A Semi-Local Paradigm for Wavelet Denoising

Richard Charnigo, Jiayang Sun, and Raymond Muzic

***Abstract*—Wavelet denoising methods have been proven useful for many one- and two-dimensional problems. Most existing methods can in principle be carried over to three-dimensional problems, such as the denoising of volumetric positron emission tomography (PET) images, but they may not be sufficiently flexible in allowing some regions of an image to be denoised more aggressively than others. In this paper, we propose a *semi-local* paradigm for wavelet denoising. The semi-local paradigm involves the division of an image into suitable blocks, which are then individually denoised. To denoise the blocks, we use our modification of the generalized cross validation (GCV) technique of Jansen and Bultheel [1] to choose thresholding parameters; we also present risk estimators to guide some of the other choices involved in the implementation. Experiments with phantom PET images show that the semi-local paradigm provides superior denoising compared to standard application of the GCV technique. An asymptotic analysis demonstrates that, under some regularity conditions, semi-local denoising is asymptotically consistent on the logarithmic scale. The paper concludes with a discussion on the nature of semi-local denoising and some topics for future research.**

***Index Terms*—Imaging, logarithmic consistency, positron emission tomography, thresholding.**

I. Introduction

LET y denote an image corrupted by noise, and let f denote its noncorrupted analogue; we may think of f as a discretized version of an unobservable true image. Consider the problem of denoising y (i.e., estimating f from y). For a well-chosen invertible operator W, it may be easier to estimate Wf from Wy and apply W^{-1} than to estimate f from y directly. This is often the case when W induces a change from the image domain to a new domain in which the true image has a sparse representation with respect to the basis functions. In such a situation, the small coefficients in Wy are presumed to be manifestations of the noise, so a reasonable strategy is to reduce or eliminate these small coefficients through application of some thresholding rule governed by one or more thresholding parameters.

Since piecewise smooth true images have sparse representations with respect to *wavelet* basis functions (see [2]–[4]), wavelet denoising methods have been intensively researched over the past two decades. Donoho and Johnstone [5], [6] provided an asymptotically optimal choice for a single thresholding parameter under soft thresholding and an assumption of white noise. Following their work, Nason [7] and Jansen, Malfait, and Bultheel [8] proposed cross validation (CV) and generalized cross validation (GCV) techniques, respectively, for thresholding parameter selection. Johnstone and Silverman [9] suggested dealing with correlated noise by grouping wavelet coefficients according to their resolution levels and then selecting thresholding parameters separately for each resolution level. Jansen and Bultheel [1] subsequently revised the GCV technique to accommodate correlated noise. Thresholding rule specification has been further examined by Antoniadis and Fan [10] and by Figueiredo and Nowak [11]. Empirical Bayes frameworks have been proposed by Jansen and Bultheel [12] and by Johnstone and Silverman [13]. Nason [14] has designed a procedure for determining the number of resolution levels at which thresholding is performed. Issues of basis selection have been studied by Candes and Donoho [15] and by Ferrando and Kolasa [16].

Most wavelet denoising methods were designed for one- and two-dimensional problems but can in principle be extended to three-dimensional problems. However, many of these existing methods deal with the *whole* image when setting up a mechanism for the thresholding of wavelet coefficients. Hence, these methods may not be sufficiently flexible to handle realistic three-dimensional problems. For example, medical images may have extremely localized structures of particular importance, perhaps depicting tumors or other anomalies; if the existence and nature of these structures are not known well enough *a priori* to accommodate sophisticated modeling, the structures may be inadvertently removed or degraded along with the noise.

A few more recent works, aimed at making wavelet denoising more flexible, include the following. Chang *et al.* [17] suggested that image features such as edges be modeled in order to make denoising more adaptive to the local characteristics of an image. Sendur and Selesnick [18] proposed bivariate shrinkage with local variance estimation. Cai [19], Hall *et al.* [20], and Chen and Bui [21] advocated block thresholding, in which thresholding decisions are made not for individual wavelet coefficients but for small groups of neighboring wavelet coefficients.

In this paper, we present an alternative approach for making wavelet denoising more flexible. Drawing an analogy between the thresholding of wavelet coefficients and dealing with an overfitted regression model, one may consider the notion of a *local* paradigm for wavelet denoising *in the image domain*: each element of y could be operated on individually, replaced by a fitted value based on a wavelet denoising of a neighborhood containing that element. Unfortunately, such a local paradigm would be impractical with existing computational resources, at

Manuscript received April 24, 2004; revised March 10, 2005x. The work of R. Charnigo and J. Sun was supported in part by a grant from the National Science Foundation. The Associate Editor coordinating the review of this manuscript and approving it for publication was Dr. Truong Q. Nguyen.

R. Charnigo is with the Department of Statistics, University of Kentucky, Lexington, KT, 40506-0027 USA (e-mail: richc@ms.uky.edu).

J. Sun is with the Department of Statistics, Case Western Reserve University, Cleveland, OH, 44106-7054 USA (e-mail: jiayang@sun.cwru.edu).

R. Muzic is with Nuclear Medicine, University Hospitals of Cleveland, Cleveland, OH, 44106 USA (e-mail: muzic@uhrad.com).

Digital Object Identifier 10.1109/TIP.2005.863037

least for three-dimensional problems: for a $128 \times 128 \times 47$ element PET image, more than 770 000 wavelet transform operations would be involved! Thus, we propose a *semi-local* paradigm for wavelet denoising, in which the image is divided into suitable blocks that are then individually denoised. This allows some parts of the image to be denoised more aggressively than others, providing a flexible trade-off between variance reduction (i.e., elimination of noise) and bias control (i.e., preservation of localized structures). In a medical image, for instance, we would like to denoise more aggressively in regions of normal physiological activity or anatomical structure but more conservatively in the presence of abnormal activity or structure.

Our semi-local paradigm may be contrasted with the block thresholding in [19]–[21]. There, the blocks are subsets of coefficients *in the wavelet domain*, organized after the entire image has been subjected to a wavelet transform; the idea is to borrow information from the other coefficients in a block when deciding whether to threshold a particular coefficient. Our blocks are subsets of elements *in the image domain*, organized so that different parts of the image are subjected to their own wavelet transforms; this allows more flexible denoising that directly takes into account the local characteristics of the image.

This paper is organized as follows. The statistical model and some other preliminaries are described in Section II. Our semi-local paradigm is developed in Section III, where prescriptions for the choices of thresholding and other key parameters are discussed. The GCV technique [1] is modified to choose thresholding parameters block-by-block in the image domain. The modification is that nontrivial lower bounds, based on the local characteristics of the image, are placed on the ranges over which minima of the GCV criteria are sought. To guide some of the other choices required for implementation, we identify useful risk estimators as well as generalizations that allow more emphasis to be placed on bias control. In Section IV, visual and quantitative results from experiments with phantom PET images show that the semi-local paradigm is demonstrably more effective than standard application of the GCV technique. In Section V, semi-local denoising is shown to provide asymptotically consistent estimators of the logarithms of (unknown) risk-minimizing thresholding parameters. Thus, the semi-local paradigm provides a computationally feasible and theoretically justified approach for varying the degree of denoising across different regions of an image without *a priori* knowledge about the image's structures. Possibilities for future research, including an extension that would make semi-local denoising even more adaptive, are described in Section VI. Details of the theoretical justification are provided in the Appendix.

II. Preliminaries

A. Statistical Model

Consider the additive error model

$$Y = f + \epsilon \tag{1}$$

in which we observe noisy D-dimensional data y (a realization of Y) on a lattice $\mathcal{L}$ contained in the D-dimensional discrete set $\{0, 1/2^M, \ldots, (2^M - 1)/2^M\}^D$, a grid consisting of $2^M \times \cdots \times 2^M$ points. The true D-dimensional image $f(x)$ is defined on a convex subspace of $\Re^D$ containing $\mathcal{L}$, and f denotes the restriction of $f(x)$ to $\mathcal{L}$. The noise ϵ is assumed to be stationary: the covariance between any two components of ϵ depends only on the distance between the corresponding points of $\mathcal{L}$. Our goal is to estimate f by denoising y.

Wavelet denoising methods rely on the principle that many images $f(x)$ have sparse representations in terms of wavelet basis functions $\phi(x)$ and $\psi_{l,o,i}(x)$ [13]

$$f(x) \approx v_L \phi(x) + \sum_{l=1}^{L-1} \sum_{o=1}^{2^D-1} \sum_{i=1}^{n_l} v_{l,o,i} \psi_{l,o,i}(x) \tag{2}$$

where many of the coefficients $v_{l,o,i}$ are zero (or nearly zero). In (2), L is the number of resolution levels and is related to M by the equation $L = M + 1$, while the image dimension D determines $2^D - 1$ orientations at each resolution level $l \leq L-1$. There are $n_l := (2^{M-l})^D$ basis functions at each resolution level $l \leq L - 1$ and orientation o, so that the index i pertains to the position of a corresponding wavelet basis function at level $l \leq L - 1$ and orientation o. When $D = 1$, $\phi(x)$ is a father wavelet function and the $\psi_{l,o,i}(x)$ are related to a mother wavelet function $\psi(x)$ through dilation and translation [8]. When $D > 1$, $\phi(x)$ is a D-fold tensor product of father wavelet functions, and the $\psi_{l,o,i}(x)$ are D-fold tensor products involving dilated and translated father and mother wavelet functions. The basis functions at the top resolution levels (i.e., corresponding to small values of l) are associated with the finest and most localized structures of $f(x)$.

B. GCV Technique

The GCV technique for wavelet denoising was proposed by Jansen and Bultheel [1] for one- and two-dimensional problems but can be extended to three-dimensional problems such as the denoising of volumetric medical images. The GCV technique is actually applied repeatedly, once for each level and orientation in the wavelet domain; this is because the assumption of stationary noise carries over into the wavelet domain but only within each level and orientation [1]. The GCV technique is appealing because no estimation of $Var(\epsilon)$ in model (1) is required, nor is *a priori* knowledge needed about any structures in $f(x)$.

Let $w_{l,o,1}, \ldots, w_{l,o,n_l}$ denote the wavelet coefficients in Wy associated with a specific level l and orientation o at which we wish to do thresholding. Given the value of the thresholding parameter $\delta_{l,o}$, the soft-thresholded version of the i^{th} coefficient at level l and orientation o is

$$w_{l,o,i;\delta_{l,o}} := \operatorname{sgn}(w_{l,o,i}) \left(|w_{l,o,i}| - \delta_{l,o}\right)_+ . \tag{3}$$

For ease of exposition, we will hereafter abbreviate $w_{l,o,i;\delta_{l,o}}$ to $w_{l,o,i;\delta}$. The GCV technique entails finding nonzero $\delta_{l,o}$ to minimize

$$GCV_{l,o}(\delta_{l,o}) := \frac{n_l^{-1} \sum_{i=1}^{n_l} (w_{l,o,i} - w_{l,o,i;\delta})^2}{\left(n_l^{-1} \sum_{i=1}^{n_l} 1_{w_{l,o,i;\delta}=0}\right)^2}. \tag{4}$$

Soft thresholding is then performed at level l and orientation o using the selected value of $\delta_{l,o}$.

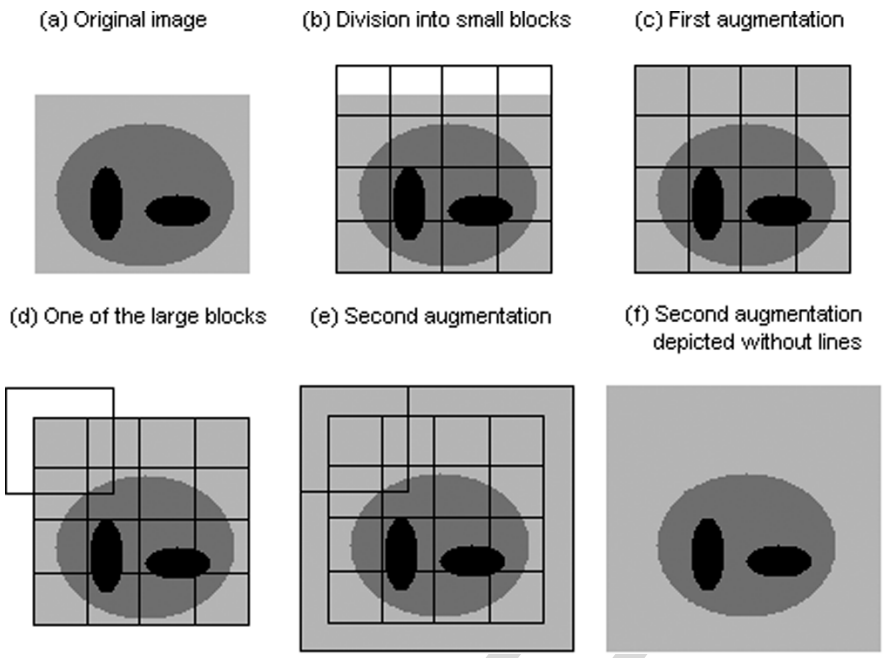

Fig. 1. Image augmentation. (a) Original image; (b) division into small blocks; (c) first augmentation; (d) one of the large blocks; (e) second augmentation; and (f) second augmentation depicted without lines.

The numerator of the GCV criterion (4), being proportional to the sum of squared differences between the unthresholded and thresholded coefficients, increases with $\delta_{l,o}$. So does the denominator, which is proportional to the square of the number of coefficients that are zero after thresholding.

An important issue is the selection of lower and upper bounds $L_{l,o}$ and $U_{l,o}$ for the range of $\delta_{l,o}$ values over which we seek a minimum of the GCV criterion (4). By default one may take $L_{l,o} := 0$ and $U_{l,o} := \max_i |w_{l,o,i}|$, although the former choice may not work well. We will discuss how to choose lower bounds in the next section, when we incorporate a modification of the GCV technique into our semi-local paradigm.

III. emi-Local Paradigm

A. Image Augmentation

We begin this section by describing how to perform image augmentation, a process that prepares an image for semi-local denoising by ensuring that all inputs to wavelet transform operators are well-defined.

Given an integer $J \geq 2$, we divide an image y into small nonoverlapping blocks of size $2^J \times \cdots \times 2^J$. Panel (a) of Fig. 1 shows a two-dimensional image y, and panel (b) shows its division into small nonoverlapping blocks. As is evident from panel (b) of Fig. 1, some of the small blocks will be incomplete if the dimensions of the lattice $\mathcal{L}$ on which y is defined are not multiples of 2^J. Thus, we create a *first augmented image* y^a, as depicted in panel (c) of Fig. 1. We define y^a on a lattice $\mathcal{L}^a$ whose dimensions are multiples of 2^J, and we do so in such a way that the restriction of y^a to $\mathcal{L}$ is y. When we divide this first augmented image into small nonoverlapping blocks of size $2^J \times \cdots \times 2^J$, all of the small blocks will be complete.

Next, we centrally embed each small block in a larger block of size $2^{J+1} \times \cdots \times 2^{J+1}$. Panel (d) of Fig. 1 shows the large block encasing the small block in the upper left corner of y^a. As is clear from panel (d), when a small block is on the boundary of y^a, the corresponding large block will be incomplete. Therefore, we create a *second augmented image* y^A, as depicted in panel (e) of Fig. 1. We define y^A on a lattice $\mathcal{L}^A$ in such a way that the restriction of y^A to $\mathcal{L}^a$ is y^a. Of course, the restriction of y^A to $\mathcal{L}$ is y. All of the large blocks in this second augmented image will be complete. Panel (f) of Fig. 1 shows the second augmented image without the lines indicating block divisions.

Example: If y is a three-dimensional positron emission tomography (PET) image of size $128 \times 128 \times 47$ and $J = 3$, we may construct y^a simply by appending a layer of zeros to y as a 48th horizontal slice. Since 128 and 48 are multiples of $8 = 2^3$, this first augmented image may be divided into small nonoverlapping blocks of size $2^3 \times 2^3 \times 2^3$, and all of the small blocks will be complete. However, a large block of size $2^4 \times 2^4 \times 2^4$ encasing a small block on the boundary of y^a will not be complete. Thus, we may define y^A to be an array of size $136 \times 136 \times 56$ whose central $128 \times 128 \times 48$ subarray is y^a and whose other elements are zeros. With the construction of this second augmented image, all of the large blocks will be complete.

The manner in which the augmentations are done may depend on the application at hand. For medical images obtained via PET, computerized tomography (CT), magnetic resonance imaging (MRI), or single photon emission computerized tomography (SPECT), augmentations via zero-padding are natural because zero represents an absence of physiological activity or anatomical structure. For other types of images, we may perform image augmentation by appending contextually meaningful nonzero values.

B. Semi-Local Denoising

The basic idea is intuitive. Starting with a noisy image y, we will obtain a first augmented image y^a that can be divided into

complete nonoverlapping small blocks and a second augmented image y^A such that all of the small blocks can be embedded in complete large blocks. We will then subject each large block in y^A to a wavelet denoising; details will be given in the next subsection. After each large block has been denoised, we will extract the denoised small block from its center. We will reassemble these denoised small blocks to yield a denoised version of y^a that we may call $\tilde{y}^a$. Our end product $\tilde{y}$ will be the restriction of $\tilde{y}^a$ to $\mathcal{L}$.

The rationale for operating on the large blocks and then extracting the denoised small blocks rather than operating directly on the small blocks themselves is as follows. First, proceeding in this way is more closely analogous to local regression, in which all the observations in a neighborhood are used to obtain a fitted value at a particular location; we do not want the denoised versions of adjacent small blocks to be obtained in complete isolation from each other. Second, if there are any edge effects when we denoise the large blocks, they will not be manifested in $\tilde{y}$.

Remark: It is possible to take the large blocks to have size $2^{J+I} \times \cdots \times 2^{J+I}$ for some $I > 1$. We leave $I = 1$ because this choice is the most computationally efficient and satisfactorily addresses both of the issues mentioned in the preceding paragraph.

C. Key Aspects of Semi-Local Denoising

We must decide how to determine J, representing *the size of the small blocks* into which an image is divided; how to choose *a wavelet basis* with regard to which wavelet coefficients will be computed for each large block; how to determine *the number of resolution levels* K at which wavelet coefficients will be thresholded; and, how to select *the thresholding parameters* for denoising each large block.

We will address first the selection of thresholding parameters, then how to choose J and K, and finally how to choose a wavelet basis.

1) Selecting the Thresholding Parameters: Assume for the moment that the wavelet basis and the values of J and K have been chosen. The choice of J has determined how many large blocks there are in the second augmented image y^A. Let B denote this number of large blocks, and let b serve as an index for the large blocks (i.e., $b = 1, \ldots, B$). Thus, we will write y_b^A to denote the b^{th} large block, $w_{l,o,i;b}$ to denote a wavelet coefficient associated with y_b^A, $\delta_{l,o;b}$ to denote a thresholding parameter associated with y_b^A, $L_{l,o;b}$ and $U_{l,o;b}$ to denote lower and upper bounds for this thresholding parameter, and $w_{l,o,i;b,\delta}$ to denote a thresholded wavelet coefficient associated with y_b^A.

We modify the GCV technique [1] not merely by working with the large blocks but by choosing the lower bounds $L_{l,o;b} > 0$ *according to the values* in y_b^A, allowing us to denoise either more conservatively or more aggressively based on the local characteristics of the image. Specifically, for every level l and orientation o such that $l \leq K$, let $q_{l,o;b}$ denote an upper quantile of the $|w_{l,o,i;b}|$, and let $g_l(y_b^A, y)$ be a nonnegative function of y_b^A and y that is large when the values in y_b^A dictate conservative denoising (e.g., when these values are too extreme or too variable) and small when the values in y_b^A call for aggressive denoising (e.g., when these values are modest and not too variable). We set

$$L_{l,o;b} := q_{l,o;b} \exp\left[-g_l\left(y_b^A, y\right)\right] \tag{5}$$

and seek $\delta_{l,o;b}$ between $L_{l,o;b}$ and $U_{l,o;b} := \max_i |w_{l,o,i;b}|$ to minimize

$$GCV_{l,o;b}(\delta_{l,o;b}) := \frac{n_{l;b}^{-1} \sum_{i=1}^{n_{l;b}} (w_{l,o,i;b} - w_{l,o,i;b,\delta})^2}{\left(n_{l;b}^{-1} \sum_{i=1}^{n_{l;b}} 1_{w_{l,o,i;b,\delta}=0}\right)^2} \tag{6}$$

where $n_{l;b} := (2^{J+1-l})^D$. We will refer to (6), used in conjunction with the data-dependent lower bound in (5), as the *modified GCV criterion*. We may add a tiny quantity to the sum in the denominator of (6) to ensure that no division by zero takes place. In our experiments with PET images, we used 10^{-8} both as this tiny quantity and as a tolerance to determine how many of the $w_{l,o,i;b,\delta}$ could be regarded as equal to zero.

For practical purposes, given the bounds $L_{l,o;b}$ and $U_{l,o;b}$, we may choose among $\delta_{l,o;b}$ on the grid

$$L_{l,o;b} + \frac{2n-1}{2^{P+1}}(U_{l,o;b} - L_{l,o;b}) \tag{7}$$

where $n = 1, \ldots, 2^P$ and P is selected to *balance* precision against computation time.

Remark: By requiring that $L_{l,o;b} > 0$ we also avoid a difficulty related to zero padding. When y is augmented through zero padding, some of the large blocks will have wavelet coefficients that are zero even *before* thresholding. The presence of wavelet coefficients that are zero before thresholding will inflate the denominator of the modified GCV criterion (6), so that the numerator becomes dominant. Consequently, if we allow $L_{l,o;b} = 0$, undesirably small values of $\delta_{l,o;b}$ may be selected.

Example, Continued: For a PET image, let $q_{l,o;b}$ and η_b be the upper 0.05 quantiles of the absolute wavelet coefficients $|w_{l,o,i;b}|$ and of the values in y_b^A, respectively. Set

$$g_l\left(y_b^A, y\right) := \left(\frac{\eta_b}{\zeta_l}\right) \log 2 \tag{8}$$

where each ζ_l is an upper quantile of the values in y. When η_b is small, indicating little activity in y_b^A, then $L_{l,o;b}$ is large; thus, in regions where we want to be aggressive in our denoising, we force $\delta_{l,o;b}$ to be large. On the other hand, when η_b is large, indicating considerable activity, then $L_{l,o;b}$ is small; in regions where we need to be more conservative, we do not force $\delta_{l,o;b}$ to be large.

2) Selecting J and K: Pilot experiments with test images can suggest choices of J and K for denoising similar images in the future. Indeed, the results of our experiments in the next section provide insight about suitable choices of J and K for images depicting structures that are heterogeneous in size or intensity. If it were anticipated that several combinations of J and K would produce comparable results, the choices of J and K could be made to minimize computation time. The computation time is related to the total number of thresholding parameters to be selected, which is proportional to $K/(2^J)^D$.

In the absence of pilot experiments, we may *estimate* the risk under squared error loss for denoising y (i.e., estimate the expected value of the mean squared error over repeated sampling) based on different choices of J and K, ultimately favoring those choices of J and K for which the estimated risk is lowest. This entails estimating the risks for denoising the y_b^A $(b = 1, \ldots, B)$, which can be accomplished by combining estimates of the risks associated with different levels and orientations in the wavelet domain. Explicitly, for each y_b^A and for each level l and orientation o, we will want to estimate the expected value of

$$R_{l,o;b}(\delta_{l,o;b}) := \sum_{i=1}^{n_{l;b}} \frac{[v_{l,o,i;b} - w_{l,o,i;b,\delta}]^2}{n_{l;b}} \tag{9}$$

where the $v_{l,o,i;b}$ are analogous to the $v_{l,o,i}$ in (2) but pertain to y_b^A rather than to y.

The modified GCV criterion (6) may be viewed as a proxy for the risk (9) when $\delta_{l,o;b} > 0$ but not when $\delta_{l,o;b} = 0$, as the numerator of (6) is then automatically zero. Hence, we need another proxy for the risk (9). One option is Stein's Unbiased Risk Estimator (SURE) [22]. However, there is now an extra cost: we must estimate the noise variance for the wavelet coefficients, which we will denote by $\sigma^2_{l,o;b}$. In a general setting, Donoho and Johnstone [6] have suggested estimating the noise variance as the square of the median absolute wavelet coefficient divided by 0.6745, which is based on the premise that most of the wavelet coefficients are dominated by noise and that the noise is approximately Gaussian.

We now introduce a generalization of SURE that allows for differential penalization of bias (i.e., oversmoothing) and variance (i.e., undersmoothing). Let

$$SURE_{l,o;b,c} := c \sum_{i=1}^{n_{l;b}} \frac{[w_{l,o,i;b} - w_{l,o,i;b,\delta}]^2}{n_{l;b}} + (2-c) 2\hat{\sigma}^2_{l,o;b} \left[0.5 - \sum_{i=1}^{n_{l;b}} \frac{1_{w_{l,o,i;b,\delta}=0}}{n_{l;b}} \right].$$

The parameter c may take values between 0 and 2. If $c = 1$, then we recover the standard SURE. If $c > 1$, then we have a linear combination of SURE and a term that penalizes bias. In medical applications, where oversmoothing is much worse than undersmoothing, it will generally be preferable to select $c > 1$. Our experiments, described below, suggest that $c = 1.9$ is in closer agreement with visual perception of image quality than $c = 1$.

We may also work with the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) rather than SURE. The corresponding expressions for $AIC_{l,o;b,c}$ and $BIC_{l,o;b,c}$ are given below (see [23]–[25] for the standard AIC and BIC)

$$AIC_{l,o;b,c} := c \sum_{i=1}^{n_{l;b}} \frac{[w_{l,o,i;b} - w_{l,o,i;b,\delta}]^2}{n_{l;b}} + (2-c) 2\hat{\sigma}^2_{l,o;b} \left[1 - \sum_{i=1}^{n_{l;b}} \frac{1_{w_{l,o,i;b,\delta}=0}}{n_{l;b}} \right],$$

$$BIC_{l,o;b,c} := c \sum_{i=1}^{n_{l;b}} \frac{[w_{l,o,i;b} - w_{l,o,i;b,\delta}]^2}{n_{l;b}} + (2-c) \hat{\sigma}^2_{l,o;b} \log(n_{l;b}) \left[1 - \sum_{i=1}^{n_{l;b}} \frac{1_{w_{l,o,i;b,\delta}=0}}{n_{l;b}} \right].$$

3) Choosing a Wavelet Basis: One could select a basis using the same approach suggested for choosing J and K: compute estimates of the risk for denoising y (or generalizations of such estimates) and select the basis yielding the smallest estimate. In our experiments, however, we have retained the $S8$ basis functions [26].

IV. Experiments With Phantom PET Images

In this section, we deal with two volumetric PET images, both depicting the Galaxi Hot Spheres phantom (Biomec, Inc.; Cleveland, Ohio), a cylindrical object that can be thought of as a crude proxy for a human torso and that is used for testing imaging hardware and software.

The horizontal cross sections of the Galaxi Hot Spheres phantom correspond to one of three spatial distributions of activity: uniform (no tumors), multi-resolution (tumors of six different sizes but all of the same high intensity), and multi-contrast (tumors of two different sizes and three different intensities). We will concentrate on each image as a whole as well as on the multi-resolution and multi-contrast distributions in particular.

Emission data were acquired using a model 921 ECAT EXACT scanner (Siemens/CTI; Knoxville, Tennessee); refer to Wienhard *et al.* [27] for a discussion of this scanner's capabilities. The images were reconstructed via filtered backprojection using Version 7.2.1 of the scanner manufacturer's ECAT software.

Using the phantom manufacturer's specifications, we also constructed a target image. This allowed us to calculate the signal-to-noise ratio (SNR) for each of the two phantom images via

$$\text{SNR} := 10 \log_{10} \left(\frac{\|f\|^2}{\|y - f\|^2} \right) \tag{10}$$

where $\|f\|^2 := \sum (f_i)^2$ and $\|y - f\|^2 := \sum (y_i - f_i)^2$. The first phantom image was based on an acquisition time of five minutes and had SNR 6.86. The second was based on an acquisition time of thirty minutes and had SNR 8.10.

Each phantom image was subjected to three different kinds of denoisings.

1) GCV: The GCV technique was applied to the whole image (embedded in a $128 \times 128 \times 128$ array) without

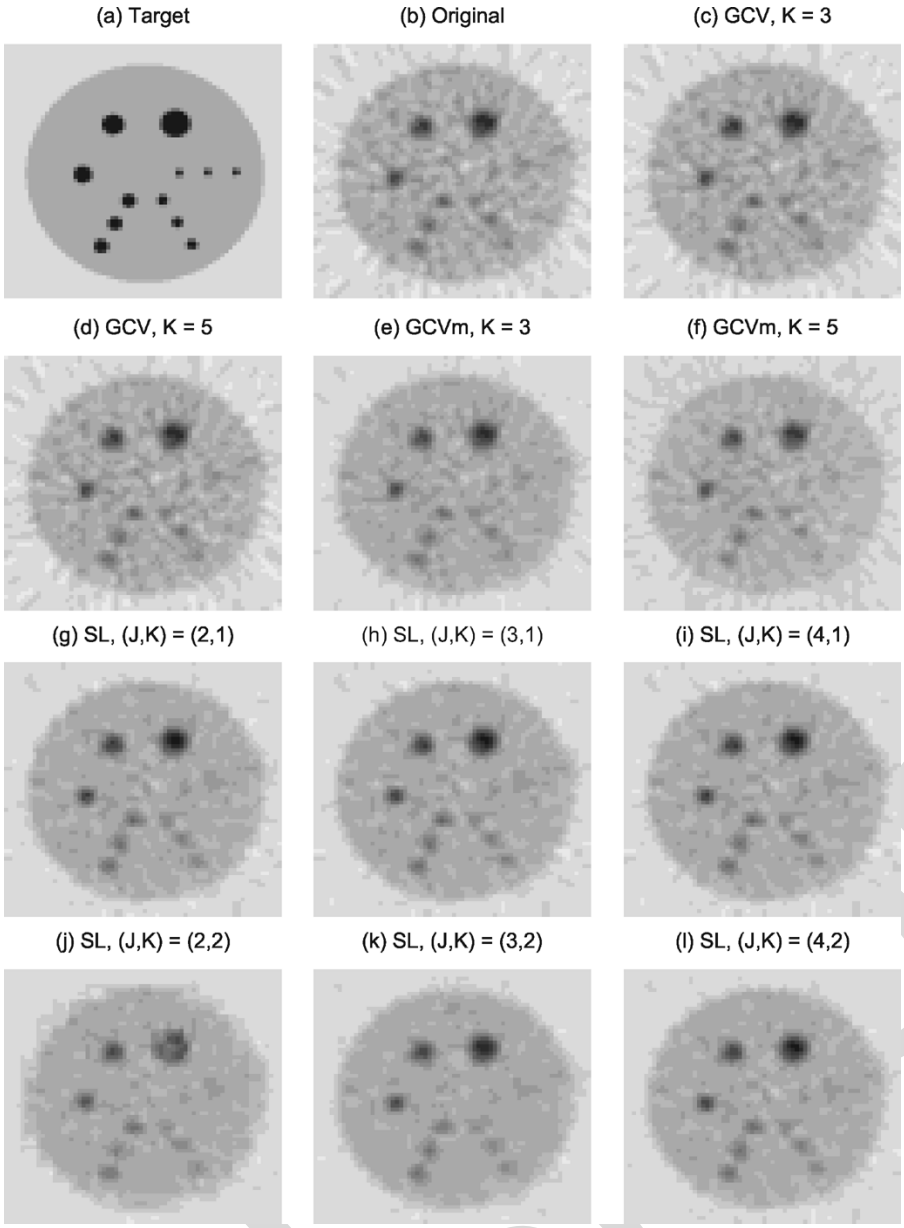

Fig. 2. Denoisings, multi-resolution activity. (a) Target; (b) original; (c) GCV, $\mathrm{K} = 3$; (d) GCV, $\mathrm{K} = 5$; (e) GCVm, $\mathrm{K} = 3$; (f) GCVm, $\mathrm{K} = 5$; (g) SL, $(\mathrm{J}, \mathrm{K}) = (2, 1)$; (h) SL, $(\mathrm{J}, \mathrm{K}) = (3, 1)$; (i) SL, $(\mathrm{J}, \mathrm{K}) = (4, 1)$; (j) SL, $(\mathrm{J}, \mathrm{K}) = (2, 2)$; (k) SL, $(\mathrm{J}, \mathrm{K}) = (3, 2)$; and (l) SL, $(\mathrm{J}, \mathrm{K}) = (4, 2)$.

modification; that is, we set $L_{l,o} := 0$. We considered $K = 1, \ldots, 5$.

2) GCVm: The GCV technique was applied to the whole image (embedded in a $128 \times 128 \times 128$ array) but modified through incorporation of nontrivial lower bounds. We set $L_{l,o} := \lambda_l q_{l,o}$ for various λ_l, where $q_{l,o}$ is the upper 0.05 quantile of the $|w_{l,o,i}|$, and we considered $K = 1, \ldots, 5$. Here, λ_l plays a role like that of the $\exp[-g_l(y_b^A, y)]$ multiplier in (5).

3) SL: The semi-local paradigm was employed with $g_l(y_b^A, y)$ as in (8) and $q_{l,o;b}$ as the upper 0.05 quantile of the $|w_{l,o,i;b}|$. We considered $(J, K) = (3,1), (3,2), (4,1), (4,2)$ with various values of ζ_1 and ζ_2, and we considered $(J, K) = (2,1), (2,2)$ with $\zeta_1 = 8000$ and $\zeta_2 = 2000$ (which were roughly the upper 0.05 and 0.20 quantiles of the values in y, respectively).

All denoisings involved $S8$ basis functions [26]. We fixed the balancing parameter P in (7) to be $P = 6$ for the first two kinds of denoisings and $P = 5$ for the third.

A. Visual Results

We present two figures that pertain to the phantom image acquired in thirty minutes. Fig. 2 depicts a horizontal slice of multi-resolution activity, while Fig. 3 displays multi-contrast activity.

The first six panels of each figure show: (a) the target image, which we produced using the phantom manufacturer's specifications; (b) the original image produced by the ECAT software; (c) the results of GCV denoising with $K = 3$; (d) the

results of GCV denoising with $K = 5$; (e) the results of modified GCV denoising with $K = 3$ ($\lambda_1 = \lambda_2 = \lambda_3 = 0.40$); and, (f) the results of modified GCV denoising with $K = 5$ ($\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0.40$).

The last six panels show the results of semi-local denoising with: (g) $(J, K) = (2, 1)$; (h) $(J, K) = (3, 1)$; (i) $(J, K) = (4, 1)$; (j) $(J, K) = (2, 2)$; (k) $(J, K) = (3, 2)$; and, (l) $(J, K) = (4, 2)$. In all cases, $\zeta_1 = 8000$; where applicable, $\zeta_2 = 2000$.

Fig. 2 shows that GCV denoising is ineffective. Modified GCV denoising with $K = 3$ is fairly effective, although the small tumors at the lower right lack definition. Modified GCV denoising with $K = 5$ oversmooths, as evidenced by the radial artifacts just outside the boundary of the phantom object. On the other hand, semi-local denoising with $(J, K) = (2, 1), (3, 1), (4, 1)$ removes a substantial amount of noise while leaving the small tumors with some definition. The result obtained from $(J, K) = (4, 1)$ is perhaps marginally better than the others and is arguably the best among all displayed in Fig. 2. Semi-local denoising with $(J, K) = (2, 2)$ yields noticeable distortions. When $J = 2$, the wavelet coefficients corresponding to each large block occupy only four resolution levels, and enough information about the true image is contained in the second resolution level that taking $K = 2$ does more harm than good. The choice of $(J, K) = (3, 2)$ does not produce such noticeable distortions but clearly oversmooths. Semi-local denoising with $(J, K) = (4, 2)$ slightly oversmooths.

Fig. 3 shows that the best results for multi-contrast activity are obtained from semi-local denoising with $(J, K) = (4, 2)$ and modified GCV denoising with $K = 3$; the tumors are slightly more pronounced in the former. Semi-local denoising with $(J, K) = (3, 2)$ produces results that are fairly appealing but perhaps slightly oversmoothed.

B. Quantitative Results

Tables I and II give percent reductions in mean squared error (MSE) for the various denoisings of the phantom images. The first number in each entry pertains to a denoising of the image acquired in five minutes, while the second number pertains to a denoising of the image acquired in thirty minutes. Calculations of MSE reductions were possible because we had available a target image.

In Table I, we see that standard application of the GCV technique has little impact; the MSE reductions are less than 7% for the image acquired in five minutes and less than 6% for the image acquired in thirty minutes. Modified GCV denoising works well with high lower bounds for $\delta_{l,o}$ at the first three resolution levels; MSE reductions of up to 45% and 34% are obtained. Thresholding at the fourth and fifth resolution levels erases information about the true image along with the noise.

In Table II, we see that semi-local denoising with $(J, K) = (3, 1)$ yields MSE reductions of up to 39% and 24%; with $(J, K) = (4, 1)$, the reductions reach 38% and 23%. Semi-local denoising with $(J, K) = (3, 2)$ yields MSE reductions of up to 53% and 39%; with $(J, K) = (4, 2)$, the reductions reach 54% and 40%. For given J and K, the results are robust over a wide range of ζ_1 and ζ_2.

Since the computation time increases about eightfold in going from $J = 3$ to $J = 2$, we limited our experimentation with $(J, K) = (2, 1)$ and $(J, K) = (2, 2)$ to working with the values $\zeta_1 = 8000$ and $\zeta_2 = 2000$. The outcomes for $(J, K) = (2, 1)$ are not demonstrably better than those for $(J, K) = (3, 1)$ and $(J, K) = (4, 1)$, while the results for $(J, K) = (2, 2)$ are plainly worse than those for $(J, K) = (3, 2)$ and $(J, K) = (4, 2)$.

C. Assessing the Results

Table III compares visual quality, MSE reduction, and MSE reduction in high-activity regions for the denoisings depicted in Figs. 2 and 3. The relationship between visual quality and MSE reduction is imperfect. Some images with lower MSE are not as good as other images with higher MSE because localized regions of high activity have been degraded, a fact that is also evident from the disparities between MSE reduction and high-activity MSE reduction.

In the context of interventional fast magnetic resonance imaging, Salem *et al.* [28] found that a measure of distance computed via a Perceptual Difference Model (PDM) better corresponded to visual quality than MSE. The PDM distance is defined for two-dimensional images and does not extend directly to three-dimensional images. However, one may compute the PDM distance for the two-dimensional cross sections of a three-dimensional image and then aggregate the results. In our experiments, we found that the aggregated PDM distance, like MSE, understated the degradation of localized regions of high activity. On the other hand, with $c = 1.9$, $SURE_{l,o;b,c}$ provided verdicts about the semi-locally denoised images that reasonably accorded with visual quality.

Using version 6.1.2 of S-PLUS for Linux on an AMD Athlon processor (858 MHz) workstation, the computation time for semi-local denoising with $(J, K) = (3, 1)$ was thirty minutes; computation times for other J and K were roughly consistent with the $K/(2^J)^D$ proportionality mentioned earlier. It is possible to reduce the computation time (e.g., by using an approximate minimizer instead of a grid search to select the thresholds), but the $K/(2^J)^D$ proportionality will not change. Thus, considering both computation time and the end products, the most viable choices of J and K for images similar to those we have examined are $(J, K) = (4, 1), (4, 2)$. In addition, the fact that superior results were not obtained when $J = 2$ suggests that a fully local paradigm for wavelet denoising might not be better than a semi-local paradigm even if it were computationally feasible.

V. Asymptotic Validity

A PET image of size $128 \times 128 \times 47$ is large because of its three-dimensionality but is not large in the asymptotics sense; with reference to the lattice $\mathcal{L}$ on which the image is defined in (1), M is only 7. However, since we will wish to employ the semi-local paradigm for other purposes, it is worthwhile to investigate asymptotic issues. Our investigation takes place in the one-dimensional uncorrelated-noise setting examined by Jansen and Bultheel [29], who described the asymptotic behavior of the risk-minimizing threshold for a certain class of functions $f(x)$ defined on [0,1]. Jansen, Malfait, and Bultheel [8] had invoked

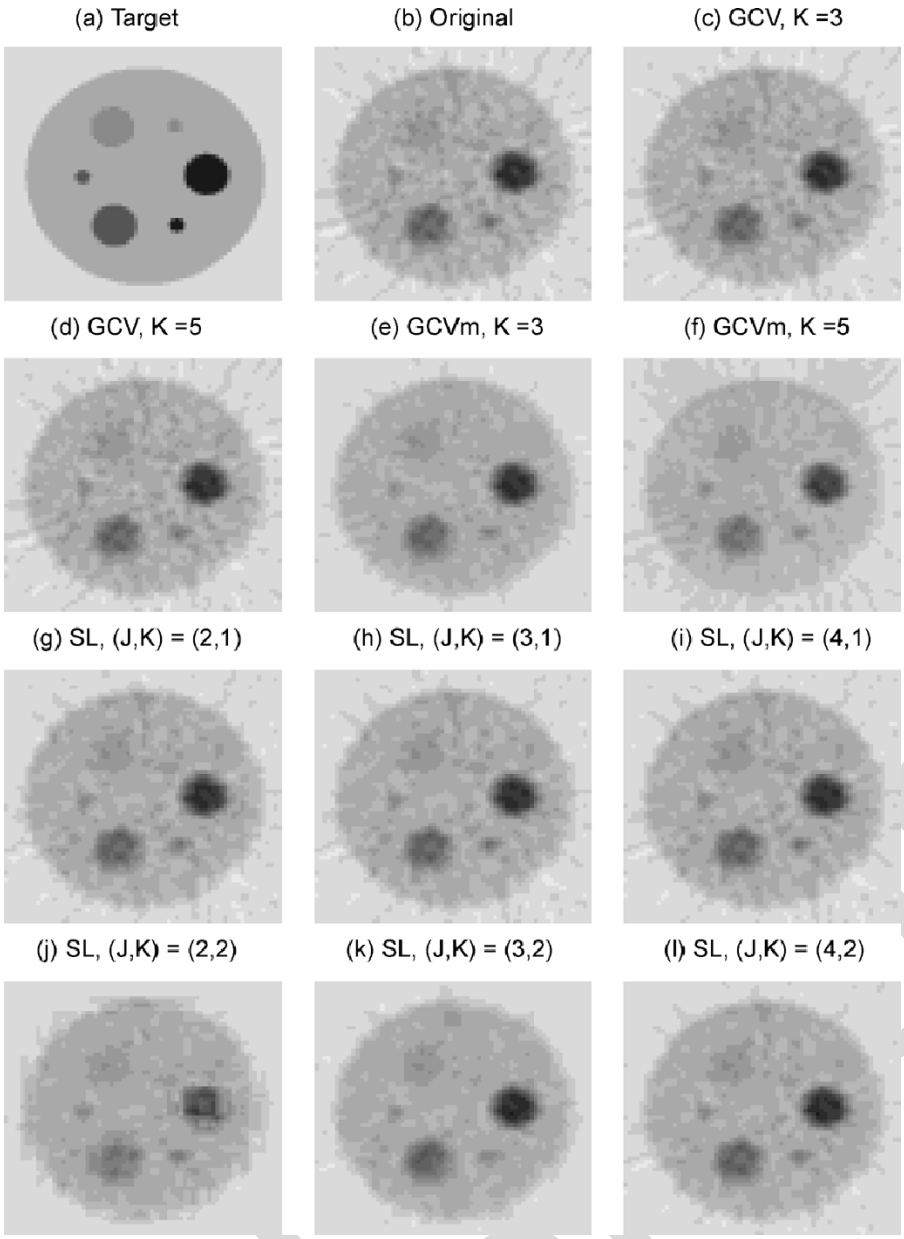

Fig. 3. Denoisings, multi-contrast activity. (a) Target; (b) original; (c) GCV, $\mathrm{K} = 3$; (d) GCV, $\mathrm{K} = 5$; (e) GCVm, $\mathrm{K} = 3$; (f) GCVm, $\mathrm{K} = 5$; (g) SL, $(\mathrm{J}, \mathrm{K}) = (2, 1)$; (h) SL, $(\mathrm{J}, \mathrm{K}) = (3, 1)$; (i) SL, $(\mathrm{J}, \mathrm{K}) = (4, 1)$; (j) SL, $(\mathrm{J}, \mathrm{K}) = (2, 2)$; (k) SL, $(\mathrm{J}, \mathrm{K}) = (3, 2)$; and (l) SL, $(\mathrm{J}, \mathrm{K}) = (4, 2)$.

this asymptotic behavior to justify the GCV technique in the one-dimensional uncorrelated-noise setting.

Let $Y = f + \epsilon$ consist of 2^M observations, defined on the lattice $\mathcal{L} := \{0, 1/2^M, \ldots, (2^M - 1)/2^M\}$. Just as in the more general setting, we divide the data into small blocks containing 2^J elements, each of which is centrally embedded in a large block containing 2^{J+1} elements; the large blocks are then denoised, following which the denoised small blocks are extracted and reassembled. Now, however, we seek a single threshold δ_b for *all* of the wavelet coefficients associated with a large block y_b^A. To make explicit the dependence on M, we hereafter write $\delta_{b,M}$ for δ_b and $y_{b,M}^A$ for y_b^A. Let $\delta_{b,M}^*$ denote the value of the threshold that minimizes the risk for denoising $y_{b,M}^A$. Similarly, let ${\delta_{b,M}^S}^*$ denote the value of the threshold that would minimize the risk for denoising the small block embedded in $y_{b,M}^A$ if we were directly denoising that small block.

Our main result is as follows. Suppose that, for some $\alpha > 0$, $f(x)$ is Lipschitz-α (refer to the Appendix). Consider a sequence of large blocks, denoted by $y_{b,M}^A$, with J dependent on M such that

$$j_M := \frac{J}{M} \to j \in (0, 1). \tag{11}$$

Relation (11) implies that the large blocks grow in size as M increases *and* that there are more large blocks as M increases. Let $\hat{\delta}_{b,M}$ be a sequence of thresholds applied to denoise the large blocks. These thresholds need not be obtained through minimization of a GCV criterion, but we assume that

$$\log \hat{\delta}_{b,M} - \log {\delta_{b,M}}^* \to 0 \tag{12}$$

TABLE II
MSE REDUCTION: FIVE MINUTES/THIRTY MINUTES (SL)

Denoising	Values of J and K			
SL	$(3,1)$	$(3,2)$	$(4,1)$	$(4,2)$
$\{\zeta_1, \quad \zeta_2\}$				
{2000,2000}	33.1 / 19.8	46.9 / 33.4	27.3 / 16.8	42.3 / 31.3
{4000,2000}	n.a.	50.0 / 35.8	n.a.	47.5 / 34.9
{4000,4000}	36.2 / 22.2	50.4 / 36.6	32.6 / 20.3	49.3 / 36.7
{6000,2000}	n.a.	51.3 / 36.9	n.a.	50.2 / 36.6
{6000,4000}	n.a.	51.7 / 37.7	n.a.	52.0 / 38.4
{6000,6000}	37.8 / 23.3	50.5 / 37.4	35.4 / 22.1	52.1 / 38.6
{8000,2000}	n.a.	52.1 / 37.5	n.a.	51.6 / 37.4
{8000,4000}	n.a.	52.5 / 38.2	n.a.	53.3 / 39.2
{8000,6000}	n.a.	51.3 / 37.9	n.a.	53.4 / 39.4
{8000,8000}	38.3 / 23.8	50.3 / 37.8	36.8 / 23.0	53.2 / 39.0
{10000,2000}	n.a.	52.4 / 37.7	n.a.	52.3 / 37.9
{10000,4000}	n.a.	52.8 / 38.5	n.a.	54.0 / 39.7
{10000,6000}	n.a.	51.6 / 38.2	n.a.	54.2 / 39.8
{10000,8000}	n.a.	50.6 / 38.0	n.a.	53.9 / 39.4
{10000,10000}	38.7 / 24.0	50.4 / 37.7	37.5 / 23.4	53.7 / 39.1

Note 1: For SL denoisings with $K = 1$, we list the results on the lines where $\zeta_1 = \zeta_2$.
Note 2: With $(J, K) = (2, 1)$ and $\zeta_1 = 8000$, MSE reductions are 38.4% and 24.4%. With $(J, K) = (2, 2)$ and $\zeta_1 = 8000, \zeta_2 = 2000$, MSE reductions are 43.3% and 20.8%.

TABLE I
MSE REDUCTION: FIVE MINUTES/THIRTY MINUTES (GCV, GCVm)

Denoising	Value of K			
GCV	2	3	4	5
	6.0 / 5.1	6.5 / 5.7	6.4 / 5.5	6.2 / 5.1
GCVm	2	3	4	5
$\{\lambda_l, \lambda_K\}$				
{0.1,0.1}	17.3 / 13.3	18.3 / 14.5	17.6 / 13.3	14.6 / 9.4
{0.2,0.1}	23.0 / 16.6	27.4 / 21.1	26.9 / 20.2	22.6 / 14.5
{0.2,0.2}	26.3 / 19.9	27.7 / 21.4	25.6 / 18.5	18.7 / 9.5
{0.3,0.1}	27.7 / 19.3	34.6 / 26.4	34.1 / 25.4	28.3 / 17.6
{0.3,0.2}	30.9 / 22.6	34.8 / 26.6	32.9 / 23.8	24.4 / 12.6
{0.3,0.3}	33.5 / 25.2	34.9 / 26.7	31.4 / 21.8	19.8 / 6.6
{0.4,0.1}	31.4 / 21.5	40.3 / 30.5	39.7 / 29.5	32.2 / 19.4
{0.4,0.2}	34.7 / 24.8	40.5 / 30.8	38.5 / 27.8	28.3 / 14.2
{0.4,0.3}	37.2 / 27.4	40.6 / 30.8	37.0 / 25.8	23.7 / 8.2
{0.4,0.4}	39.2 / 29.4	40.6 / 30.7	35.4 / 23.6	18.5 / 1.1
{0.5,0.1}	34.4 / 23.3	44.8 / 33.8	44.1 / 32.5	34.7 / 19.9
{0.5,0.2}	37.6 / 26.6	45.0 / 34.1	42.8 / 30.8	30.7 / 14.8
{0.5,0.3}	40.2 / 29.1	45.1 / 34.1	41.3 / 28.8	26.1 / 8.7
{0.5,0.4}	42.2 / 31.1	45.1 / 34.0	39.7 / 26.6	20.9 / 1.5
{0.5,0.5}	43.7 / 32.7	44.9 / 33.7	37.9 / 24.2	15.5 / -4.0

Note 1: The first number of each index $\{\lambda_l, \lambda_K\}$ is the common value for $\lambda_1, \ldots, \lambda_{K-1}$, while the second number is the value for λ_K.
Note 2: Results for $K = 1$ are 3.7/2.7% (GCV), 10.8/6.7% (GCVm, $\lambda_1 = 0.1$), 16.6/10.1% (GCVm, $\lambda_1 = 0.2$), 21.3/12.8% (GCVm, $\lambda_1 = 0.3$), 25.0/15.0% (GCVm, $\lambda_1 = 0.4$), and 28.0/16.8% (GCVm, $\lambda_1 = 0.5$).

TABLE III
VISUAL QUALITY, MSE REDUCTION, HIGH-ACTIVITY MSE REDUCTION: THIRTY MINUTES

Denoising	Visual quality	MSE red.	High MSE red.
GCV, $K = 3$		5.7	-0.5
GCV, $K = 5$		5.1	-2.0
GCVm, $K = 3$	Multi-contrast	30.7	-3.5
GCVm, $K = 5$		1.1	-28.0
SL, $(J, K) = (2, 1)$	Multi-resolution	24.4	0.1
SL, $(J, K) = (3, 1)$	Multi-resolution	23.8	0.4
SL, $(J, K) = (4, 1)$	Multi-resolution	23.0	0.3
SL, $(J, K) = (2, 2)$		20.8	-20.0
SL, $(J, K) = (3, 2)$	Multi-contrast	37.5	0.0
SL, $(J, K) = (4, 2)$	Multi-contrast	37.4	-1.1

Note 1: High activity is defined as at least four times the baseline level; this includes all tumors in a multi-resolution slice and the two tumors of highest intensity in a multi-contrast slice.
Note 2: Entries in the "Visual quality" column indicate which denoisings were judged most successful with regard to the multi-resolution and multi-contrast activity slices.

in probability as $M \to \infty$. Property (12) will be referred to as *logarithmic consistency*. Then, under conditions that we will specify below, the semi-local sequence

$$\tilde{\delta}^S_{b,M} := \max\{L_{b,M}, \hat{\delta}_{b,M}\} \tag{13}$$

is logarithmically consistent for ${\delta^S_{b,M}}^*$, where

$$L_{b,M} := q_{b,M} \exp\left[-g\left(y^A_{b,M}, y\right)\right]. \tag{14}$$

Here, $q_{b,M}$ is (for some fixed γ) the upper γ quantile of the $|w_{l,o,i;b,M}|$ across *all* resolution levels, and $g(y^A_{b,M}, y)$ is a non-negative function of $y^A_{b,M}$ and y. We will refer to $\tilde{\delta}^S_{b,M}$ as the *semi-local estimator* based on the *underlying estimator* $\hat{\delta}_{b,M}$.

Remark: The motivation for logarithmic consistency is that the optimal values ${\delta_{b,M}}^*$ and ${\delta^S_{b,M}}^*$ escalate with M, so that a reasonable estimator $\hat{\delta}_{b,M}$ may approach ${\delta_{b,M}}^*$ in relative terms (i.e., on the logarithmic scale) but not in absolute terms (i.e., on the ordinary scale). On the other hand, logarithmic consistency is a stronger requirement than mere attainment of the correct asymptotic order; for instance, $\hat{\delta}_{b,M} = O_p(\sqrt{M})$ does not imply the logarithmic consistency of $\hat{\delta}_{b,M}$.

Theorem 5.1: Consider a sequence of large blocks $y^A_{b,M}$, each defined on a subset of $\mathcal{L}$. Assume that $f(x)$ is Lipschitz-α, that J grows as indicated in (11), and that the noise term ϵ is Gaussian with covariance matrix $\sigma^2 I_{2^M}$, where I_{2^M} is the identity matrix of dimension 2^M. Assume also that the mother wavelet function has at least $\lceil\alpha\rceil$ vanishing moments, where $\lceil\alpha\rceil$ denotes the smallest integer greater than or equal to α.

Let $\hat{\delta}_{b,M}$ be logarithmically consistent for ${\delta_{b,M}}^*$. Then $\tilde{\delta}^S_{b,M}$ as defined by (13), with lower bound sequence $L_{b,M}$ as defined by (14), is logarithmically consistent for ${\delta^S_{b,M}}^*$.

Theorem 5.1 is, in fact, valid for any other choice of lower bounds $\tilde{L}_{b,M}$ satisfying $\tilde{L}_{b,M} = o_p(\sqrt{M})$. The proof of Theorem 5.1 is provided in the Appendix along with the definition of the Lipschitz-α condition.

We have established that, in the one-dimensional uncorrelated-noise setting of [8] and [29], the semi-local estimator inherits logarithmic consistency from its underlying estimator. Semi-local denoising will enjoy logarithmic consistency in more general settings provided that: (a) the underlying estimators at each level and orientation are logarithmically consistent; and, (b) the function $f(x)$ is such that analogues to (17) and (18) in the Appendix hold at each level and orientation.

VI. Discussion

The semi-local paradigm provides a computationally feasible way to vary the degree of denoising across different regions of an image without *a priori* knowledge of the image's structures. Our experiments have shown that semi-local denoising is superior to standard application of the GCV technique for images possessing structures with multi-resolution heterogeneity (mimicking tumors of different sizes) or multi-contrast heterogeneity (mimicking tumors of different sizes and intensities). Under some regularity conditions, semi-local denoising provides logarithmically consistent estimators of risk-minimizing thresholding parameters.

While we emphasized PET images in this paper, the semi-local paradigm is applicable to medical images obtained via other modalities as well as to images that arise in nonmedical applications. We note here that our initial assumption of stationary noise can be weakened; since a modification of the GCV technique is being applied to the large blocks rather than to the whole image, we only need for the noise to be approximately stationary in the large blocks.

In modifying the GCV technique of Jansen and Bultheel [1], we adopted their choice to use a soft thresholding rule instead of a hard thresholding rule; as indicated in Jansen, Malfait, and Bultheel [8], the asymptotic justification of the GCV technique depends on the soft thresholding rule. However, a hard thresholding rule may be more reasonable in some applications, as it may be undesirable to shrink large wavelet coefficients at all; in such applications, a proxy for the risk other than the GCV criterion would be needed to make a fair comparison between soft and hard thresholding. More generally, continuing study of asymptotic issues for semi-local denoising is needed.

Our experiments with the phantom PET images suggest that semi-local denoising is robust with respect to the scaling factor ζ_l appearing in (8). On the other hand, the choices of J and K are crucial. Thus, it is advisable to seek guidance from pilot experiments or to compute a quantity like $SURE_{l,o;b,c}$. Regarding $SURE_{l,o;b,c}$, further research is needed to: (a) determine a general procedure for selecting c; and, (b) establish whether taking data-dependent values of $\delta_{l,o;b}$ as inputs causes $SURE_{l,o;b,c}$ to be differentially optimistic for different choices of J and K. Issue (b) relates to the idea of "predictive" error estimation in data mining, where using the same data to both fit and evaluate a model may result in an unduly favorable evaluation. Some practical suggestions, in any case, are to: i) carry out and view multiple denoisings and ii) carefully explore both the original and denoised images to identify any features that are not physically reasonable. With respect to i), we envisage software employing a "sliding" mechanism that would allow practitioners to move through various semi-locally denoised images in a continuous manner. The relevance of ii) was illustrated in our experiments, where we identified distortions that were manifested when $(J, K) = (2, 2)$.

A natural extension of the semi-local paradigm would allow K to vary from one large block to another. Fan *et al.* [30] have remarked that it would be desirable but impractical to change the number of levels at which thresholding is done across different regions of an image. With the semi-local paradigm, however, this is possible. The sensitivity of our experimental results to the choice of K indicates that care must be taken in developing the procedure by which K would be selected for each large block.

Details of the Theoretical Justification

Appendix

We begin by defining the Lipschitz-α condition (see [29]).

Definition 1.1: We say that $f(x)$ is *Lipschitz-α* on [0,1] if there exists a positive constant C such that, for each $t \in [0,1]$, there exists a polynomial $p_t(x)$ with $|f(x) - p_t(x)| \le C|t-x|^\alpha$ for all $x \in [0,1]$.

We are now ready to prove Theorem 5.1; the intermediate results used in the proof are derived afterward.

Proof of Theorem 5.1: Let $\xi > 0$ be given. We have

$$\begin{aligned} &P\left\{\left|\log \tilde{\delta}^S_{b,M} - \log {\delta^S_{b,M}}^*\right| \ge \xi\right\} \\ &\quad \le P\left\{\tilde{\delta}^S_{b,M} \ne \hat{\delta}_{b,M}\right\} + P\left\{\left|\log \hat{\delta}_{b,M} - \log {\delta_{b,M}}^*\right| \ge \frac{\xi}{2}\right\} \\ &\qquad + P\left\{\left|\log {\delta_{b,M}}^* - \log {\delta^S_{b,M}}^*\right| \ge \frac{\xi}{2}\right\}. \end{aligned}$$

Corollary 1.1, the logarithmic consistency of $\hat{\delta}_{b,M}$, and Proposition 1.2 below imply that each term on the right side of the inequality converges to zero. QED

Jansen and Bultheel [29] established that the optimal threshold for denoising the full data set is

$$\sqrt{\frac{4\alpha}{2\alpha+1}\sigma^2 M \log 2}\,[1 + o(1)]. \tag{15}$$

Let $I \subset [0,1]$ be an interval of length 2^{-E} for some positive integer E. The arguments used to derive (15) imply that the optimal threshold for denoising the part of the data set defined on $I \cap \mathcal{L}$ is

$$\sqrt{\frac{4\alpha}{2\alpha+1}\sigma^2 (M-E) \log 2}\,[1 + o(1)]. \tag{16}$$

Assuming that the $o(1)$ term in (16) is uniform across all I associated with the large blocks $y^A_{b,M}$, we conclude that

$$\delta_{b,M}{}^* = \sqrt{\frac{4\alpha}{2\alpha+1}\sigma^2(j_M M+1)\log 2}\,[1+o(1)] \qquad (17)$$

which yields the following proposition.

Proposition 1.1: Let $\tilde{L}_{b,M}$ be any sequence such that $\tilde{L}_{b,M} = o_p(\sqrt{M})$. Then, under the conditions of Theorem 5.1,

$$P\left\{\log \tilde{L}_{b,M} < \log \delta_{b,M}{}^* - 1\right\} \to 1 \text{ as } M \to \infty.$$

Proof: The desired result follows immediately from (17). QED

Using a result by Jaffard [3] and Theorem 1 of Jansen and Bultheel [29], we calculate that

$$|v_{l,o,i;b,M}| \le C_M 2^{\left(\frac{l}{2}\right)-(j_M M+1-l)\alpha} \qquad (18)$$

for coefficients $v_{l,o,i;b,M}$ corresponding to the Lipschitz-α pieces of $f(x)$, where $C_M = O(1)$. In establishing their Theorem 1, Jansen and Bultheel [29] assumed that the uncorrupted wavelet coefficients could be calculated via a continuous wavelet transform applied to $f(x)$ rather than via a discrete wavelet transform applied to f. Assuming that (18) also holds under the discrete wavelet transform, we obtain the following corollary.

Corollary 1.1: With $L_{b,M}$ as indicated in (14), we have

$$P\left\{\log L_{b,M} < \log \delta_{b,M}{}^* - 1\right\} \to 1$$

and

$$P\left\{\tilde{\delta}^S_{b,M} = \hat{\delta}_{b,M}\right\} \to 1,$$

where $\hat{\delta}_{b,M}$ and $\tilde{\delta}^S_{b,M}$ are as in (12) and (13).

Proof: Choose l^* such that $\sum_{l=1}^{l^*} 2^{-l} > 1-\gamma/2$. We have $2^{(l/2)-(j_M M+1-l)\alpha} \to 0$ uniformly for $l \le l^*$ as $M \to \infty$. Hence, the upper $\gamma/2$ quantile of the $|w_{l,o,i;b,M}|$ with $l \le l^*$ is $O_p(1) = o_p(\sqrt{M})$. Since the upper $\gamma/2$ quantile of the $|w_{l,o,i;b,M}|$ with $l \le l^*$ is greater than or equal to the upper $\gamma - \gamma^2/4$ quantile of the $|w_{l,o,i;b,M}|$ without restriction on l, it follows that $q_{b,M}$ is $O_p(1) = o_p(\sqrt{M})$. This fact and the logarithmic consistency of $\hat{\delta}_{b,M}$ imply the desired results.QED

Our second proposition concerns the relationship between $\delta_{b,M}{}^*$ and $\delta^S_{b,M}{}^*$.

Proposition 1.2: Under the conditions of Theorem 5.1,

$$\log \delta_{b,M}{}^* - \log \delta^S_{b,M}{}^* \to 0.$$

Proof: The same reasoning that leads to (17) yields

$$\delta^S_{b,M}{}^* = \sqrt{\frac{4\alpha}{2\alpha+1}\sigma^2 j_M M \log 2}\,[1+o(1)].$$

Now the proposition follows from the fact that

$$\log\left(\frac{j_M M+1}{j_M M}\right) \to 0.$$

QED.

Remark: With regard to the semi-local estimator, one may adopt the perspective that logarithmic consistency for $\delta_{b,M}{}^*$ is more relevant than logarithmic consistency for $\delta^S_{b,M}{}^*$ because we are thresholding wavelet coefficients from the large blocks to denoise the small blocks rather than thresholding wavelet coefficients from the small blocks themselves. Proposition 1.2 implies that the semi-local estimator is also consistent for $\delta_{b,M}{}^*$, so that semi-local denoising is justified under this perspective as well.

ACKNOWLEDGMENT

The authors acknowledge the following individuals, who provided software, useful information, or helpful discussion: M. Casey, P. Faulhaber, D. Huo, M. Jansen, C. Loader, L. Mehta, G. Nason, R. S. Pilla, K. Salem, H. Turbell, and D. Wilson. They also thank T. Nguyen and two anonymous referees.

REFERENCES

[1] M. Jansen and A. Bultheel, "Multiple wavelet threshold estimation by generalized cross validation for images with correlated noise," *IEEE Trans. Image Processing*, vol. 8, pp. 947–953, Jul. 1999.
[2] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, Jul. 1989.
[3] S. Jaffard, "Pointwise smoothness, two-microlocalization and wavelet coefficients," *Publications Matematiques*, vol. 35, pp. 155–168, 1991.
[4] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[5] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[6] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.
[7] G. P. Nason, "Wavelet shrinkage using cross validation," *J. R. Statist. Soc. B*, vol. 58, pp. 463–479, 1996.
[8] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Process.*, vol. 56, pp. 33–44, 1997.
[9] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. R. Statist. Soc. B*, vol. 59, pp. 319–351, 1997.
[10] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *J. Amer. Statist. Assoc.*, vol. 96, pp. 939–967, 2001.
[11] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. Image Processing*, vol. 10, pp. 1322–1331, Sep. 2001.
[12] M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Statist. Assoc.*, vol. 96, pp. 629–639, 2001.
[13] I. M. Johnstone and B. W. Silverman, "Empirical Bayes Selection of Wavelet Thresholds," Stanford Univ., Stanford, CA, 2002.
[14] G. P. Nason, "Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage," *Statist. Comput.*, vol. 12, pp. 219–227, 2002.
[15] E. J. Candes and D. L. Donoho, "Recovering edges in ill-posed inverse problems: optimality of curvelet frames," *Ann. Statist.*, vol. 30, pp. 784–842, 2002.
[16] S. E. Ferrando and L. A. Kolasa, "Averages of best wavelet basis estimates for denoising," *J. Comput. Appl. Math.*, vol. 136, pp. 357–367, 2001.
[17] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522–1531, Sep. 2000.
[18] L. Sendur and I. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Lett.*, vol. 9, pp. 438–441, 2002.
[19] T. Cai, "Adaptive wavelet estimation: a block thresholding and oracle inequality approach," *Ann. Statist.*, vol. 27, pp. 898–924, 1999.
[20] P. Hall, G. Kerkyacharian, and D. Picard, "On the minimax optimality of block thresholded wavelet estimators," *Statistica Sinica*, vol. 9, pp. 33–50, 1999.

[21] G. Chen and T. Bui, "Multiwavelets denoising using neighboring coefficients," *IEEE Signal Processing Lett.*, vol. 10, pp. 211–214, Jul. 2003.
[22] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.
[23] H. Akaike, "Information theory and the maximum likelihood principle," in *Proc. 2nd Int. Symp. Information Theory*, B. N. Petrov and F. Csaki, Eds., Budapest, 1973.
[24] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.
[25] X. Shen, H.-C. Huang, and N. Cressie, "Nonparametric hypothesis testing for a spatial signal," *J. Amer. Statist. Assoc.*, vol. 97, pp. 1122–1140, 2002.
[26] I. Daubechies, "Orthonormai bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.
[27] K. Wienhard, L. Eriksson, S. Grootoonk, M. Casey, U. Pietrzyk, and W. D. Heiss, "Performance evaluation of the positron scanner ECAT EXACT," *J. Comput. Assist. Tomogr.*, vol. 16, pp. 804–813, 1992.
[28] K. Salem, J. Lewin, A. Aschoff, J. Duerk, and D. Wilson, "Validation of a human vision model for image quality evaluation of fast interventional magnetic resonance imaging," *J. Electron. Imag.*, vol. 11, pp. 224–235, 2002.
[29] M. Jansen and A. Bultheel, "Asymptotic behavior of the minimum mean squared error threshold for noisy wavelet coefficients of piecewise smooth signals," *IEEE Trans. Signal Processing*, vol. 49, pp. 1113–1118, Jun. 2001.
[30] J. Fan, P. Hall, M. Martin, and P. Patil, "Adaptation to high spatial inhomogeneity using wavelet methods," *Statistica Sinica*, vol. 9, pp. 85–102, 1999.
[31] M. Jansen, GCV Software, 2000.
[32] G. P. Nason, WaveThresh3 Software, 1998.
[33] K. Salem, Perceptual Difference Model Software (Version 1.4), 2001.

Richard Charnigo. Please provide biographical information.

Jiayang Sun. Please provide biographical information.

Raymond Muzic. Please provide biographical information.