contains a single count.

Recall the estimated variance in (4.6)

$$\widehat{\sigma_I^2}(N) \;=\; \frac{N}{N-1}\sum_{m=1}^{N}\left(\theta_I^{(m)} - \frac{1}{N}\theta_I\right)^2. \tag{A.1}$$

We will show that $\widehat{\sigma_I^2}(N) \to \widehat{\sigma_I^2}$, defined in (4.14) above, as $N \to \infty$. Note that (4.13) is precisely the limiting form of the PRESS-optimal filter, as $N \to \infty$. It then follows that (4.13) is the filter that minimizes the PRESS criterion as $N \to \infty$.

First, expand $\widehat{\sigma_I^2}(N)$

$$\widehat{\sigma^2}_I(N) \;=\; \frac{N}{N-1}\left(\sum_{m=1}^{N}\left(\theta_I^{(m)}\right)^2 - \frac{1}{N}\theta_I^2\right). \tag{A.2}$$

Here we have used the fact that $\sum_{m=1}^{N}\theta_I^{(m)} = \theta_I$. Next, expand the wavelet coefficients $\theta_I^{(m)}$ and θ_I

$$\widehat{\sigma_I^2}(N) \;=\; \frac{N}{N-1}\left(\sum_{m=1}^{N}\left(\sum_{k,l} w_I(k,l)c^{(m)}(k,l)\right)^2 - \frac{1}{N}\left(\sum_{k,l} w_I(k,l)f(k,l)\right)^2\right), \tag{A.3}$$

with $c^{(m)}(k,l)$ the count in k,l-th pixel of the m-th low-count image. Expanding further we reach

$$\widehat{\sigma_I^2}(N) \;=\; \sum_{k_1,k_2,l_1,l_2} w_I(k_1,l_1)w_I(k_2,l_2)\left[\frac{N}{N-1}\left(\sum_{m=1}^{N} c^{(m)}(k_1,l_1)c^{(m)}(k_2,l_2) - \frac{1}{N}f(k_1,l_1)f(k_2,l_2)\right)\right]. \tag{A.4}$$

Now consider just the innermost bracketed term in (A.4). Since $f(k_1,l_1)$ and $f(k_2,l_2)$ are finite numbers, we have

$$\lim_{N\to\infty}\;\frac{1}{N}f(k_1,l_1)f(k_2,l_2) \;=\; 0. \tag{A.5}$$

Also, assuming that N is sufficiently large so that each low-count image is either all zero or only contains a single count, we have two cases:

$$\sum_{m=1}^{N} c^{(m)}(k_1,l_1)c^{(m)}(k_2,l_2) \;=\; \begin{cases} f(k_1,l_1), & \text{if } k_1 = k_2, l_1 = l_2 \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

Combining the above results, we see that

$$\lim_{N\to\infty} \widehat{\sigma_I^2}(N) \;=\; \sum_{k,l} w_I^2(k,l) f(k,l), \tag{A.7}$$

and the proof is complete. □

Acknowledgments: The authors thank Dave Nowak for his helpful discussions regarding nuclear medicine imaging and for providing the nuclear medicine heart study data, Ivan Selesnick for discussing the subtleties of implementing the discrete wavelet transform, and Robert Hellman (Division of Nuclear Medicine, Medical College of Wisconsin) for providing the spine study data.

References

[1] J. A. Sorenson and M. E. Phelps, *Physics in Nuclear Medicine*. New York: Grune & Stratton, 1980.

[2] J. W. Goodman and J. F. Belsher, "Fundamental limitations in linear invariant restoration of atmospherically degraded images," in *Proc. Soc. Photo-Opt. Instrum.*, pp. 141-154, 1976.

[3] H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A*, vol. 7, no. 7, pp. 1266–1278, 1990.

[4] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*. New York: John Wiley and Sons, 1992.

[5] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 7, pp. 165–177, 1985.

[6] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Soc.*, vol. 90, pp. 1200–1224, Dec. 1995.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transforms," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[8] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons, 1990.

[9] R. D. Nowak, "Optimal signal estimation using cross-validation," *IEEE Signal Processing Letters*, vol. 3, no. 12, pp. 23–25, 1996.

[10] I. Daubechies, *Ten Lectures on Wavelets*. New York: SIAM, 1992.

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

[12] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellsley, MA: Wellsley-Cambridge, 1996.

[13] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *App. and Comp. Harmonic Analysis*, vol. 1, pp. 100–115, Dec. 1993.

[14] E. L. Lehmann, *Theory of Point Estimation*. Pacific Grove, CA: Wadsworth & Brooks/Cole, 1983.

[15] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discret wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[16] M. Girardi and W. Sweldens, "A new class of of unbalanced haar wavelets that form an unconditional basis for L_p on general measure spaces," *preprint*, 1997.

[17] P. Abry and P. Flanrdin, "On the initialization of the discrete wavelet transform algorithm," *IEEE Signal Processing Letters*, vol. 1, pp. 32–34, Feb. 1994.

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

[19] I. Daubechies, *Ten Lectures on Wavelets*. Philidelphia: SIAM CBMS-NSF Series in Applied Mathematics, no. 61, 1992.

[20] R. Nowak, R. Hellman, D. Nowak, and R. Baraniuk, "Wavelet domain filtering for nuclear medicine imaging," in *Proc. IEEE Med. Imaging Conf.*, pp. 279–290, 1996.