Hyperanalytic Denoising

Sofia C. Olhede

Abstract—**A new threshold rule for the estimation of a deterministic image immersed in noise is proposed. The full estimation procedure is based on a separable wavelet decomposition of the observed image, and the estimation is improved by introducing the new threshold to estimate the decomposition coefficients. The observed wavelet coefficients are thresholded, using the magnitudes of wavelet transforms of a small number of "replicates" of the image. The "replicates" are calculated by extending the image into a vector-valued hyperanalytic signal. More than one hyperanalytic signal may be chosen, and either the hypercomplex or Riesz transforms are used, to calculate this object. The deterministic and stochastic properties of the observed wavelet coefficients of the hyperanalytic signal, at a fixed scale and position index, are determined. A "universal" threshold is calculated for the proposed procedure. An expression for the risk of an individual coefficient is derived. The risk is calculated explicitly when the "universal" threshold is used and is shown to be less than the risk of "universal" hard thresholding, under certain conditions. The proposed method is implemented and the derived theoretical risk reductions substantiated.**

Index Terms—**Hilbert transform, image denoising, 2-D analytic, wavelets.**

I. INTRODUCTION

THIS paper treats the problem of estimating an unknown deterministic image immersed in noise, based on estimating a separable wavelet decomposition of the observed image, by using the hyperanalytic extension. Global image "replicates" that can be argued to have the same local structure as the observed image are defined from the vector-valued hyperanalytic extension. The 2-D wavelet transform of the vector-valued signal is calculated. If the magnitude of the wavelet coefficients of the hyperanalytic signal exceeds a given threshold, at a fixed local index value, the wavelet coefficient of the deterministic image is estimated by the observed wavelet coefficient, otherwise it is estimated by zero. The distribution of this magnitude can be determined, and a suitable threshold chosen. Given the estimated wavelet coefficients, the image is estimated by inverting the wavelet transform. Theoretical risk reductions are obtained for the estimation of individual coefficients, under certain conditions.

In 1-D signal estimation, Donoho and Johnstone [1] first proposed estimation procedures based on wavelet decompositions. Heuristically, such estimation procedures correspond to separating the coefficients into "signal dominated" and "noise dominated" coefficients. The "noise dominated" coefficients are eliminated (thresholded), or subjected to some other form of shrinkage [2], to estimate the signal coefficients with reduced mean square error (MSE). A possible choice of threshold is to use the "universal" threshold, constant across coefficients. The "universal" threshold for large sample sizes gives a risk close to that given by using an "oracle" [1]. If the decomposition is highly compressed, hard thresholding combined with the "universal" threshold will achieve good estimation in terms of low MSE.

Unfortunately, 2-D separable wavelets do not provide optimal compression for some simple 2-D signal structures that are often observed in images. Alternative methods have been developed to achieve better compression of images containing simple local structures, i.e., work by Starck *et al.* [3]. Methods that achieve a substantial degree of compression, can afford to treat each decomposition coefficient individually, and without a great deal of sophistication. To achieve better estimation of the deterministic image, more sophisticated techniques for the estimation of the coefficients, may also be used.

As an example of the latter, using the tendency of signal structure to produce nonzero wavelet coefficients in connected regions of space and scale, improves estimation. Examples of such methods include work by Cai and Silverman [4], Dragotti and Vetterli [5], Pižurica *et al.* [6], Crouse *et al.* [7], Fryzlewicz [8], and Olhede and Walden [9]. Using such methods may give substantial estimation improvements, in terms of improved visual reconstructions, as realistic image structures are represented by a complete set of image coefficients. The aforementioned methods may also give reduced MSE.

In 1-D, Olhede and Walden [9] aimed to improve the estimation of a wavelet coefficient, by using the local structure of the signal. They calculated a second signal from the observed signal, that roughly had the same local first order structure, using the Hilbert transform (HT). Furthermore, the wavelet coefficients of the observed signal and its HT at a fixed scale and time are (up to $O(1/N)$) uncorrelated. Olhede and Walden thresholded each wavelet coefficient, depending on the combined observed magnitude of the wavelet coefficient, and the magnitude of the wavelet coefficient of the HT of the signal. The method is denoted "analytic" denoising, as it thresholds a coefficient given the value of the decomposition of the analytic signal [10], at that local point. The procedure, as both the HT and the wavelet transform are linear transformations and commute, can equivalently be viewed as using a second variable that is formed by a weighted combination of coefficients at the same scale, that are nearby in time, when deciding to threshold.

For "analytic" thresholding to work well, when estimating signal dominated coefficients, the magnitude of the wavelet coefficient of the HT of the deterministic signal must be simultaneously large, to that of the wavelet coefficient of the signal. As the

The author is with the Department of Mathematics, Imperial College London, London SW7 2AZ U.K. (e-mail: s.olhede@imperial.ac.uk).
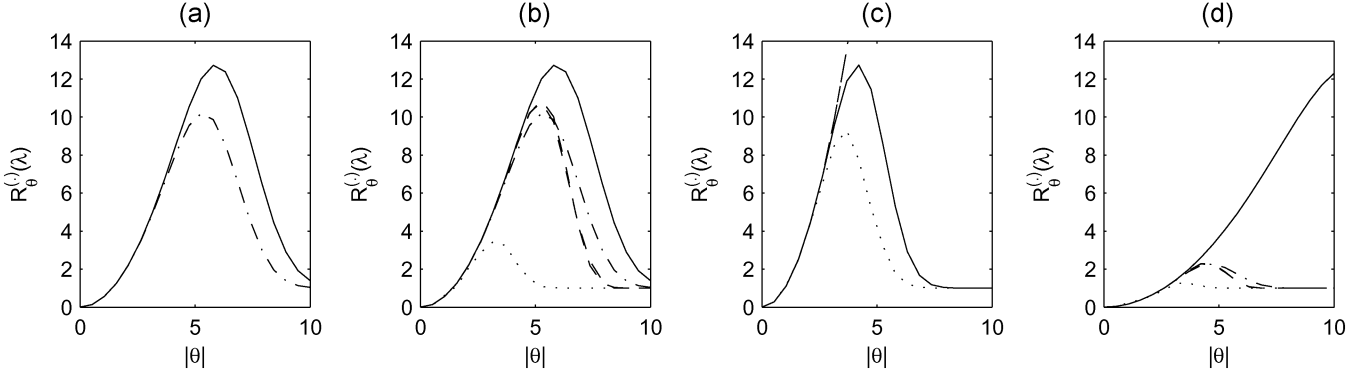
Fig. 1. Risk of a standardized (see Section III-F) estimated wavelet coefficient using a "universal" thresholding rule. The risk of (dash-dotted line) "analytic" hard thresholding and the risk of (solid line) classical hard thresholding [see subplot (a)] are plotted for $s = c$, a, $\theta_{\boldsymbol{\xi},1}^{(s)} = \theta_{\boldsymbol{\xi},2}^{(a)} = |\theta|/\sqrt{2}$. The risk of (dash-dotted) "analytic" hard thresholding, (dashed lines) Riesz thresholding, (dotted line) hypercomplex thresholding, and (solid line) classical hard thresholding are plotted in subplots (b)–(d). In subplot (b), $\theta_{\boldsymbol{\xi},1}^{(r)} = \sqrt{\theta_{\boldsymbol{\xi},2}^{(r)2} + \theta_{\boldsymbol{\xi},3}^{(r)2}} = |\theta|/\sqrt{2}$ for the Riesz threshold, while $\theta_{\boldsymbol{\xi},1}^{(h)} = \theta_{\boldsymbol{\xi},2}^{(h)} = \theta_{\boldsymbol{\xi},3}^{(h)} = \theta_{\boldsymbol{\xi},4}^{(h)} = |\theta|/\sqrt{2}$, for the Hypercomplex threshold. In plot (c), $\theta_{\boldsymbol{\xi},2}^{(r)} = \theta_{\boldsymbol{\xi},3}^{(r)} = 0$ for the Riesz threshold while for $s = r$, h, $\theta_{\boldsymbol{\xi},1}^{(s)} = \theta_{\boldsymbol{\xi},2}^{(s)} = |\theta|/\sqrt{2}$ and $\theta_{\boldsymbol{\xi},3}^{(s)} = \theta_{\boldsymbol{\xi},4}^{(s)} = 0$ for the hypercomplex threshold. In plot (d), $\theta_{\boldsymbol{\xi},1}^{(s)} = |\theta| \cos(3\pi/8)$, $\theta_{\boldsymbol{\xi},2}^{(r)} = \theta_{\boldsymbol{\xi},3}^{(r)} = |\theta| \sin(3\pi/8)/\sqrt{2}$ for the Riesz threshold, while $\theta_{\boldsymbol{\xi},1}^{(h)} = \theta_{\boldsymbol{\xi},2}^{(h)} = |\theta| \cos(3\pi/8)$ and $\theta_{\boldsymbol{\xi},3}^{(h)} = \theta_{\boldsymbol{\xi},4}^{(h)} = |\theta| \sin(3\pi/8)$ for the Hypercomplex threshold. For the "analytic" procedure in (b)–(d), we use $\theta_{\boldsymbol{\xi},2}^{(a)} = \theta_{\boldsymbol{\xi},4}^{(h)}$ as the single quadrature component, corresponding to the total Hilbert transform.

HT can be considered to have the same time-frequency structure as the original signal [11], this can be argued. When estimating the noise dominated coefficients, the expected value of the combined magnitudes of the coefficients needs to be small, for the procedure to perform well. To determine a suitable threshold, the (approximate) distribution of the magnitude, must be known. For large sample sizes, using the analytic signal, this is the case. Fig. 1(a) shows the risk of a coefficient using "analytic" denoising, based on the wavelet transform of the signal and its HT taking the same magnitude, under different signal-to-noise ratios (SNRs) [θ denotes the standardized mean (see Section III-F)]. Fig. 1(a) verifies that with a "universal" threshold, the risk of an "analytic" hard thresholded coefficient estimate, is less than that of a hard thresholded coefficient estimate.

We seek to extend "analytic" thresholding to 2-D, and this will require defining additional images, based on 2-D "analytic" signals. The additional images need to have the same local structure as the original image, and suitable stochastic properties. There are several 2-D extensions to the HT, where more than a single additional image, is usually defined. We refer to any complete set of such images, as a set of *quadrature components*. We use two special sets: the Riesz transforms (RTs, Section II-C), and the hypercomplex transforms (HCTs, Section II-D), of the image. We define the local magnitude of the wavelet coefficients using a set of quadrature components (Section III-A), that together with the original image form a hyperanalytic signal. We propose a threshold criterion, to estimate the wavelet coefficients of the image, depending on the value of the magnitude of the wavelet transform of the hyperanalytic signal. Once the wavelet transform is inverted, this yields an estimate of the image, and the method is denoted hyperanalytic denoising.

We discuss the form of the local magnitude for stylized image structure: i.e., the behavior of the threshold criterion for oscillatory structures and edges (Section III-B). The approximate distribution of the decomposition of the Riesz and Hypercomplex components of noise alone is determined, at a fixed value of the local index (Section III-C). This permits the calculation of

"universal" thresholds, for both the RT and HCT based methods (Section III-D and E). The risks of the two different thresholding strategies are calculated for a given threshold, (Section III-F), where under certain conditions the risk associated with hyperanalytic thresholding, is less than that of hard thresholding. We implement the procedure (see Section IV), and compare results with the hidden Markov model method (HMM) [7], and hard thresholding. We observe a reduced MSE in several examples, when using the proposed image denoising strategies, and observe improvements in the visual reconstructions. Hyperanalytic denoising is thus shown to give a simple and competitive method of improving existing denoising strategies.

II. IMAGE MODEL AND NONSTATIONARY STRUCTURE

A. Image Structure

We model the observed image $[Y_{\mathbf{x}}]_{\mathbf{x}}$ for $\mathbf{x} = [x_1, x_2]^T \in D$, where $D = [0, N-1]^2$, and Δx denotes the sampling period via

$$Y_{\mathbf{x}} = q(x_1 \Delta x, x_2 \Delta x) + \epsilon_{\mathbf{x}}, \quad \mathbf{x} \in D. \qquad (1)$$

We collect the observed image in a matrix $\boldsymbol{Y} = [Y_{\mathbf{x}}]_{\mathbf{x} \in D}$, and similarly define $\boldsymbol{q} = [q_{\mathbf{x}}]_{\mathbf{x} \in D} = [q(x_1 \Delta x, x_2 \Delta x)]_{\mathbf{x} \in D}^T$, as well as $\boldsymbol{\epsilon} = [\epsilon_{\mathbf{x}}]_{\mathbf{x} \in D}$. The noise is modeled by $\epsilon_{\mathbf{x}} \sim N(0, \sigma^2)$, where $\sim$ denotes distributed as, and $\text{Cov}(\epsilon_{\mathbf{x}}, \epsilon_{\mathbf{y}}) = \sigma^2 \delta_{\mathbf{x},\mathbf{y}}$, $\mathbf{x}$, $\mathbf{y} \in D$. A decomposition of the image in terms of a wavelet basis is formed [12]. We take $\psi_{j,1,\mathbf{k}}(\mathbf{x})$, $\psi_{j,2,\mathbf{k}}(\mathbf{x})$, $\psi_{j,3,\mathbf{k}}(\mathbf{x})$, and $\psi_{j,4,\mathbf{k}}(\mathbf{x}) = \phi_{j,\mathbf{k}}(\mathbf{x})$ as the tensor products of functions $\psi_{j,k}(x)$ and $\phi_{j,k}(x)$. If we let $W_{j,u,\mathbf{k}}^{(q)}$ denote the discrete wavelet transform (DWT) of $q(\cdot)$ with tensor product wavelet u, at scale j, and position $\mathbf{k} = [k_1, k_2]^T$, i.e., the inner product of $q(\mathbf{x})$ with the wavelet $\psi_{j,u,\mathbf{k}}(\mathbf{x})$, then we may represent $q(\cdot)$ as

$$q(\mathbf{x}\Delta x) = \sum_{j,\mathbf{k}} W_{j,1,\mathbf{k}}^{(q)} \psi_{j,1,\mathbf{k}}(\mathbf{x}) + \sum_{j,\mathbf{k}} W_{j,2,\mathbf{k}}^{(q)} \psi_{j,2,\mathbf{k}}(\mathbf{x})$$
$$+ \sum_{j,\mathbf{k}} W_{j,3,\mathbf{k}}^{(q)} \psi_{j,3,\mathbf{k}}(\mathbf{x}) + \sum_{\mathbf{k}} V_{J,\mathbf{k}}^{(q)} \phi_{J,\mathbf{k}}(\mathbf{x}). \qquad (2)$$

$V_{J,\mathbf{k}}^{(q)} \equiv W_{J,4,\mathbf{k}}^{(q)}$ is then associated with smooth variation in the image $q(\mathbf{x})$, $W_{j,2,\mathbf{k}}^{(q)}$ is associated with smooth variation in x_2 and rapid variation in x_1, etc. j denotes scale 2^j, where $1 \leq j \leq J_0 = \log_2(N)$, while $\mathbf{k}$ denotes a spatial localization in the plane. If an image with N^2 coefficients is observed, then for any fixed value j, $0 \leq k_l \leq N_j - 1$, $l = 1, 2$, where $N_j = N/2^j$. We collect the local indices in a vector-valued index of $\boldsymbol{\xi} = [j, u, \mathbf{k}]^T$.

The DWT is implemented by repeated filtering of the observed signal with two filters, the scaling filter $\{g_l : l = 0, \ldots, \tilde{L} - 1\}$ and the wavelet filter $\{h_l : l = 0, \ldots, \tilde{L} - 1\}$, in both spatial directions separately [13], where $\tilde{L}$ denotes the length of the filters $\{h_l\}$ and $\{g_l\}$. The transform is initialized by equating the image to the finest scale representation of the image, i.e., $V_{0,\mathbf{x}}^{(q)} \equiv q_{\mathbf{x}}$. The transform at index $\boldsymbol{\xi}$ can also be implemented using a single filter $\{h_{\boldsymbol{\xi}}\}$. The decomposition will be halted at level $j = J \leq J_0 = \log_2(N)$, and the scaling coefficients $\{V_{J,\mathbf{k}}^{(q)}\}$ are determined at this level to complete the representation. Hence, for $j < J$, only $\{W_{j,u,\mathbf{k}}^{(q)}\}$, for $u = 1, 2, 3$, are calculated. For more details on the DWT (see, for example, Percival and Walden [13]), while a good exposition of image decompositions can be found in Mallat [12].

Having observed $Y_{\mathbf{x}}$, rather than $q_{\mathbf{x}}$, we calculate the DWT coefficients $\{W_{\boldsymbol{\xi}}^{(Y)}\}$, and threshold these to obtain an estimate of $\{W_{\boldsymbol{\xi}}^{(q)}\}$, denoted $\{\widehat{W}_{\boldsymbol{\xi}}^{(q)}\}$. Wavelets will compress images of sufficient regularity, a statement that can be made precise in terms of Besov spaces. For some locally simple image structures, a more compressed representation can be made [3]. Hence, for images containing, say edges, the deterministic image energy in the DWT will be spread over more coefficients than strictly necessary. As the expected magnitude of the signal generated coefficients in the less than optimally compressed representation will be less than the mean magnitude of coefficients representing the same signal structure in a more compressed alternative decomposition, it is important that the estimation procedure does not fail to retain signal generated coefficients.

B. Quadrature Components

In a version of the 1-D estimation algorithms for the estimation of a time domain signal suggested by Cai and Silverman [4], a wavelet coefficient at scale j and position k, was estimated using a shrinkage rule. The rule was defined in terms of the combined magnitude of the observed coefficient at $[j, k]$, and the magnitudes of the immediate time-neighbors at the same scale, i.e., at $[j, l]$ for $l \neq k$. This procedure performs well if a signal contribution present at the $[j, k]$ index exhibits clustering in adjacent coefficients, i.e., the wavelet coefficients will also have large means at $[j, l]$. Furthermore, using the DWT at other time points, the noise of the other wavelet coefficients is uncorrelated, over $l \neq k$. The procedure is, thus, based on determining an additional set of variables, where the full joint distribution of the additional variables and the coefficient to be estimated, is tractable, and the full set of variables aid the estimation of the single coefficient. In 1-D, Olhede and Walden [9] defined a single additional random variable at each $[j, k]$ using the HT, to serve the same purpose as the set of DWTs in Cai and Silverman

[4]. This did not necessitate assuming local homogeneity of coefficients in time, but was based on the local properties of the HT. We shall discuss the mechanism of "analytic" denoising in some detail, to generalise the concepts involved to 2-D.

Let the Fourier transform (FT) of a d-dimension signal $q(\mathbf{x})$ be denoted by

$$Q(\mathbf{f}) = \int\limits_{\mathbb{R}^d} q(\mathbf{x}) e^{-2i\pi \mathbf{f}^T \mathbf{x}} d^d \mathbf{x} = |Q(\mathbf{f})| e^{-2i\pi \varphi_Q(\mathbf{f})}$$

this defining the magnitude, $|Q(\mathbf{f})|$, and phase, $\varphi_Q(\mathbf{f})$, of $q(\mathbf{x})$ in the Fourier domain. Given a 1-D signal $q(x)$, the HT of $q(x)$, is equivalently defined by

$$\mathcal{H}q(x) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{q(y)}{x - y} dy, \quad (\mathcal{H}Q)(f) = (-i) Q(f) \operatorname{sgn}(f). \tag{3}$$

We now present the following arguments, as to why the local structure of the HT, is similar to that of the original signal. First, we collect $q(x)$ and $\mathcal{H}q(x)$ in a complex-valued signal, denoted by the *analytic signal* [10], and this is given by $q^{(+)}(x) = q(x) + i\mathcal{H}q(x)$. For $f' > 0$, if $q(x) = \cos(2\pi f' x)$, then $q^{(+)}(x) = \exp(2\pi i f' x)$. If a signal corresponds to a modulated oscillation with slowly changing period and amplitude, a similar result holds under some suitable conditions [10]. Thus, in this simple case, $\mathcal{H}q(x)$ can be argued to have the same time-frequency structure as $q(x)$. For other signals commonly in signal processing, the (complex-valued) analytic signal is used to construct the time-frequency representation of the real signal, i.e., the Wigner–Ville distribution [11]. When using such representations, implicitly it is assumed that $q(x)$ and $\mathcal{H}q(x)$ have the same local structure. We also note that

$$q(x) = 2 \int\limits_{0}^{\infty} |Q(f)| \cos\left(2\pi \left(fx - \varphi_Q(f)\right)\right) df$$

$$\mathcal{H}q(x) = 2 \int\limits_{0}^{\infty} |Q(f)| \sin\left(2\pi \left(fx - \varphi_Q(f)\right)\right) df. \tag{4}$$

Thus, the frequency contribution of $q(x)$, previously associated with $\cos(2\pi(\varphi_Q(f) - fx))$, is shifted in cycle or phase, by $\pi/2$ in $\mathcal{H}q(x)$. Thus, we are recovering a similar signal, as the frequency description is the same, but a given frequency contribution is assigned to slightly different time locations. Based on this, if somewhat heuristic justification, we argue that $\mathcal{H}q(x)$ should have roughly the same time-frequency support as $q(x)$, and we refer to this property, as property 1) of the HT. The property implies that the mean of the DWT coefficient of $\mathcal{H}q(x)$, should be nonzero when the mean of the DWT coefficient of $q(x)$ is nonzero. Property 1) needs to be extended to 2-D.

Furthermore, the stochastic properties of the DWT of the analytic signal are useful, and allow us to define a suitable threshold procedure. These are related to a number of properties of the HT, that we shall enumerate and suitably extend. First, we note property 2), that the HT satisfies, namely $\mathcal{H}q(x)$ is orthogonal to $q(x)$, i.e., $\int \mathcal{H}q^*(x)q(x)dx = 0$. This ensures that the DWT of the noise and the DWT of the HT of the noise are approximately

uncorrelated at $[j, k]$.[1] Second, we note property 3), namely the magnitude of the HT of $q(x)$ at any given frequency $f \neq 0$, is the same as that of the original signal. This ensures that the DWT of the observed signal and its HT have approximately the same variance. Finally, we note property 4), the HT is linear in the signal. This establishes that the DWT coefficients of the signal and the HT of the signal are jointly normal, given (1). Properties 1)–4) follow from (3). The DWT of the signal is then estimated by the observed DWT if the magnitude of the DWT of the analytic signal exceeds a given threshold. For a complex-valued decomposition where the real and imaginary parts are correlated, common practice would involve transforming the variables into an uncorrelated pair of variables, and then using the magnitude of the complex coefficients defined by the uncorrelated pair, to threshold the wavelet decomposition [15]. While such a procedure will produce a variable where the distribution of the magnitude of the signal-free coefficient is tractable, when the signal is present, the first-order properties of the new magnitude will in general not be suitable. For "analytic" denoising, simultaneously the distribution of the noise only coefficients is tractable, and the magnitude is interpretable in terms of deterministic structure of the observed image. Once the coefficient has been estimated using "analytic" thresholding, the estimate of the signal will be based on the thresholded DWT of the observed signal alone. Thus, discontinuities in $q(x)$ can still be reconstructed, and are not necessarily blurred out by the HT.

The "analytic denoising" procedure will now be generalized to 2-D. We need to define a set of images, i.e., more than one, that will be indexed by $l = 1, \ldots, L$, to serve the same purpose that the HT of the signal did in 1-D. To construct these images we use a set of linear transformations, so that property 4) will be satisfied. The HT and the signal in 1-D formed a natural representation in terms of the "analytic" signal, where the real and imaginary components were phase shifted versions of each other, or "in quadrature." We shall, therefore, refer to images constructed from 2-D extensions of the HT as "quadrature components," where the method of construction is denoted by s.

Definition 2.1 (s-Quadrature Components): Any set of images $\{\breve{q}^{(s,l)}(\mathbf{x})\}_{l=1}^{L}$ that satisfy the following conditions.

1) The space and spatial frequency support of $[\breve{q}^{(s,1)}(\mathbf{x}), \ldots, \breve{q}^{(s,L)}(\mathbf{x})]$, is similar to that of $q(\mathbf{x}) = \breve{q}^{(s,0)}$.
2) Each $\breve{q}^{(s,l)}(\mathbf{x})$ is orthogonal ("out of phase") to the original signal, or $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q^*(\mathbf{x}) \breve{q}^{(s,l)}(\mathbf{x}) d^2\mathbf{x} = 0$, and for all separable $q_S(\mathbf{x}) = q_1(x_1) q_2(x_2)$, also $\int_{-\infty}^{\infty} q_S^*(\mathbf{x}) \breve{q}_S^{(s,l)}(\mathbf{x}) dx_m = 0$, for $m = 1, 2, l = 1, \ldots, L$.
3) The combined energy assigned to each frequency $\mathbf{f}$, from the full set of quadrature components at all points of $\mathbf{f} \in \mathbb{R}^2$ except for a finite set of frequencies, satisfies the equation $\sum_{l=1}^{L} |\breve{Q}^{(s,l)}(\mathbf{f})|^2 = C_L^{(s)} |Q(\mathbf{f})|^2$, where $0 < C_L^{(s)} < \infty$ is constant.
4) Each $\breve{q}^{(s,l)}(\mathbf{x})$, for $l = 1, \ldots, L$, is constructed by a linear transformation of $q(\mathbf{x})$, are denoted the *s-quadrature components* of $q(\mathbf{x})$.

[1]Note that using the dual-tree [14] at level $j = 1$ gives real and imaginary wavelet coefficients whose correlation is 0.5997, not even *approximately* zero. The filters from http://taco.poly.edu/WaveletSoftware/dt1D.html were used.

The four properties of the continuous s-quadrature components [properties 1)–4)], have been chosen to ensure the correct joint distribution of the DWT of the noisy image, and the DWTs of the noisy s-quadrature components. From property 1) we note that the magnitudes of the DWTs of the quadrature components are large, when the image has a contributing DWT. Property 2) ensures the DWTs of the image and its quadrature components are uncorrelated (independent given Gaussian noise), and the quadrature components are with the noise model of (1), also pairwise uncorrelated to each other. Property 3) ensures the variances of the DWTs are analytically determinable. Property 4) determines the joint distribution of the DWTs of the image and the quadrature components, at a given local point, with the model of (1). Of course, while the definition of quadrature components now seems justifiable, this does not guarantee the existence of such objects. We shall give two different specific examples of quadrature components, based on extending the analytic signal to 2-D.

The transformation that constructs quadrature component $\breve{q}^{(s,l)}(\mathbf{x})$, for each value of l, is implemented in the spatial domain by a convolution (denoted by $**$), using kernel $v^{(s,l)}(\mathbf{x})$. The integral will be approximated using a Riemann sum, where $v^{(s,l)}(\mathbf{x})$ is replaced by a periodic discrete linear filter $v_{D,\mathbf{x}}^{(s,l)}$. The FT of $v^{(s,l)}(\mathbf{x})$ is denoted by $V^{(s,l)}(\mathbf{f})$, while taking the discrete Fourier transform (DFT) of $v_{D,\mathbf{x}}^{(s,l)}$, yields the object $V_D^{(s,l)}(\mathbf{f})$. The discrete implementation of the calculation of the quadrature components is outlined in Appendix A. Define $\breve{q}_{\mathbf{x}}^{(s,l)} = (v_{D,:}^{(s,l)} ** q)_{\mathbf{x}}$, and $\breve{q}^{(s,l)}(\mathbf{x}) = (v^{(s,l)} ** q)(\mathbf{x})$. An error of order $O(1/N)$ is introduced by the Riemann approximation, i.e., $\breve{q}^{(s,l)}(\mathbf{x}\Delta x) = \breve{q}_{\mathbf{x}}^{(s,l)} + O(1/N)$. Note that the properties of the continuous s-quadrature components are only preserved up to $O(1/N)$, once $\{\breve{q}_{\mathbf{x}}^{(s,l)}\}$ are calculated, and used instead of $\breve{q}^{(s,l)}(\mathbf{x})$.

C. Riesz Transforms

The RTs have been used in combination with the continuous wavelet transform (CWT) by Metikas and Olhede [16], [17]. The RTs of $q(\cdot)$, denoted by $\breve{q}^{(r,1)}(\mathbf{x})$ and $\breve{q}^{(r,2)}(\mathbf{x})$ are calculated by convolving $q(\cdot)$ with the Riesz kernels $v^{(r,l)}(\mathbf{x})$, for $l = 1, 2$, given with $x = \sqrt{x_1^2 + x_2^2}$ and $f = \sqrt{f_1^2 + f_2^2}$ by

$$v^{(r,l)}(\mathbf{x}) = \frac{x_l}{2\pi x^3}, \quad V^{(r,l)}(\mathbf{f}) = -i\frac{f_l}{f}, \; l = 1, 2$$

$$\breve{q}^{(r,l)}(\mathbf{x}) = \left(v^{(r,l)} ** q\right)(\mathbf{x}), \; l = 1, 2. \tag{5}$$

The RTs satisfy properties 2)–4) of quadrature components (see, for example, [17, pp. 15–16]). We implement the RTs discretely (see Appendix A), and so the calculated discrete images are only approximately satisfying 2)–4). Given the RTs combine to have the same norm as $q(\mathbf{x})$, $C_2^{(r)} = 1$. To establish property 1), cf [17], consider the interpretation of the RTs for unidirectional structure. If the image admits the representation with $\mathbf{u} = [\cos(\nu) \sin(\nu)]^T$, $\nu \in [0, \pi)$, for some 1-D function $\widetilde{Q}_U(f)$, of

$$q_U(\mathbf{x}) = \int_0^{\infty} \widetilde{Q}_U(f) \cos(2\pi f \mathbf{u}^T \mathbf{x}) df \tag{6}$$

then the interpretation of the RTs is clear. We use polar co-ordinates and let $\mathbf{f} = f[\cos(\phi)\sin(\phi)]$. The FT of $q_U(\mathbf{x})$ is $Q_U(\mathbf{f}) = \widetilde{Q}_U(f)/2[\delta(\phi - \nu) + \delta(\phi + \nu)]$, and we note

$$\left[\breve{q}_U^{(r,1)}(\mathbf{x})\breve{q}_U^{(r,2)}(\mathbf{x})\right] = [\cos(\nu)\sin(\nu)]\int_0^\infty \widetilde{Q}_U(f)\sin(2\pi f\mathbf{u}^T\mathbf{x})df. \tag{7}$$

Thus, the two quadrature components represent the *same* 1-D directional variation as $q_U(\mathbf{x})$, with the *same* directionality as $q_U(\mathbf{x})$, but where the variations in direction ν have been shifted in phase by $\pi/2$. Thus, (informally) for unidirectional variation, the RTs have the same spatial and spatial frequency support as the original signal. In general, we cannot assume that the signal is globally unidirectional, but as convolutions commute, we can think of either implementing the DWT, or the RTs first (implementing the subsampling of the DWT last). Once we have localized the image, in scale and space, it is more reasonable to assume that the local structure of the image is unidirectional. We, thus, argue from (7) that the DWT of the RTs have the same local structure as the DWT of $q(\mathbf{x})$, for locally unidirectional structure. Note that we are *not* assuming that $q(\mathbf{x})$ is periodic or oscillatory at this stage. For nondirectional structure, the RTs still have identical frequency support to the original image. The RTs have similar spatial support to the original image [see (5)] as the spatial decay of the filter is $O(1/x^2)$. Using the RTs when there is no local unidirectional variation, just like for "analytic" denoising, can be viewed as taking local weighted combinations of the local coefficients. If the signal has locally similar structure, then again the magnitudes of the RTs should be large, if there is signal present. Using the RTs to form the hyperanalytic signal is denoted by putting $s = r$.

D. Hypercomplex Transforms

A second set of 2-D HTs, are the HCTs. Denote the partial HT [10] in direction x_l by $\mathcal{H}_l$. Three additional quadrature components are defined by the three HCTs

$$\breve{q}^{(h,1)}(\mathbf{x}) = \mathcal{H}_1\{q\}(\mathbf{x}), \quad \breve{q}^{(h,2)}(\mathbf{x}) = \mathcal{H}_2\{q\}(\mathbf{x})$$
$$\breve{q}^{(h,3)}(\mathbf{x}) = \mathcal{H}_2\mathcal{H}_1\{q\}(\mathbf{x}). \tag{8}$$

In Olhede and Metikas [17, pp. 12–13], it was shown that the HCTs satisfy properties 2)–4), and note $C_3^{(h)} = 3$. We implement the HCTs using the discrete HT (DHT) (see Appendix A), and so the discretely filtered additional images are only approximately satisfying conditions 2)–4). If the image is naturally expressed as separable, the three HCTs of $q_S(\cdot)$ can be argued to have the same structure as the original image, using the same arguments as in Section II-B. By using the hyperanalytic signal, we hope to alleviate observed problems (see, for example, Starck *et al.* [3, p. 671]), when estimating nonseparable images, using a wavelet decomposition calculated using a separable wavelet basis. We argue that the local structure of the HCTs of the image, is the same as the local structure of the image, by the following. Define the partial FT (PFT) of $q(\mathbf{x})$ in

direction x_1, by $Q_1(f_1, x_2) = |Q_1(f_1, x_2)|e^{-2\pi\varphi_1(f_1, x_2)}$, with $Q_1(f_1, x_2) = \int_{-\infty}^\infty q(\mathbf{x})e^{-2i\pi f_1 x_1}dx_1$. The image is written as

$$q(\mathbf{x}) = 2\int_0^\infty |Q_1(f_1, x_2)|\cos\left(2\pi\left(f_1 x_1 - \varphi_1(f_1, x_2)\right)\right)df_1 \tag{9}$$

$$\breve{q}^{(h,1)} \equiv \mathcal{H}_1 q(\mathbf{x}) = 2\int_0^\infty |Q_1(f_1, x_2)|$$
$$\times \cos\left(2\pi\left(f_1 x_1 - \varphi_1(f_1, x_2) - \frac{1}{4}\right)\right)df_1. \tag{10}$$

Thus, $\breve{q}^{(h,1)}(\mathbf{x})$ corresponds to replicating all variation in x_1, for any fixed value of x_2, shifted in phase by $\pi/2$ and, *mutatis mutandis*, the analogous statements hold for $\breve{q}^{(h,2)}(\mathbf{x})$ and $\breve{q}^{(h,3)}(\mathbf{x})$. If $q(\mathbf{x})$ corresponds to a particular time-frequency structure as a signal in x_1, for fixed values of x_2, then $\breve{q}^{(h,1)}(\mathbf{x})$ will replicate the same structure, but shifted in phase in x_1. Thus, the same structure is replicated in $\breve{q}^{(h,1)}(\mathbf{x})$ compared to $q(\mathbf{x})$, in either of the two axes separately, and for this reason we consider property 1) met. Thus, if the separable DWT coefficient of $q(\mathbf{x})$ is nonzero at a local index, then so should also the DWTs of $\{\breve{q}^{(h,l)}\}$ be, for $l = 1, \ldots, 3$. We denote the hypercomplex components by taking $s = h$. Finally a naive 2-D extension of "analytic" denoising [9], would be to define a single extra quadrature component corresponding to $\breve{q}^{(h,3)}(\mathbf{x})$, phase shifting in both spatial directions simultaneously. We discuss the risk of this procedure in Section III-F, and show that this exceeds those of the other proposed hyperanalytic denoising procedures. The procedure is denoted by $s = a$.

For a fixed choice of s, we form the DWT of all $L + 1$ images, $\{[\breve{q}_{\mathbf{x}}^{(s,l)}]_{\mathbf{x}\in D}\}_{l=0}^L$. We define $\breve{W}_{\boldsymbol{\xi}}^{(q,s,l)} = W_{\boldsymbol{\xi}}^{(\breve{q}^{(s,l)})}$, $l = 0, \ldots, L$. Also, let $\breve{W}_{\boldsymbol{\xi}}^{(\epsilon,s,l)}$ be the DWT of $\{[\breve{\epsilon}_{\mathbf{x}}^{(s,l)}]_{\mathbf{x}\in D}\}_{l=0}^L$, and collect these random variables in the vector $\mathbf{n}^{(s,u)} = [n_1^{(s,u)}, \ldots, n_{L+1}^{(s,u)}]^T = [\breve{W}_{\boldsymbol{\xi}}^{(\epsilon,s,0)}, \ldots, \breve{W}_{\boldsymbol{\xi}}^{(\epsilon,s,L)}]^T$. The explicit dependence on $\mathbf{k}$ and j in $\mathbf{n}^{(s,u)}$ is dropped, as the distribution of $\mathbf{n}^{(s,u)}$, will *only* depend on s, the choice of quadrature components, and u, the tensor product index. Given the outlined properties of the quadrature components, it seems reasonable to consider a signal as "locally present" if the magnitudes of the DWTs of all the quadrature components, i.e., $\{|\breve{W}_{\boldsymbol{\xi}}^{(q,s,l)}|^2\}$, are large. Before we do so, to define an estimation procedure, and justify this measure of local presence, we shall need to demonstrate that the combined magnitudes of the DWTs of the quadrature components of noise behave "well."

E. Stochastic Properties of the DWTs of the Quadrature Components

Proposition 1 (Energy of Quadrature Components): At a fixed index value $\boldsymbol{\xi}$, the total energy of the DWT of the discrete quadrature components of white noise, with variance σ^2, is given by

$$E\left(\sum_{l=2}^{L+1} n_l^{(s,u)2}\right) = C_L^{(s)}\sigma^2 + O(1/N), \quad u = 1, 2, 3, 4. \tag{11}$$

Proof: See Appendix C. The error term follows from the discrete implementation of the hyperanalytic signal. The result follows from properties 2)–4) of Definition 2.1. ∎

Proposition 2 (Covariance of Transforms): At a fixed index value $\boldsymbol{\xi}$, the covariance of the DWT of white noise, and the DWT of any of the discrete quadrature components of white noise is of order $O(1/N)$.

Proof: See Appendix C. The error term follows from the Riemann approximation to the integral, and the result follows from property 2). ∎

Thus, at any given value of $\boldsymbol{\xi}$, the DWTs of $\epsilon_{\mathbf{x}}$ and $\{\breve{\epsilon}_{\mathbf{x}}^{(s,l)}\}$, at any fixed $l = 1, \ldots, L$, are approximately uncorrelated. The combined energy of the DWTs of $\{\breve{\epsilon}_{\mathbf{x}}^{(s,l)}\}$, for $l = 1, \ldots, L$, is a multiplicative constant of the energy of the DWT of $\epsilon_{\mathbf{x}}$. We have, thus, demonstrated that the energy of the DWT of the noise, and its s-quadrature components, is a multiple of the variance. So, in brief, we have made the case that the magnitude of the DWTs of quadrature components will be large if there is signal present. The energy of the DWTs of the noise components is controlled and proportional to the variance. We are now in the position where we may define an estimator of a wavelet coefficient.

III. ESTIMATION

A. Defining Estimates

We define a local s-magnitude, in terms of the combined magnitudes of the DWTs, of the full set of $\{\breve{q}^{(s,l)}(\mathbf{x})\}_{l=0}^{L}$. The local s-magnitude quantifies the local presence of the signal.

Definition 3.1 (The s-Magnitude of a DWT Coefficient): We define the s-magnitude of the DWT of signal $q(\mathbf{x})$, $W_{\boldsymbol{\xi}}^{(q)}$, using quadrature components denoted by s, via:
$M_{\boldsymbol{\xi}}^{(q,s)2} = (1/(C_L^{(s)} + 1)) \sum_{l=0}^{L} \breve{W}_{\boldsymbol{\xi}}^{(q,s,l)2}$.

To relate the s-magnitude definition to previous local threshold strategies [4], let for some fixed $L_1 \in \mathbb{N}$ $B_i^j = \{k : k = i + l, l = -L_1, \ldots L_1\}$. Then the s-magnitude, $M_{\boldsymbol{\xi}}^{(q,s)2}$, is a 2-D analogy to $S_{j,k}^2 = \sum_{l \in B_k^j} W_{[j,l]}^{(q)2}$, used by Cai and Silverman [4, p. 132], to block threshold DWTs. Each coefficient $W_{\boldsymbol{\xi}}^{(q)}$, will be estimated by hard thresholding the observed coefficient, depending on the value of $M_{\boldsymbol{\xi}}^{(Y,s)2}$, with threshold λ^2, according to

$$\widehat{W}_{\boldsymbol{\xi}}^{(q,s)}(\lambda^2) = \begin{cases} W_{\boldsymbol{\xi}}^{(Y)}, & \text{if } M_{\boldsymbol{\xi}}^{(Y,s)2} \geq \frac{\sigma^2}{C_L^{(s)} + 1} \lambda^2 \\ 0, & \text{if } M_{\boldsymbol{\xi}}^{(Y,s)2} < \frac{\sigma^2}{C_L^{(s)} + 1} \lambda^2 \end{cases}. \quad (12)$$

This choice of estimator is different from commonly used thresholding rules, as the rule is specified in terms of $M_{\boldsymbol{\xi}}^{(Y,s)2}$ *not* $|W_{\boldsymbol{\xi}}^{(Y)}|$. When $W_{\boldsymbol{\xi}}^{(q)}$ is non-negligible, but not large compared to the noise-level, the estimated coefficient may be zero, even if $W_{\boldsymbol{\xi}}^{(q)}$ represents parts of a curve or directional structure. Removing $W_{\boldsymbol{\xi}}^{(Y)}$ from the estimated image will produce artifacts in the reconstruction, and inflate the MSE of $\widehat{q}_{\mathbf{x}}$. We have argued that the combined magnitudes of the DWTs of the quadrature components are consistently large, when $W_{\boldsymbol{\xi}}^{(q)}$ is part of the deterministic image. We have shown the noise to be uncorrelated (up to $O(1/N)$), across quadrature components, and so the magnitude is not consistently large, if there is no

signal present. Thus, the estimator $\widehat{W}_{\boldsymbol{\xi}}^{(q,s)}(\lambda^2)$, is likelier to keep, rather than annihilate, energy part of a structure, while we still control the risk when estimating noise dominated coefficients. These heuristic arguments will be made more precise in Section III-B and F.

We define, for any fixed choice of s-quadrature components, using tensor product wavelet u, $\boldsymbol{W}_{\boldsymbol{\xi}}^{(s)} = [\breve{W}_{\boldsymbol{\xi}}^{(Y,s,0)}, \ldots, \breve{W}_{\boldsymbol{\xi}}^{(Y,s,L)}]^T$, as the DWT of the discrete implementation of the quadrature components of $\boldsymbol{Y}$ at $\boldsymbol{\xi}$. We also let $\boldsymbol{\mu}_{\boldsymbol{\xi}}^{(s)} = [\mu_{\boldsymbol{\xi},1}^{(s)}, \ldots, \mu_{\boldsymbol{\xi},L+1}^{(s)}]^T = [\breve{W}_{\boldsymbol{\xi}}^{(q,s,0)}, \ldots, \breve{W}_{\boldsymbol{\xi}}^{(q,s,L)}]^T$, be the DWT of the quadrature components of $q(\mathbf{x})$, i.e., $\breve{q}^{(s,l)}(\mathbf{x})\}$, at $\boldsymbol{\xi}$. $\mu_{\boldsymbol{\xi},1}^{(s)} = W_{\boldsymbol{\xi}}^{(q)}$ does *not* depend on the choice of s. The estimator of $W_{\boldsymbol{\xi}}^{(q)}$ using the s indexed components and the threshold λ^2 is denoted by $\widehat{\mu}_{\boldsymbol{\xi},1}^{(s)}(\lambda^2) = \widehat{W}_{\boldsymbol{\xi}}^{(q,s)}(\lambda^2)$.

B. Magnitude of Typical Image Features

Deterministic images are often modeled as the combination of edges and texture (see, for example, work by Vese and Osher [18]). We calculate the wavelet coefficients of some stylized image features, and need to specify the local characteristics of the features, in the spatial region that the wavelet filter is supported over. Given the wavelet filters are compact the properties of the signal outside this spatial region are not important. For a fixed value of j, let $\mathbf{x} = 2^j(\mathbf{k} + \mathbf{1}) - \mathbf{1}$ denote a spatial location. For fixed values of $\mathbf{k}$ and j, and, thus, $\mathbf{x}$, each texture component is modeled locally in spatial variable $\mathbf{x}'$, as an AM/FM signal [16, eq. 27]

$$t(\mathbf{x}') = a_t(\mathbf{x}) \cos\left(2\pi \mathbf{f}_0^T(\mathbf{x})\mathbf{x}'\right),$$
$$\mathbf{x}' \in \left[\mathbf{x} - (2^{j-1}\tilde{L} - 1)\mathbf{1}, \mathbf{x}\right] \quad (13)$$

where $\tilde{L}$ denotes the length of the wavelet filter, and each edge component is modeled locally by

$$e(\mathbf{x}') = a_e(x_1)\delta\left(\cos(\theta)x_1' + \sin(\theta)x_2' - c\right) \quad (14)$$

for $\mathbf{x}' \in [\mathbf{x} - (2^{j-1}\tilde{L} - 1)\mathbf{1}, \mathbf{x}]$. $a_e(x)$ and $a_t(\mathbf{x})$ are assumed to be slowly varying. In general we do not expect to observe sinusoids or discontinuities that for very slowly varying $a_e(\cdot)$ span the entire observed image. To be able to carry out theoretical calculations, stylized image structures are analysed, that observed images would subsequently approximate. Given the DWT truncates images in space and scale, and the hyperanalytization commutes with the DWT (implementing subsampling last), these results are still useful for real images. Our subsequent risk calculations will demonstrate that the success of the method depends on the expected value of the quadrature components exhibiting suitable properties.

We define the maximum overlap discrete wavelet transform (MODWT) coefficients of the image s-quadrature components, $\{\widetilde{W}_{\boldsymbol{\xi}}^{(q,s,l)}\}$. These are the DWT coefficients calculated without subsampling, with a new normalization at each level j, see Percival and Walden [13, ch. 4]. The FT of the MODWT filter $\tilde{h}_{j,u,\mathbf{k}}$ is denoted by $\widetilde{H}_{j,u}(\mathbf{f}) = |\widetilde{H}_{j,u}(\mathbf{f})|e^{-2i\pi\widetilde{\varphi}_{j,u}(\mathbf{f})}$, this defining the modulus, or $|\widetilde{H}_{j,u}(\mathbf{f})|$, and phase, $\widetilde{\varphi}_{j,u}(\mathbf{f})$, of $\widetilde{H}_{j,u}(\mathbf{f})$, the MODWT wavelet filter. Let the region of frequency space where $\widetilde{H}_{j,u}(\mathbf{f})$ is mainly supported, be denoted by $\Omega_{j,u}$, and let $\boldsymbol{\xi}_{\mathbf{x}} = $

$[j, u, \mathbf{x}]^T$. The DWT coefficients of a generic signal $q(\mathbf{x}')$ can be extracted from the MODWT coefficients of $q(\mathbf{x}')$, using the relations $W_{j,u,\mathbf{k}}^{(q)} = 2^j \widetilde{W}_{j,u,\mathbf{x}}^{(q)}$ [13, p. 203].

Lemma 1 (RT Magnitude of Local Oscillation): If the signal locally takes the form of (13) then the r-magnitude of the DWT, defined in Definition 3.1, is given by

$$M_{\boldsymbol{\xi}}^{(t,r)2} = 2^{2j-1} a_t^2(\mathbf{x}) I\left(\mathbf{f}_0(\mathbf{x}) \in \Omega_{j,u}\right) + \rho_1 + O\left(\frac{1}{N}\right) \quad (15)$$

where $I(\cdot)$ is the indicator function. ρ_1 is an error term introduced because of the leakage of the wavelet filters in the frequency domain. If a sufficiently long wavelet filter is used, this term can be ignored; see Nielsen [19], for more discussion on avoiding leakage. The $O(1/N)$ term is due to the approximation of the DFT of the $\cos(\cdot)$ function, see Appendix B.

Proof: See Appendix B. ∎

Lemma 2 (HCT Magnitude of Local Oscillation): If the signal locally takes the form of (13), then the h-magnitude given in Definition 3.1, is

$$M_{\boldsymbol{\xi}}^{(t,h)2} = 2^{2j-1} a_t^2(\mathbf{x}) I\left(\mathbf{f}_0(\mathbf{x}) \in \Omega_{j,u}\right) + \rho_2 + O\left(\frac{1}{N}\right) \quad (16)$$

where ρ_2, is an error term depending on the leakage of the wavelet filters in the frequency domain, and the $O(1/N)$ follows due to the approximation of the DFT.

Proof: See Appendix B. ∎

Lemma 3 (RT Magnitude of Discontinuity): If the signal locally takes the form of (14), then the r-magnitude given in Definition 3.1, is given by

$$
\begin{aligned}
W_{\boldsymbol{\xi}}^{(e)} = {} & 2^j A_e(0) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \widetilde{H}_{j,u}\left(\cos(\theta)\nu_1, \sin(\theta)\nu_1\right) \\
& \times e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 \\
& + \rho_3 + O\left(\frac{1}{N}\right) \\
U_{\boldsymbol{\xi}}^{(e)} = {} & 2^j A_e(0)(-i) \\
& \times \int_{\nu_{1,\min}}^{\nu_{1,\max}} \operatorname{sgn}(\nu_1) \widetilde{H}_{j,u}\left(\cos(\theta)\nu_1, \sin(\theta)\nu_1\right) \\
& \times e^{2\pi i \nu_1 (\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 \\
& + \rho_4 + O\left(\frac{1}{N}\right) \\
M_{\boldsymbol{\xi}}^{(e,r)2} = {} & \frac{1}{2} W_{\boldsymbol{\xi}}^{(e)2} + \frac{1}{2} U_{\boldsymbol{\xi}}^{(e)2} + \rho_5 + O\left(\frac{1}{N}\right) \quad (17)
\end{aligned}
$$

where ρ_3, ρ_4, and ρ_5 depend on the smoothness of $a_e(\cdot)$, while $\nu_{1,\min}$ and $\nu_{1,\max}$ are given in Appendix B. The $O(1/N)$ term follows from approximating the DFT of $e(\mathbf{x})$.

Proof: See Appendix B. ∎

Lemma 4 (HCT Magnitude of Discontinuity): If the signal locally can be approximated by (14), then the magnitude defined in definition 3.1, is given by

$$M_{\boldsymbol{\xi}}^{(e,r)2} = \frac{1}{2} W_{\boldsymbol{\xi}}^{(e)2} + \frac{1}{2} U_{\boldsymbol{\xi}}^{(e)2} + \rho_6 + O\left(\frac{1}{N}\right) \quad (18)$$

where ρ_6 depends on the smoothness of $a_e(\cdot)$, and the forms of $W_{\boldsymbol{\xi}}^{(e)}$ and $U_{\boldsymbol{\xi}}^{(e)}$ are given in lemma 3. The $O(1/N)$ term follows as in Lemma 3.

Proof: See Appendix B. ∎

For oscillatory signals, the magnitude, hence, reflects signal presence at $\boldsymbol{\xi}$, and the calculations motivate using the local s-magnitudes, when thresholding local oscillations. Equation (17) illustrates the problem experienced by an edge in a 2-D separable representation: only if $\theta = 0$ or $\theta = m\pi/2$, for $m \in \mathbb{Z}$, will the edge live in $\Omega_{j,2}$ or $\Omega_{j,3}$, i.e., constant in one direction and variable in the other. In this case the representation will be extremely compressed (note that the proof needs to be adjusted for $\theta = 0$).

From (17), it is clear that the compact spatial support of $\widetilde{h}_{\boldsymbol{\xi}_\mathbf{x}}$ ensures that the energy of $U_{\boldsymbol{\xi}}^{(e)}$ and $W_{\boldsymbol{\xi}}^{(e)}$ is mainly concentrated near $\cos(\theta)x_1 + \sin(\theta)x_2 = c$. If we represent $\widetilde{H}_{j,u}(\cdot)$ in terms of a magnitude and a phase, the difference between $\widetilde{\varphi}_{j,u}(\cos(\theta)\nu, \sin(\theta)\nu)$ and $2\pi\nu(\cos(\theta)x_1 + \sin(\theta)x_2 - c)$ will determine exactly at which spatial indices $W_{\boldsymbol{\xi}}^{(e)}$ and $U_{\boldsymbol{\xi}}^{(e)}$ have non-negligible magnitudes (see also the discussion in Gopinath on this phenomenon [20, p. 1794]). The DWT of the quadrature components will be large near the discontinuity, and the quadrature components can be used to improve the estimation of local discontinuities. Using hyperanalytic denoising, line discontinuities are better reconstructed [see Fig. 2(a)–(c)]. A curved discontinuity can be approximated as the aggregation of amplitude modulated line discontinuities, and so improvements in estimation can be observed also for curved structures [see Fig. 2(e)–(g)].

C. Distribution of Noise and "Universal" Thresholds

Let $K = N^2$ be the total number of coefficients of the original observed image. For a fixed s, the distribution of $\mathbf{n}^{(s,u)}$ must be determined, to obtain a "universal" threshold [21], λ_K^2. We let $M_K^{(s,\max)2} = \max_{\boldsymbol{\xi}}(C_L^{(s)} + 1) M_{\boldsymbol{\xi}}^{(s)2}/\sigma^2$. Downie and Silverman [21] proposed that a "universal" threshold should satisfy taking a value, such that $\lim_{K\to\infty} P(M_K^{(s,\max)2} \leq \lambda_K^2) = C$, for some constant $0 < C < 1$. As K increases, the expected number of coefficients exceeding the threshold is some small but finite nonzero value. We choose to use a slightly more conservative threshold for both the Riesz and Hypercomplex magnitudes. We cannot quite achieve analogous results to Downie and Silverman, as the set of DWT coefficients $\mathbf{W}_{\boldsymbol{\xi}}^{(s)}$, are correlated across indices $\boldsymbol{\xi}$. We adopt arguments similar to those given by Johnstone and Silverman [22], and Olhede and Walden [9], to justify the choice of threshold.

We do not aim to determine the full covariance structure of the full set of wavelet coefficients, of the observed image and the quadrature components. Define $\mathcal{M}_K^{(s)2}$ as the maximum of K *independent* variates with the same marginal distribution as

$\{(C_L^{(s)}+1)M_{\boldsymbol{\xi}}^{(s)2}/\sigma^2\}$. By determining a threshold for $\mathcal{M}_K^{(s)2}$, we obtain a conservative threshold for $M_K^{(s,\max)2}$, as

$$P\left(M_K^{(s,\max)2} \le \lambda_K^{(s)2}\right)$$
$$= P\left(\cap_{i=1}^K \left[\frac{M_{\boldsymbol{\xi}_i}^{(s)2}}{\sigma^2/\left(C_L^{(s)}+1\right)} \le \lambda_K^{(s)2}\right]\right)$$
$$\ge \prod_{i=1}^K P\left(\frac{M_{\boldsymbol{\xi}_i}^{(s)2}}{\sigma^2/\left(C_L^{(s)}+1\right)} \le \lambda_K^{(s)2}\right)$$
$$= P\left(\mathcal{M}_K^{(s)2} \le \lambda_K^{(s)2}\right)$$

by corollary 2, from Dykstra [23]. $C_L^{(s)}+1$ can be interpreted as the degrees of freedom associated with $M_{\boldsymbol{\xi}}^{(s)2}$. As shown in the subsequent section, the DWT of the quadrature components for any fixed value of $\boldsymbol{\xi}$ are approximately uncorrelated, and the $\boldsymbol{A}_i$ matrices of Dykstra are defined to take the nonuniform variances of $\{n_l^{(r,u)}\}$, into account.

D. Distribution of the Magnitude

We seek to determine the approximate distribution of $\mathbf{n}^{(s,u)}$ to derive the distribution of $M_{\boldsymbol{\xi}}^{(s)2}$. We, as usual, denote by $\stackrel{\mathcal{L}}{=}$ as equality in law [24].

Lemma 5 (Distribution of Riesz Coefficients): The DWT coefficients of the original signal and the discrete RTs of Gaussian white noise are distributed as

$$\mathbf{n}^{(r,u)} \stackrel{\mathcal{L}}{=} \boldsymbol{Z}^{(r,u)} + O(1/N), \quad u=1,2,3,4,$$
$$\boldsymbol{Z}^{(r,u)} \sim \mathcal{N}_3\left(\mathbf{0}, \sigma^2 \mathrm{diag}\left[1 \quad a^{(r,u)} \quad 1-a^{(r,u)}\right]\right) \quad (19)$$

where diag denotes a diagonal square matrix, $a^{(r,u)} = 1/2$, $u=1,4$, $a^{(r,2)} = (1/2)+2\tan^{-1}(1/2)+(1/2)\tan^{-1}(2)$ and $a^{(r,3)} = (1/2)-2\tan^{-1}(1/2)-(1/2)\tan^{-1}(2)$.

Proof: For the proof, see Appendix C. The $O(1/N)$ terms follow by the discrete approximation to the RT. ∎

Lemma 6 (Distribution of Riesz Magnitude): The magnitude square of the DWT of the discrete RTs of Gaussian white noise, denoted by $M_{\boldsymbol{\xi}}^{(\epsilon,r)2}$, are distributed as

$$\frac{M_{\boldsymbol{\xi}}^{(\epsilon,r)2}}{\sigma^2/2} \stackrel{\mathcal{L}}{=} \frac{\boldsymbol{Z}^{(r,u)T}\boldsymbol{Z}^{(r,u)}}{\sigma^2} + O(1/N) \stackrel{\mathcal{L}}{=} T_1 + O(1/N)$$
$$T_1 \sim \chi_1^2 + a^{(r,u)}\chi_1^2 + \left(1-a^{(r,u)}\right)\chi_1^2 \quad (20)$$

$j=1,\ldots,J$, $k_1,k_2=0,\ldots,N_j$, $u=1,\ldots,4$, where if $u=1,4$, T_1 has the distribution

$$f_{T_1}(t) = e^{-t}\frac{2}{\sqrt{\pi}}\int_0^{\sqrt{\frac{t}{2}}} e^{w^2}dw \quad (21)$$

while if $u=2,3$, the moment generating function of T_1 is given in Appendix C. By using formulae derived by [25] the probability of obtaining large values may be found.

Proof: For the proof see Appendix C. ∎

Lemma 7 (Distribution of HCT Coefficients): The DWT coefficients of the discrete HCT of Gaussian white noise are distributed as

$$\mathbf{n}^{(h,u)} \stackrel{\mathcal{L}}{=} \boldsymbol{Z}^{(h,u)} + O(1/N)$$
$$\boldsymbol{Z}^{(h,u)} \sim \mathcal{N}_4\left(\mathbf{0}, \sigma^2\mathrm{diag}[1 \ 1 \ 1 \ 1]\right), \quad u=1,2,3,4. \quad (22)$$

Proof: For the proof see Appendix C. ∎

Given the approximate joint distribution of the DWT coefficients at $\boldsymbol{\xi}$ has been determined, it trivially follows that the magnitude is distributed as

$$\frac{M_{\boldsymbol{\xi}}^{(\epsilon,h)2}}{\sigma^2/4} \stackrel{\mathcal{L}}{=} \frac{\boldsymbol{Z}^{(h,u)T}\boldsymbol{Z}^{(h,u)}}{\sigma^2} + O(1/N) \stackrel{\mathcal{L}}{=} T_2 + O(1/N)$$
$$T_2 \sim \chi_1^2 + \chi_1^2 + \chi_1^2 + \chi_1^2 \sim \chi_4^2.$$

E. Threshold Choice

Lemma 8 (Riesz Conservative Threshold): With a threshold of $\lambda_K^{(r)2}$ as given below, if $C > -1$, it follows:

$$\lambda_K^{(r)2} = 2\log(K) + 2C\log\left(\log(K)\right)$$
$$P\left(\mathcal{M}_K^{(r)} < \lambda_K^{(r)2}\right) \to 1. \quad (23)$$

Proof: For the proof see Appendix C. ∎

From [21] we may note that the RT threshold is, thus, like that of a χ_1^2 ($C > -1$). To ensure that the probability of estimating a purely noise generated image by zeros tends to 1, we take $C = 0$, rather than $C = -1$. Given the normalized marginal magnitudes of the HCT components are χ_4^2, we may use results of [21] to note that, $\lambda_K^{(h)2} = 2\log(K) + 2\log(\log(K))$, gives an appropriate threshold. As the wavelet coefficients will be correlated *across* $\boldsymbol{\xi}$, this is conservative. As a final step of the procedure we implement cycle-spinning [13, p. 429], which is known to improve MSE results considerably. Implementing hard thresholding based on the magnitude of $W_{\boldsymbol{\xi}}^{(Y)}$, will be denoted by $s = c$ (classical thresholding). We also discuss using a single quadrature component of $n_4^{(h,u)}$, when thresholding $n_1^{(h,u)}$. This corresponds to a naive extension of "analytic" thresholding to 2-D, denoted by taking $s = a$.

F. Risk Calculations

To compare the theoretical performance of the threshold estimators proposed in this paper, we calculate the standardized risk at any fixed value of $\boldsymbol{\xi}$. We define the standardized risk for procedure s, in terms of the standardized means of the $l = 0,\ldots,L$ quadrature components $\theta_{\boldsymbol{\xi},l+1}^{(s)} = \mu_{\boldsymbol{\xi},l+1}^{(s)}/\sigma$, by

$$R_{\boldsymbol{\theta}^{(s)}}^{(s)}(\lambda) = \sigma^{-2}E\left[\left(\widehat{\mu}_{\boldsymbol{\xi},1}^{(s)}(\lambda^2) - \mu_{\boldsymbol{\xi},1}^{(s)}\right)^2\right]. \quad (24)$$

If $s = r$, then we denote by r_1 and r_2 the two different cases that correspond to $u = 1, 4$ or $u = 2, 3$. For completeness we here also provide the risk of the "analytic" denoising, not given in Olhede and Walden [9].

Theorem 1 (The Risk of a Thresholded Coefficient): The standardized risk of an individual coefficient using threshold strategy $s = c, a, r, h$, with $\theta_{\boldsymbol{\xi},i}^{(s)} = \mu_{\boldsymbol{\xi},i}^{(s)}/\sigma = \theta_i^{(s)}$, noting $\theta_1^{(s)} =$

θ_1, is independent of s, and $\Upsilon_s(\lambda) = \sum_{l=1}^{L+1}(w_l + \theta_l^{(s)})^2 < \lambda^2$ is given by

$$R_{\boldsymbol{\theta}^{(c)}}^{(c)}(\lambda) = 1 + \int_{\Upsilon_c(\lambda)} \left[\theta_1^2 - w^2\right] \phi(w)dw$$

$$R_{\boldsymbol{\theta}^{(a)}}^{(a)}(\lambda) = 1 + \int_{\Upsilon_a(\lambda)} \left[\theta_1^2 - w^2\right] \phi(w_1)\phi(w_2)d^2\boldsymbol{w} + o(1)$$

$$R_{\boldsymbol{\theta}^{(r)}}^{(r)}(\lambda) = 1 + \frac{1}{\sqrt{a^{(r,u)}\left(1 - a^{(r,u)}\right)}}$$
$$\times \int_{\Upsilon_r(\lambda)} \left[\theta_1^2 - w^2\right] \phi(w_1)\phi\left(w_2/\sqrt{a^{(r,u)}}\right)$$
$$\times \phi\left(w_3/\sqrt{1 - a^{(r,u)}}\right) d^3\boldsymbol{w} + o(1)$$

$$R_{\boldsymbol{\theta}^{(h)}}^{(h)}(\lambda) = 1 + \int_{\Upsilon_h(\lambda)} \left[\theta_1^2 - w^2\right] \phi(w_1)\phi(w_2)$$
$$\times \phi(w_3)\phi(w_4)d^4\boldsymbol{w} + o(1). \qquad (25)$$

Proof: The risk of an individual coefficient using "universal" hard thresholding has been noted by Marron *et al.* [26], while the risk of the hyperanalytic thresholds are derived in Appendix C. $s = a$ is a special case of the bi-variate thresholding investigated by Fryzlewicz [8]. The $o(1)$ terms follow as the risk is derived using the asymptotic distribution of $\mathbf{W}_{\boldsymbol{\xi}}^{(s)}$, and the $O(1/N)$ terms in the covariance of the components yield an extra term in the expression. This error term can be bounded precisely using a series expansion of the exponential, and the limits of the integration. ∎

For some examples of signal/noise distributions, the individual risk of a given coefficient is plotted in Fig. 1, for the four estimation procedures using the "universal" threshold. Fig. 1(a) shows the reduced risk of "analytic" thresholding (denoted by a), compared to hard "universal" thresholding in 1-D (denoted by c), when the means of the wavelet coefficient of the signal and of the HT of the signal are equal. The reduction in the risk provides theoretical justification for "analytic" denoising [9]. Fig. 1(b)–(d) shows the risk at $\boldsymbol{\xi}$, using the proposed threshold. The risk is calculated with $K = 256^2$, and using the "universal" threshold, for any of the proposed procedures. If the means of the DWTs of the s-quadrature components are of similar magnitude to the DWT of the signal, then the risk is reduced [see Fig. 1(b) or (d)]. The greatest weakness of the proposed methods is if the means of the s-quadrature components are completely disparate from that of the original signal. This can be noted in Fig. 1(c). The results of Section III-B, argue that this will not be the case for some typical image features. Also, the norm of a signal and the full set of quadrature components are multiples of each other. Thus, given the results of Section III-B, the means of the DWT of the signal and the DWT of the quadrature components, are unlikely to be consistently mismatched. Finally, if there is no signal present, we may bound the risk in estimating a coefficient, as K becomes large.

Corollary 1 (The Risk of a Noisy Thresholded Coefficient): The risk of an individual coefficient, when there is no signal

present, using threshold strategy $s = c, a, r, h$ with $\theta_{l+1} = 0$, for $l = 0, \ldots, L$ is given by

$$R_0^{(c)}(\lambda) = 1 - \gamma\left(\frac{1}{2}, \frac{\lambda^2}{2}\right), \quad R_0^{(a)}(\lambda) = e^{-\frac{\lambda^2}{2}}\frac{2 + \lambda^2}{2} + o(1)$$

$$R_0^{(h)}(\lambda) = e^{-\frac{\lambda^2}{2}}\left(1 + \frac{\lambda^2}{2} + \frac{\lambda^4}{8}\right) + o(1)$$

$$R_0^{(r)}(\lambda) = 1 + \gamma\left(\frac{1}{2}, \frac{\lambda^2}{2}\right) - 2\gamma\left(\frac{1}{2}, \lambda^2\right)$$
$$+ 4\lambda^2\left[\Phi(\sqrt{2}\lambda) - \Phi(\lambda)\right] + o(1), \quad u = 1, 4. \quad (26)$$

Proof: See Appendix C. We denote by $\gamma(a, x) = (1/\Gamma(a))\int_0^x s^{a-1}e^{-s}ds$. ∎

As the representation of the image will be sparse, determining the risk of a coefficient if no signal is present, is important. From the corollary, and the asymptotic forms in λ, given in Appendix C, we may note that the risk at the "universal" threshold when there is no signal present is of the same order for $R_0^{(c)}(\sqrt{2\log(K)})$ and $R_0^{(r_1)}(\sqrt{2\log(K)})$. The coefficients multiplying the order terms differ in favor of $R_0^{(c)}(\sqrt{2\log(K)})$. $R_0^{(h)}(\sqrt{2\log(K\log(K))})$ and $R_0^{(a)}(\sqrt{2\log(K)})$ correspond to different orders (larger risks). The thresholds were introduced to improve the estimation of signals whose representations were slightly more spread across coefficients than strictly necessary. The risk for any coefficient when no signal is present increases marginally, to make the difference in estimation of the noisy coefficients between hyperanalytic and hard thresholding for large samples, negligible [i.e., $O(\log(K)/K)$ rather than $O(1/K\log(K))$]. The examples will substantiate this claim.

IV. EXAMPLES

To examine the properties of the proposed methods, we have implemented simulation studies on images that can be retrieved at http://sipi.usc.edu/database/ (Tiffany and Boat), while (Lena and MRIScan) are downloaded from http://www-stat.stanford.edu/~wavelab/. We used LA wavelets length 8. To compare our results, similarly to [3], we also implemented usual hard thresholding and the wavelet-domain HMM [7], where the software is available at http://www-dsp.rice.edu/software/. The method will be denoted by $s = \text{hmm}$. We used the code hdenoise.m with default settings, and $\text{daubcqf}(8,' \min')$. We implemented the method at several SNRs of 2 (very noisy), 4 (moderately noisy), and 8 (quite clean), with a set of images, i.e., Lena (512 × 512 version), Boat (512 × 512), MRIScan (256 × 256), and the second channel of the color image of Tiffany (512 × 512). The SNR is (as usual) given by $\text{SNR}^2 = (1/N^2\sigma^2)\sum\sum q_{\mathbf{x}}^2$. Table I shows the result over repeated simulations. Each column of the table is presenting results for a different image, at three different SNRs separated by/sign. The rows present average MSEs, average standard deviations (SDs) of the MSEs, and average PSNRs over the 50 runs. We index the rows of the table by n (noisy signal), c (classical thresholding with the "universal" threshold), r (Riesz thresholding with the "universal" threshold), h (Hypercomplex thresholding with the "universal" threshold), and hmm (the HMM method), thus indicating the method used for obtaining the estimator of the image. Reduced MSEs and increased
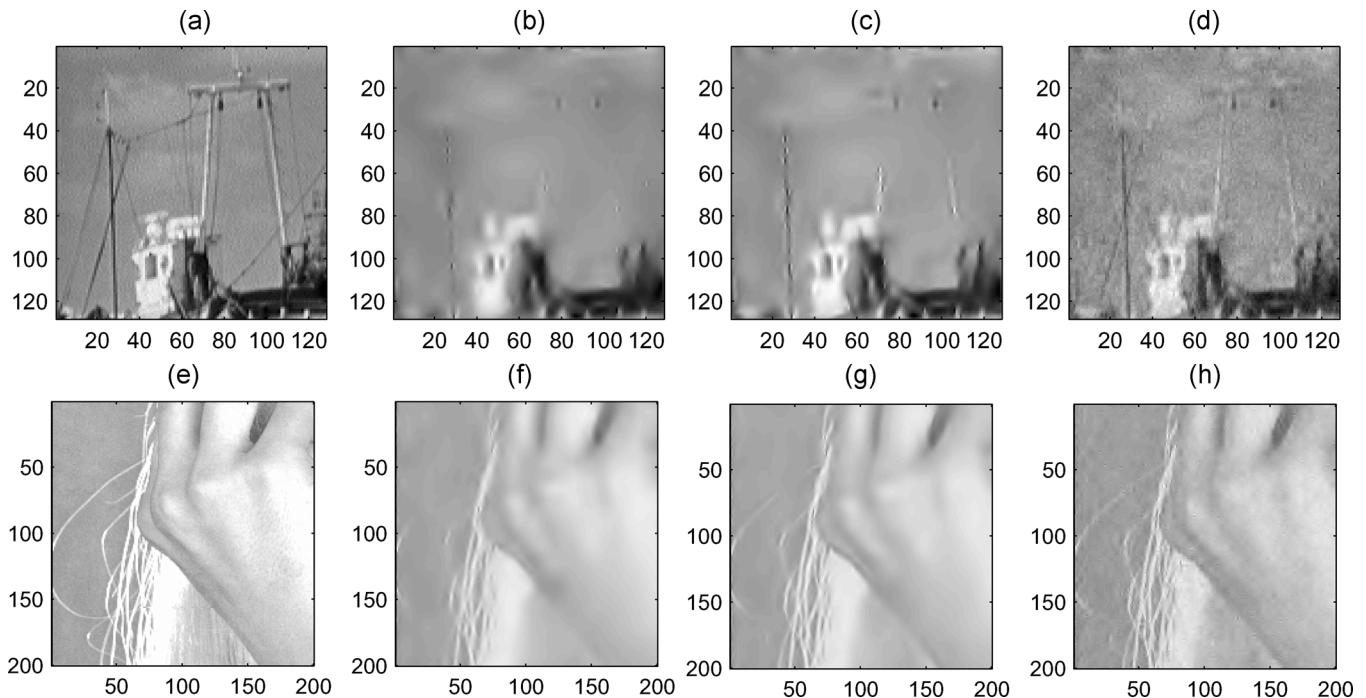
Fig. 2. Row 1: Estimating a section of the noisy boat image ($SNR = 5.56$), with (a) the clean image, (b) the "universal" hard threshold estimate, (c) the "universal" hypercomplex estimate, and (d) the HMM estimate. Row 2: Section of one of the bands of the noisy Tiffany image ($SNR = 8$), (e) with the clean image, (f) the "universal" hard threshold estimate, (g) the "universal" hypercomplex estimate, and (h) the HMM estimate.

peak signal-to-noise ratios (PSNRs) (using the definition of [15]) are observed when using the proposed method with the hyperanalytic threshold criterions. The reduction in MSE is of a respectable magnitude compared to variation across replications. Overall the hypercomplex thresholding procedure is outperforming the Riesz thresholding, as well as the other methods, apart from the boat image at high SNRs where the HMM method performs better. The Hypercomplex method is expected to outperform the Riesz method from the risk calculations, but not perhaps from our discussion in Section II-C and D. The RT threshold may appear more suitable for improving denoising in a separable basis, as it calculates a replicate in the prevalent local direction from the image, while the HCT thresholding simply decreases the risk in estimation by considering variation associated with the same time-frequency (i.e., 1-D) behavior in both axes separately. However, while the RT is mainly suitable to use on locally unidirectional structure, the HCT allows for variation in several directions at the same spatial point. Images quite frequently have multidirectional variations present even locally, and so this method is borrowing more information across the image.

Finally, to consider the visual appearance of the reconstructions, see Fig. 2(a)–(h). We show the HCT reconstructions only, as the RT and HCT reconstructions are similar. The HCT estimate in Fig. 2(c) achieves greater continuity in reconstructing entire features [see, for example, the mast] than that achieved by Fig. 2(b), i.e., classical thresholding, and preserves more detailed features. We observe similar results in Fig. 2(g) compared to Fig. 2(f). Observe, for example, the curved loose strand of hair in Fig. 2(g). The HMM estimate has more noisy background reconstruction [see Fig. 2(d) and (h)]. This feature has also been observed by Starck *et al.* [3, p. 680]. The HMM method captures image detail well, as can be seen from some fine linear structures in the background of the boat image in Fig. 2(c). However, the retained noise makes it harder to make out these detailed features. An interesting feature from comparing Fig. 2(a) to (c) is that the bright outlines of some of the masts, are preserved in the HCT estimate, but not by the other methods. Overall the proposed method, despite its simplicity, represents a good improvement to standard "universal" denoising. It is surprisingly competitive to more sophisticated methods, as is verified by the reported SNRs.

In addition to the SNRs chosen for the full range of images, we implemented the procedure at the SNR chosen by Starck *et al.* [3], namely adding Gaussian noise with a standard deviation of 20 to the raw Lena image, or using a SNR of 5.58. The PSNR we observed in the noisy image (21.58) is less than theirs (22.13), but that is to be expected in two different noisy replications. We found for the methods discussed in this paper, that averaged over 100 replications the PSNRs were $s = c\,(29.22)$, $s = r\,(30.12)$, $s = h\,(30.93)$, and $s = \mathrm{hmm}\,(30.48)$. Starck *et al.* obtained for $s = c\,(28.35)$ and $s = \mathrm{hmm}\,(30.80)$. Starck *et al.* obtained PSNRs between 29.99 to 31.95 by using local ridgelets and curvelets. Clearly, the hyperanalytic denoising methods proposed by this paper perform on par with competitive algorithms. Some additional image examples can also be found in [27]. In addition, the proposed procedure is both cheap to implement and extremely simple to code.

V. CONCLUSION

This paper has proposed a new thresholding estimator for the estimation of the separable DWT. We determined both the

TABLE I
AVERAGE RESULTS OVER 50 RUNS. THE SYMMLET WAVELETS (OR LA WAVELETS WERE USED) AND $J = 3$

Example (SNR)	method	Boat (2/4/8)	Lena (2/4/8)	Tiffany (2/4/8)	MRIScan (2/4/8)
average MSE $\times 10^{-8}$	n	95.4/23.8/6.0	95.4/23.8/6.0	95.4/23.8/6.0	381.5/95.4/23.8
average MSE $\times 10^{-8}$	c	8.0/4.4/2.2	5.2/2.6/1.1	3.8/2.2/1.0	56.9/24.5/9.1
average MSE $\times 10^{-8}$	r	7.1/3.8/1.8	4.5/2.2/0.93	3.4/1.9/0.9	46.9/19.7/7.3
average MSE $\times 10^{-8}$	h	**6.3**/3.2/1.6	**3.9/1.9/0.77**	**3.1/1.7/0.8**	**40.4/16.6/6.1**
average MSE $\times 10^{-8}$	hmm	6.7/**3.0/1.4**	6.3/2.1/ 0.85	5.4/1.8/**0.8**	48.9/20.9/7.5
sd MSE $\times 10^{-10}$	n	26.3/6.6/1.6	26.3/6.6/1.6	26.3/6.6/1.6	187.9/47.0/11.7
sd MSE $\times 10^{-10}$	c	6.7/3.2/1.2	7.3/2.3/0.8	4.9/2.5/0.8	110.0/43.4/15.3
sd MSE $\times 10^{-10}$	r	6.2/2.7/1.0	5.6/2.1/0.7	3.8/2.4/0.7	89.2/36.8/11.6
sd MSE $\times 10^{-10}$	h	6.6/2.2/0.9	4.7/2.0/0.6	3.3/2.3/0.6	79.1/31.0/8.7
sd MSE $\times 10^{-10}$	hmm	11.3/1.8/0.7	16.0/2.3/0.5	16.4/1.2/2.7	82.1/56.9/7.4
PSNR	n	11.36/17.38/23.40	10.67/16.69/22.71	7.65/13.67/19.69	16.96/22.98/29.00
PSNR	c	22.12/24.68/27.70	23.34/26.34/29.88	21.64/24.02/27.39	25.22/28.88/33.18
PSNR	r	22.63/25.39/28.49	23.96/27.09/30.77	22.09/24.73/27.99	26.06/29.82/34.17
PSNR	h	**23.14**/ 26.04/29.23	**24.56/27.74/31.58**	**22.48/25.06/28.55**	**26.71/30.57/34.92**
PSNR	hmm	22.90/**26.38/29.62**	22.48/27.31/31.19	20.12/24.78/28.26	25.88/29.57/34.01

deterministic (for some stylized image features) and stochastic properties of the suggested new thresholding estimator. We established "universal" thresholds. We calculated the risk theoretically, and for some specific choices of the mean local coefficients, provided plots of the risk, showing that the proposed methods outperform classical denoising theoretically. We implemented the procedure on several examples, at several SNRs, comparing the methods with the hidden Markov model, as well as "universal" hard thresholding. We found that the proposed algorithms offered improvements in most cases. Given the simplicity of the implementation, and visually pleasing reconstructions, hyperanalytic denoising methods offer a computationally cheap improvement to existing methodology.

APPENDIX A
DIGITAL IMPLEMENTATION

The discrete Fourier transform (DFT) and its inverse are given with $\boldsymbol{\Omega} = [-(1/2\Delta x), (1/2\Delta x)]^2$ by

$$Q_D(f_1, f_2) = \Delta x^2 \sum_{x_1=0}^{N-1} \sum_{x_2=0}^{N-1} q(x_1 \Delta x, x_2 \Delta x) e^{-2i\pi \mathbf{f}^T \mathbf{x} \Delta x}$$

$$q_{\mathbf{x}} = \iint_{\boldsymbol{\Omega}} Q_D(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2 \mathbf{f}. \qquad (27)$$

The value of $Q_D(f_1, f_2)$ at frequencies $f_1 \in \{((N/2)/N\Delta x), ((N/2 + 1)/N\Delta x), ((N/2 + 2)/N\Delta x), \ldots, ((N-1)/N\Delta x)\}$ is equivalent to the value of $Q_D(f_1, f_2)$ at frequencies $f_1 \in \{-((N/2)/N\Delta x), -((N/2 - 1)/N\Delta x), -((N/2 - 2)/N\Delta x), \ldots, -(1/N\Delta x)\}$, and the equivalent statement, *mutatis mutandis*, hold for f_2. Let $Q_{D,\mathbf{p}} = Q_D(p_1/(N\Delta x), p_2/(N\Delta x))$, with $\mathbf{p} = [p_1, p_2]^T$, then

$$\breve{q}_{\mathbf{x}}^{(s,l)} = \iint_{\boldsymbol{\Omega}} Q_D(\mathbf{f}) V_D^{(s,l)}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x} \Delta x} d^2 \mathbf{f}$$

$$= \frac{1}{\Delta x^2 N^2} \sum_{p_1=0}^{N-1} \sum_{p_2=0}^{N-1} Q_{D,\mathbf{p}} V_{D,\mathbf{p}}^{(s,l)} e^{2i\pi \mathbf{p}^T \mathbf{x}} + O\left(\frac{1}{N}\right)$$

$s = r, h$, and $l = 1, \ldots, L$. As the discrete RTs/HCTs, thus, have an error of $1/(N\Delta x)$ associated with approximating the continuous versions, the discrete implementation will not exactly meet the requirements of quadrature components. We define

$$r_{\mathbf{p}} = \begin{cases} \sqrt{(p_1/(N\Delta x))^2 + (p_2/(N\Delta x))^2} \\ \quad \text{if } \mathbf{p} \in [0, N/2]^2/\{\mathbf{0}\} \\ \sqrt{((N-p_1)/(N\Delta x))^2 + (p_2/(N\Delta x))^2} \\ \quad \text{if } \mathbf{p} \in [N/2+1, N-1] \times [0, N/2] \\ \sqrt{(p_1/(N\Delta x))^2 + ((N-p_2)/(N\Delta x))^2} \\ \quad \text{if } \mathbf{p} \in [0, N/2] \times [N/2+1, N-1] \\ \sqrt{((N-p_1)/(N\Delta x))^2 + ((N-p_2)/(N\Delta x))^2} \\ \quad \text{if } \mathbf{p} \in [N/2+1, N-1]^2. \end{cases} \qquad (28)$$

We let $V_{D,\mathbf{0}}^{(r,1)} = V_{D,\mathbf{0}}^{(r,2)} = 0$ and define for $\mathbf{p} \neq \mathbf{0}$, with $\mathbf{V}_{D,\mathbf{p}}^{(r)} = [V_{D,\mathbf{p}}^{(r,1)}, V_{D,\mathbf{p}}^{(r,2)}]$

$$\boldsymbol{V}_{D,\mathbf{p}}^{(r)} = \begin{cases} -i(r_{\mathbf{p}} N\Delta x)^{-1} \mathbf{p} \quad \text{if } p_1, p_2 = 0, \ldots, N/2, \\ -i(r_{\mathbf{p}} N\Delta x)^{-1}[N - p_1, p_2] \\ \quad \text{if } p_1 = N/2+1, \ldots, N-1, p_2 = 0, \ldots, N/2 \\ -i(r_{\mathbf{p}} N\Delta x)^{-1}[p_1, N - p_2] \\ \quad \text{if } p_1 = 0, \ldots, N/2, p_2 = N/2+1, \ldots, N-1 \\ -i(r_{\mathbf{p}} N\Delta x)^{-1}[N - p_1, N - p_2] \\ \quad \text{if } p_1, p_2 = N/2+1, \ldots, N-1. \end{cases}$$

The discrete implementation of the HCT filters are based on the partial HTs. Define $l' = 1$ if $l = 2$ and $l' = 2$ if $l = 1$. Then, for all $p_{l'} \in [0, N-1]$

$$V_{D,\mathbf{p}}^{(h,l)} = \begin{cases} 0, & \text{if } p_l = 0, N/2, \\ -i, & \text{if } p_l = 1, \ldots, N/2-1 \\ i, & \text{if } p_l = N/2+1, \ldots, N-1 \end{cases}, \quad l = 1, 2$$

and $V_{D,\mathbf{p}}^{(h,3)} = V_{D,\mathbf{p}}^{(h,1)} V_{D,\mathbf{p}}^{(h,2)}$. Implementing the discrete HCTs introduces an error term of $O(1/N)$, just like the error introduced in the discrete RTs.

APPENDIX B
DWT OF TYPICAL DETERMINISTIC IMAGE FEATURES

Proof of Lemmas 1 and 2: We use periodic boundary treatment (see Mallat [12, pp. 282–292]) when implementing the MODWT

$$\widetilde{W}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(q)} = \int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f}) Q_D(\mathbf{f}) e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} \qquad (29)$$

with $\boldsymbol{\xi}_{\mathbf{x}} = [j, u, \mathbf{x}]^T$. Let $\mathbf{f}_0(\mathbf{x}) = f_0(\mathbf{x}) [\cos(\phi_0(\mathbf{x})) \sin(\phi_0(\mathbf{x}))]^T$, where we take $f_0(\mathbf{x}) > 0$, and $0 \le \phi_0(\mathbf{x}) < \pi$. By direct calculation using (29), and approximating the DFT of the $\cos(2\pi \mathbf{f}_0(\mathbf{x})^T \mathbf{x}')$ by $\delta(\cdot)$ distributions (this incurring the $O(1/N)$ error term), it then follows that:

$$\mu_{\boldsymbol{\xi},1}^{(s)} = 2^j \int\int_{\boldsymbol{\Omega}} \frac{a_t(\mathbf{x})}{2} \left(\delta\left(\mathbf{f}_0(\mathbf{x}) - \mathbf{f}\right) + \delta\left(\mathbf{f}_0(\mathbf{x}) + \mathbf{f}\right)\right)$$
$$\times \widetilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$
$$= 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right|$$
$$\times \cos\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right)\right)$$
$$+ O\left(\frac{1}{N}\right), \quad s = r, h$$
$$\mu_{\boldsymbol{\xi},2}^{(r)} = 2^j \int\int_{\boldsymbol{\Omega}} \frac{a_t(\mathbf{x})}{2}(-i)\frac{f_1}{f} \left(\delta(\mathbf{f}_0(\mathbf{x}) - \mathbf{f}) + \delta(\mathbf{f}_0(\mathbf{x}) + \mathbf{f})\right)$$
$$\times \widetilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$
$$= 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right| \cos\left(\phi_0(\mathbf{x})\right)$$
$$\times \sin\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right)\right) + O\left(\frac{1}{N}\right)$$

and similarly
$$\mu_{\boldsymbol{\xi},3}^{(r)} = 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right| \sin\left(\phi_0(\mathbf{x})\right)$$
$$\times \sin\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}(\mathbf{f}_0)\right)\right) + O\left(\frac{1}{N}\right).$$

Thus, $(1/2)\sum_l \mu_{\boldsymbol{\xi},l+1}^{(r)2} = 2^{2j-1} a_t^2(\mathbf{x})|\widetilde{H}_{j,u}(\mathbf{f}_0(\mathbf{x}))|^2 + O(1/N)$. ρ_1 is introduced when $|\widetilde{H}_{j,u}(\mathbf{f}_0(\mathbf{x}))|$ is approximated by $|\widetilde{H}_{j,u}(\mathbf{f}_0(\mathbf{x}))| = I(\mathbf{f}_0(\mathbf{x}) \in \Omega_{j,u}) + \rho'_1$, where ρ'_1 depends on the wavelet used.

By direct calculation, using (29), it follows that:

$$\mu_{\boldsymbol{\xi},2}^{(h)} = 2^j \int\int_{\boldsymbol{\Omega}} \frac{a_t(\mathbf{x})}{2}(-i)\mathrm{sgn}(f_1)$$
$$\times \left(\delta\left(\mathbf{f}_0(\mathbf{x}) - \mathbf{f}\right) + \delta\left(\mathbf{f}_0(\mathbf{x}) + \mathbf{f}\right)\right)$$
$$\times \widetilde{H}_{j,u}(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$
$$= 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right| \mathrm{sgn}\left(\cos\left(\phi_0(\mathbf{x})\right)\right)$$
$$\times \sin\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}(\mathbf{f}_0)\right)\right) + O\left(\frac{1}{N}\right).$$

Similarly

$$\mu_{\boldsymbol{\xi},3}^{(h)} = 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right| \sin\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}(\mathbf{f}_0)\right)\right)$$
$$+ O\left(\frac{1}{N}\right)$$
$$\mu_{\boldsymbol{\xi},4}^{(h)} = 2^j a_t(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right| \mathrm{sgn}\left(\cos\left(\phi_0(\mathbf{x})\right)\right)$$
$$\times \cos\left(2\pi\left(\mathbf{f}_0^T(\mathbf{x})\mathbf{x} - \varphi_{j,u}(\mathbf{f}_0)\right)\right) + O\left(\frac{1}{N}\right).$$

Thus, it transpires that

$$\frac{1}{4}\sum_l \mu_{\boldsymbol{\xi},l+1}^{(h)2} = 2^{2j-1} a_t^2(\mathbf{x}) \left|\widetilde{H}_{j,u}\left(\mathbf{f}_0(\mathbf{x})\right)\right|^2 + O\left(\frac{1}{N}\right).$$

ρ_2 is introduced when $|\widetilde{H}_{j,u}(\mathbf{f}_0(\mathbf{x}))|$ is approximated.

Proof of Lemmas 3 & 4: Assume for simplicity of exposition, $0 < \theta \le \pi/2$, but with some notational changes no such restriction needs to be made. Define the rotation matrix by $\mathbf{r}_\theta = [\cos(\theta) - \sin(\theta), \sin(\theta)\cos(\theta)]$, and the change of variable given by: $\boldsymbol{\nu} = \mathbf{f}(-\theta) = [f_1(-\theta)f_2(-\theta)]^T = \mathbf{r}_{-\theta}\mathbf{f}$. We assume that $A_e(f)$ decays for large frequencies, $A_e(f) = A_e(0)\delta(f - 0)$ is an example of such structure. For example, if $a_e(x) = I(x \in [0,L))/\sqrt{L}$ then $A_e(f) = (e^{-i\pi fL}\sin(\pi fL))/(\pi f\sqrt{L})$, that as $L \to \infty, A_e(f)$ will concentrate to $f = 0$. $E(\mathbf{f}) = (A_e(-f_2(-\theta)/\sin(-\theta))/|\sin(\theta)|)e^{-2\pi i c f_2/\sin(\theta)}$, and with $\boldsymbol{\Omega}(\theta)$ the rotated by θ version of $\boldsymbol{\Omega}$, with $\nu_{1,\min} = \min_\nu \nu = \sec(\theta)f_1 I(|\tan(\theta)f_1| < 1/(2\Delta x))I(|f_1| < 1/(2\Delta x))$ and $\nu_{1,\max} = \max_\nu \nu = \sec(\theta)f_1 I(|\tan(\theta)f_1| < 1/(2\Delta x))I(|f_1| < 1/(2\Delta x))$, then

$$\mu_{\boldsymbol{\xi},1}^{(s)} = 2^j \int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f})\frac{A_e\left(-f_2(-\theta)/\sin(\theta)\right)}{\sin(\theta)} e^{-2\pi i c f_2/\sin(\theta)}$$
$$\times e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right), \quad s = r, h$$
$$= \int\int_{\boldsymbol{\Omega}(\theta)} \widetilde{H}_{j,u}(\mathbf{r}_\theta\boldsymbol{\nu})\frac{2^j A_e\left(-\nu_2/\sin(\theta)\right)}{\sin(\theta)}$$
$$\times e^{-2\pi i\left((\sin(\theta)\nu_1 + \cos(\theta)\nu_2)c/\sin(\theta) - \boldsymbol{\nu}^T \mathbf{r}_\theta^T \mathbf{x}\right)} d^2\boldsymbol{\nu}$$
$$+ O\left(\frac{1}{N}\right)$$
$$\stackrel{(1)}{=} 2^j A_e(0) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \widetilde{H}_{j,u}\left(\cos(\theta)\nu_1, \sin(\theta)\nu_1\right)$$
$$\times e^{2\pi i \nu_1(\cos(\theta)x_1 + \sin(\theta)x_2 - c)} d\nu_1 + \rho_3 + O\left(\frac{1}{N}\right).$$

The approximation in (1) relies on $A_e(\cdot)$ taking the form of a $\delta(\cdot)$ distribution contribution, i.e., $a_e(\cdot)$ constant over a large spatial domain, but a slowly varying $a_e(\cdot)$, will approximately produce the same result. Also, we may find approximate

descriptions for decomposition of the RTs, namely with $f = \sqrt{f_1^2 + f_2^2}$ and $\nu = \sqrt{\nu_1^2 + \nu_2^2}$

$$\mu_{\boldsymbol{\xi},2}^{(r)} = 2^j \int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f})(-i)f_1/f \frac{A_e\left(-f_2(-\theta)/\sin(\theta)\right)}{\sin(\theta)}$$

$$\times e^{-2\pi i c f_2/\sin(\theta)} e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$

$$\mu_{\boldsymbol{\xi},2}^{(r)} \stackrel{(1)}{=} A_e(0) 2^j (-i) \cos(\theta)$$

$$\times \int_{\nu_{1,\min}}^{\nu_{1,\max}} \widetilde{H}_{j,u}\left(\cos(\theta)\nu_1, \sin(\theta)\nu_1\right)$$

$$\times e^{2\pi i \nu_1(\cos(\theta)x_1 + \sin(\theta)x_2 - c)} \operatorname{sgn}(\nu_1) d\nu_1$$

$$+ \rho_4' + O\left(\frac{1}{N}\right)$$

$$= \cos(\theta) U_{\boldsymbol{\xi}}^{(e)} + \rho_4' + O\left(\frac{1}{N}\right). \tag{30}$$

The approximation in (1) relies on the envelope being constant in the spatial domain. Departures from this assumption introduces an error term, ρ_4'. Similarly, it transpires that $\mu_{\boldsymbol{\xi},3}^{(r)} = \sin(\theta) U_{\boldsymbol{\xi}}^{(e)} + \rho_7 + O(1/N)$, and, thus, the result follows as $(1/2)\sum_l \mu_{\boldsymbol{\xi},l+1}^{(r)2} = (1/2)(W_{\boldsymbol{\xi}}^{(e)2} + U_{\boldsymbol{\xi}}^{(e)2}) + \rho_5 + O(1/N)$, with a new error term ρ_5.

Also, we may find approximate descriptions for the hyper-complex components, namely

$$\mu_{\boldsymbol{\xi},2}^{(h)} = 2^j \int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f})(-i)\operatorname{sgn}(f_1) \frac{A_e\left(-f_2(\theta)/\sin(-\theta)\right)}{\sin(\theta)}$$

$$\times e^{-2\pi i c f_2/\sin(\theta)} e^{2\pi i \mathbf{f}^T \mathbf{x}} d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$

$$\stackrel{(1)}{=} 2^j A_e(0)(-i) \int_{\nu_{1,\min}}^{\nu_{1,\max}} \widetilde{H}_{j,u}\left(\cos(\theta)\nu_1, \sin(\theta)\nu_1\right)$$

$$\times e^{2\pi i \nu_1(\cos(\theta)x_1 + \sin(\theta)x_2 - c)} \operatorname{sgn}(\nu_1) d\nu_1$$

$$+ \rho_8 + O\left(\frac{1}{N}\right)$$

$$= U_{\boldsymbol{\xi}}^{(e)} + \rho_8 + O\left(\frac{1}{N}\right). \tag{31}$$

Similarly, it transpires that $\mu_{\boldsymbol{\xi},3}^{(h)} = U_{\boldsymbol{\xi}}^{(e)} + \rho_9 + O(1/N)$ and $\mu_{\boldsymbol{\xi},4}^{(h)} = W_{\boldsymbol{\xi}}^{(e)} + \rho_{10} + O(1/N)$. ρ_8, ρ_9 and ρ_{10} are constants depending on the variability of $a_e(\mathbf{x})$. Again, by squaring and adding the DWT of the quadrature terms, the result follows.

APPENDIX C
STATISTICS OF THE NORMAL VECTOR AND RISK CALCULATIONS

Set $\Delta x = 1$ when deriving the statistical properties of the coefficients. $\epsilon_{\mathbf{x}}$ then has a spectral representation: $\epsilon_{\mathbf{x}} = \int_{\boldsymbol{\Omega}} dZ_\epsilon(\mathbf{f}) e^{2i\pi \mathbf{f}^T \mathbf{x}}$, where $Z_\epsilon(\mathbf{f})$ is a complex-valued

orthogonal increment process, see [28] [p. 244], i.e., $\mathrm{E}(dZ_\epsilon(\mathbf{f})dZ_\epsilon^*(\mathbf{f}')) = 0$ if $\mathbf{f} \neq \mathbf{f}'$. The DWT is calculated by scaling and subsampling the MODWT. We define the MODWT of the quadrature components of noise: $\widetilde{\mathbf{n}}^{(s,u)} = [\widetilde{W}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(\epsilon,s,0)}, \ldots, \widetilde{W}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(\epsilon,s,L)}]^T$. $\mathbf{n}^{(s,u)}$, is a subsample of $\widetilde{\mathbf{n}}^{(s,u)}$, once suitably renormalized. Note that

$$\operatorname{cov}\left(n_{l_1}^{(s,u)}, n_{l_2}^{(s,u)}\right) = 2^{2j} \operatorname{cov}\left(\widetilde{n}_{l_1}^{(s,u)}, \widetilde{n}_{l_2}^{(s,u)}\right) \tag{32}$$

$l_1, l_2 = 1, \ldots, L + 1$. We have $\widetilde{n}_1^{(s,u)} = \int_{\boldsymbol{\Omega}} \widetilde{\widetilde{H}}_{j,u}(\mathbf{f}) dZ_\epsilon(\mathbf{f}) e^{2\pi i \mathbf{f}^T \mathbf{x}}$, $\widetilde{H}_{j,1}(\mathbf{f}) = \widetilde{H}_j(f_1)\widetilde{H}_j(f_2)$, $\widetilde{H}_{j,2}(\mathbf{f}) = \widetilde{H}_j(f_1)\widetilde{G}_j(f_2)$, $\widetilde{H}_{j,3}(\mathbf{f}) = \widetilde{G}_j(f_1)\widetilde{H}_j(f_2)$ and $\widetilde{H}_{j,4}(\mathbf{f}) = \widetilde{G}_j(f_1)\widetilde{G}_j(f_2)$. Clearly, we have $\operatorname{Var}(n_1^{(s,u)}) = \sigma^2$. We approximate the magnitude of the wavelet filters as exact bandpass filters—see, for example, Nielsen [19] for a discussion of such approximations. That is we approximate the passbands by

$$\left|\widetilde{H}_j(f)\right|^2 = \begin{cases} 1, & \text{if } |f| \in \left[\frac{1}{2^{j+1}}, \frac{1}{2^j}\right) \\ 0, & \text{if } |f| \in \left[\frac{1}{2^{j+1}}, \frac{1}{2^j}\right) \end{cases}$$

$$\left|\widetilde{G}_j(f)\right|^2 = \begin{cases} 1, & \text{if } f \in \left(-\frac{1}{2^{j+1}}, \frac{1}{2^{j+1}}\right) \\ 0, & \text{if } |f| \in \left[\frac{1}{2^{j+1}}, \frac{1}{2}\right). \end{cases} \tag{33}$$

Proof of Propositions 1 and 2:

$$\operatorname{var}\left(n_{l+1}^{(s,u)}\right)$$

$$= 2^{2j} E\left(\int\int_{\boldsymbol{\Omega}}\int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f})\widetilde{H}_{j,u}^*(\mathbf{f}')\right.$$

$$\left. e^{2\pi i(\mathbf{f}^T - \mathbf{f}'^T)\mathbf{x}} dZ_\epsilon(\mathbf{f}) dZ_\epsilon^*(\mathbf{f}')\right)$$

$$= 2^{2j} \int_{\boldsymbol{\Omega}} \left|V_D^{(s,l)}(\mathbf{f})\right|^2 \left|\widetilde{H}_{j,u}(\mathbf{f})\right|^2 d^2\mathbf{f}$$

$$= 2^{2j} \int_{-\infty}^{\infty} \left|V^{(s,l)}(\mathbf{f})\right|^2 \left|\widetilde{H}_{j,u}(\mathbf{f})\right|^2 d^2\mathbf{f} + O\left(\frac{1}{N}\right)$$

$$\equiv a_l^{(s,u)} + O\left(\frac{1}{N}\right), \quad l = 1, \ldots, L.$$

The latter defines $a_l^{(s,u)}$, a quantity that will be calculated for $s = r, h$, and $u = 1, 2, 3, 4$. For $l_1 \neq l_2$, we determine $\breve{h}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(s,l)} = \int\int_{\boldsymbol{\Omega}} V_D^{(s,l)*}(\mathbf{f})\widetilde{H}_{j,u}(\mathbf{f}) d^2\mathbf{f}$ and we note that

$$\operatorname{cov}\left(n_{l_1+1}^{(s,u)}, n_{l_2+1}^{(s,u)}\right) = 2^{2j} \int_{\boldsymbol{\Omega}} \sigma^2 V_D^{(s,l_1)*}(\mathbf{f}) V_D^{(s,l_2)}(\mathbf{f})$$

$$\times \left|\widetilde{H}_{j,u}(\mathbf{f})\right|^2 d^2\mathbf{f}$$

$$= 2^{2j} \sum_{\mathbf{x}} \breve{h}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(s,l_1)} \breve{h}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(s,l_2)*} + O\left(\frac{1}{N}\right)$$

$$= O\left(\frac{1}{N}\right) \tag{34}$$

$l_1, l_2 = 1, \ldots, L$, by property 2) of definition 2.1, as $\widetilde{h}_{\boldsymbol{\xi}_{\mathbf{x}}}$ is separable. Therefore

$$
\begin{aligned}
\mathrm{E}\left(\sum_{l=1}^{L} n_{l+1}^{(s,u)2}\right) &= 2^{2j} \int_{\boldsymbol{\Omega}} \sum_{l=1}^{L} \left|V^{(s,l)}(\mathbf{f})\right|^2 \sigma^2 \left|\widetilde{H}_{j,u}(\mathbf{f})\right|^2 d^2\mathbf{f} \\
&\quad + O\left(\frac{1}{N}\right) \\
&= 2^{2j} \int_{\boldsymbol{\Omega}} C_L^{(s)} \sigma^2 \left|\widetilde{H}_{j,u}(\mathbf{f})\right|^2 d^2\mathbf{f} + O\left(\frac{1}{N}\right) \\
&= C_L^{(s)} \mathrm{var}\left(n_1^{(s,u)2}\right) + O\left(\frac{1}{N}\right) \\
&= C_L^{(s)} \sigma^2 + O\left(\frac{1}{N}\right).
\end{aligned}
\tag{35}
$$

Hence, the total energy of the quadrate components of the noise is a constant times the variance of the original signal

$$
\begin{aligned}
&\mathrm{cov}\left(n_{l+1}^{(s,u)}, n_1^{(s,u)}\right) \\
&= 2^{2j} E\left(\int\int_{\boldsymbol{\Omega}}\int\int_{\boldsymbol{\Omega}} \widetilde{H}_{j,u}(\mathbf{f})\widetilde{H}_{j,u}^*(\mathbf{f}')\right. \\
&\qquad\qquad \left.\times dZ_{\breve{\epsilon}}(\mathbf{f})dZ_\epsilon^*(\mathbf{f}')e^{2\pi i(\mathbf{f}-\mathbf{f}')\mathbf{x}}\right) \\
&= \int_{\boldsymbol{\Omega}} \sigma^2 V^{(s,l)*}(\mathbf{f})\widetilde{H}_{j,u}(\mathbf{f})\widetilde{H}_{j,u}^*(\mathbf{f})d^2\mathbf{f} + O\left(\frac{1}{N}\right) \\
&= \sum_{\mathbf{x}} \breve{h}_{\boldsymbol{\xi}_{\mathbf{x}}}^{(s,l)} \widetilde{h}_{\boldsymbol{\xi}_{\mathbf{x}}} + O\left(\frac{1}{N}\right) \\
&= O\left(\frac{1}{N}\right), \quad l = 1, \ldots, L.
\end{aligned}
$$

Proof of Lemma 5: The noise is Gaussian and zero-mean. To deduce the Lemma, using (32), we note

$$
\begin{aligned}
&\mathrm{Var}\left(n_2^{(r,1)}\right) \\
&= 2^{2j} E\left(\int\int_{\boldsymbol{\Omega}}\int\int_{\boldsymbol{\Omega}} \widetilde{H}_j(f_1)\widetilde{H}_j(f_2)dZ_\epsilon(\mathbf{f})e^{2\pi i(\mathbf{f}-\mathbf{f}')^T\mathbf{x}}\right. \\
&\qquad \times \frac{f_1}{\sqrt{f_1^2+f_2^2}}\widetilde{H}_j^*(f_1')\widetilde{H}_j^*(f_2')dZ_\epsilon(\mathbf{f}') \\
&\qquad \left.\times \frac{f_1'}{\sqrt{f_1'^2+f_2'^2}}\right) + O\left(\frac{1}{N}\right) \\
&\mathrm{Var}\left(n_l^{(r,l)}\right) \\
&= 2^{2j}\sigma^2 \int\int_{\boldsymbol{\Omega}} \frac{f_1^2}{f_1^2+f_2^2} \left|\widetilde{H}_j(f_1)\right|^2 \left|\widetilde{H}_j(f_2)\right|^2 d^2\mathbf{f} \\
&\quad + O\left(\frac{1}{N}\right), \quad l = 2, 3.
\end{aligned}
$$

Hence, it follows that (and *mutatis mutandis* for the other values of u)

$$
\begin{aligned}
\mathrm{Var}\left(n_2^{(r,1)}\right) + \mathrm{Var}\left(n_3^{(r,1)}\right) &= \mathrm{Var}\left(n_1^{(r,1)}\right), \\
\mathrm{Var}\left(n_2^{(r,u)}\right) + \mathrm{Var}\left(n_3^{(r,u)}\right) &= \mathrm{Var}\left(n_1^{(r,u)}\right), \quad u = 1, 2, 3, 4.
\end{aligned}
$$

We may note that $a_2^{(r,1)} = a_3^{(r,1)} = a_2^{(r,4)} = a_2^{(r,4)} = 1/2$, as

$$
\begin{aligned}
\mathrm{Var}\left(n_2^{(r,1)}\right) &= 2^{2j+2}\sigma^2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \frac{f_1^2}{f_1^2+f_2^2} df_2 df_1 + O\left(\frac{1}{N}\right) \\
&= \sigma^2 \left(\frac{1}{4} - \frac{1}{8}\tan^{-1}(2) + \frac{1}{2}\tan^{-1}\left(\frac{1}{2}\right)\right. \\
&\qquad \left. - \left(\frac{1}{8} - \frac{1}{8}\tan^{-1}(2) + \frac{1}{2}\tan^{-1}\left(\frac{1}{2}\right)\right)\right) \\
&\quad + O\left(\frac{1}{N}\right) \\
&= \frac{\sigma^2}{2} + O\left(\frac{1}{N}\right) = \mathrm{Var}\left(n_3^{(r,1)}\right) \\
&= \mathrm{Var}\left(n_l^{(r,4)}\right), \quad l = 2, 3.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\left(n_2^{(r,2)}\right) &= 2^{2j+2}\sigma^2 \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \int_0^{\frac{1}{2^{j+1}}} \frac{f_1^2}{f_1^2+f_2^2} df_2 df_1 + O\left(\frac{1}{N}\right) \\
&= \sigma^2 \left(\frac{1}{2} + 2\tan^{-1}\left(\frac{1}{2}\right) - \frac{1}{2}\tan^{-1}(2)\right) \\
&\quad + O\left(\frac{1}{N}\right) \\
&= \sigma^2 a^{(r,2)} + O\left(\frac{1}{N}\right) \\
&\approx 0.8737\sigma^2.
\end{aligned}
$$

Finally, note that

$$
\begin{aligned}
\mathrm{Var}\left(n_2^{(r,3)}\right) &= 2^{2j+2}\sigma^2 \int_0^{\frac{1}{2^{j+1}}} \int_{\frac{1}{2^{j+1}}}^{\frac{1}{2^j}} \frac{f_1^2}{f_1^2+f_2^2} df_2 df_1 + O\left(\frac{1}{N}\right) \\
&= \sigma^2 \left(1 - a^{(r,2)}\right) + O\left(\frac{1}{N}\right) \\
&\equiv \sigma^2 a^{(r,3)} + O\left(\frac{1}{N}\right) \approx 0.1263\sigma^2.
\end{aligned}
$$

Clearly, we can find the variance of the second RT by permuting the order of the spatial variable in the integration, and this then completes the variance calculations. From the proofs of propositions 1 and 2 we can note that the components of $\mathbf{n}^{(r,u)}$ for $u = 1, 2, 3, 4$, are uncorrelated up to $O(1/N)$. This can also be shown by direct calculation. As $\epsilon_{\mathbf{x}}$ was zero-mean Gaussian, and we are forming linear combinations to obtain $\mathbf{n}^{(r,u)}$, the stated result follows from the expressions for the covariances of the components.

Proof of Lemma 6:

a) First consider $u = 1, 4$, so that the variance of the two Riesz components is 1/2. Then by Lemma 5 it follows directly that $(C_L^{(s)} + 1)M_{\boldsymbol{\xi}}^{(\epsilon,r)2} \overset{\mathcal{L}}{=} \mathbf{Z}^{(r,u)T}\mathbf{Z}^{(r,u)} + O(1/N)$, $T_1 = \mathbf{Z}^{(r,u)T}\mathbf{Z}^{(r,u)} \sim \chi_1^2 + (1/2)\chi_1^2 + (1/2)\chi_1^2$. T_1 has a moment generating function (MGF) given by $M_{T_1}(s) = (1/(\sqrt{1-2s}))(1/(1-s))$, and, thus, $f_{T_1}(t) = e^{-t}(2/\sqrt{\pi})\int_0^{\sqrt{t/2}} e^{u^2} du$. The probability that T_1 does not exceed λ^2 is given by

$$P(T_1 < \lambda^2) = 1 + \left[e^{-x}\int_0^{\sqrt{\frac{x}{2}}} \frac{2}{\sqrt{\pi}}e^{u^2} du \right]_{\lambda^2}^{\infty} - \int_{\lambda^2}^{\infty} \frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{x}{2}}dx \quad (36)$$

where the cdf be found in Grad & Solomon [25, p. 472], and the function is expanded as $\lambda \to \infty$, to get $P(T_1 < \lambda^2) = 1 - \sqrt{(8/\pi)}(1/\lambda)e^{-\lambda^2/2} + e^{-\lambda^2/2}(O(1/\lambda^3))$.

b) Consider now $u = 2, 3$. Wlog assume that $a^{(r,u)} > 1/2$, and otherwise relabel $a^{(r,u)}$ and $1 - a^{(r,u)}$ suitably. If $a^{(r,u)} = 1 - a^{(r,u)}$ this collapses to the case given above. Define $T' = T_1/2$ so that $T' = \sum_{i=1}^3 a_i X_i^2$ where the X_i are iid Gaussian random variates with zero mean and unit variance, where $\sum_{i=1}^3 a_i = 1$. The MGF is by Grad and Solomon [25] with $a_1 = (1 - a^{(r,u)})/2$, $a_2 = a^{(r,u)}/2$, $a_3 = 1/2$

$$M_{T'}(s) = M_{T_1}(s/2)$$
$$= \left(1 - 2\left(1 - a^{(r,u)}\right)s/2\right)^{-\frac{1}{2}}$$
$$\times \left(1 - 2a^{(r,u)}s/2\right)^{-\frac{1}{2}}(1 - 2s/2)^{-\frac{1}{2}}$$
$$= \prod_{i=1}^3 (1 - 2a_i s)^{-\frac{1}{2}}.$$

We let $c_1 = 2/(1 - a^{(r,u)})$, $c_2 = 2/a^{(r,u)}$, $c_3 = 2$, and, hence, as we assumed $1 > a^{(r,u)} > 1 - a^{(r,u)}$, we note that $c_1 \geq c_2 \geq c_3$ and this choice agrees with the notation in [25]. For future reference, note that $c_1 + c_2 = 2/(a^{(r,u)}(1 - a^{(r,u)}))$, $u = 1, \ldots, 4$, and we may rewrite $M_{T'}(s) = \prod_{j=1}^3 (1 - 2s/c_j)^{-1/2}$. Using results from Grad and Solomon we may determine: $F_{T_1}(t) = F_{T'}(t/2) = (1 - Ae^{-c_3 t/4}(1/\sqrt{t})) + o(1)$, for constant A. Thus, the probability of an observation exceeding a large threshold, can be found.

Proof of Lemma 7: $\text{Var}(n_2^{(h,1)}) = \sigma^2 + O(1/N) = \text{Var}(n_l^{(h,u)})$, $u = 1, 2, 3, 4$, $l = 1, 2, 3, 4$. This follows from the form of the partial HT [10]. From the proofs of propositions (1) and (2) we note that the components of $\mathbf{n}^{(h,u)}$ for a fixed value of $u = 1, 2, 3, 4$ are uncorrelated up to $O(1/N)$. As $\{\epsilon_{\mathbf{x}}\}$ is a Gaussian process, it follows that $\mathbf{n}^{(h,u)}$, is a Gaussian vector.

Proof of Lemma 8: K wavelet coefficients will be thresholded, where K_1 magnitudes have the distribution given when $u = 1, 4$, and K_2 have the distribution that follows from $u = 2, 3$, where $K_1 + K_2 = K$, and $K_1, K_2, K_3 = O(K)$, so that by Dykstra [23]

$$P\left(\mathcal{M}_K^{(r)2} < \lambda_K^2\right) = \left(1 - \sqrt{\frac{8}{\pi}}\frac{1}{\lambda_K}e^{-\lambda_K^2/2}\right)^{K_1}$$
$$\left(1 - Ae^{-c_3\lambda_K^2/4}\frac{1}{\lambda_K}\right)^{K_2} = \left[1 - \frac{A_2}{\lambda_K}e^{-\lambda_K^2/2}\right]^{K_3}$$

ignoring $o(1)$ terms in K for a suitable constant A_2, if $\lambda_K^2 = O(\log[K])$. Thus

$$P\left(\mathcal{M}_K^{(2)} < \lambda_K^2\right) = \left(1 - \frac{A_2 e^{-\lambda_K^2/2}}{\lambda_K}\right)^{K_3} + o(1)$$
$$= \left(1 - \frac{A_2 e^{-(2\log[K] + C_2\log[\log[K]])/2}}{\sqrt{2\log[K] + C_2\log[\log[K]]}}\right)^{K_3}$$
$$+ o(1).$$

So, $P(\mathcal{M}_K^{(2)} < \lambda_K^2) \to 1$, if $C_2/2 > -1/2$, i.e., $C_2 > -1$, and we take $C_2 = 0$.

Proof of Theorem 1: From [26, p. 293] the risk of regular hard thresholding is given as stated in Theorem 1. For the hyperanalytic procedures when $\text{var}(n_{l+1}^{(s,u)}) = \sigma^2\sigma_{l+1}^2$, for $l = 0, \ldots, L$, where $\sigma_l^2 = 1$ or $\sigma_l^2 = a_{l-1}^{(s,u)}$, the risk is

$$R_{\boldsymbol{\theta}^{(s)}}^{(s)}(\lambda)$$
$$= \theta_1^2 \int_{\sum_l w_l^2 \leq \lambda^2} \prod_l \sigma_l^{-1}\phi\left(\frac{w_l - \theta_l^{(s)}}{\sigma_l}\right)dw_l$$
$$+ \int_{\sum_l w_l^2 > \lambda^2} [w_1 - \theta_1]^2 \prod_l \sigma_l^{-1}\phi\left(\frac{w_l - \theta_l^{(s)}}{\sigma_l}\right)dw_l + o(1)$$
$$= 1 + \int_{\Upsilon_s(\lambda)} [\theta_1^2 - w_1^2] \prod_l \sigma_l^{-1}\phi\left(\frac{w_l}{\sigma_l}\right)dw_l + o(1). \quad (37)$$

Proof of Corollary 1: $R_{\mathbf{0}}^{(c)}(\lambda) = 1 - \int_{-\lambda}^{\lambda} w^2 (1\sqrt{2\pi}) e^{-(1/2)w^2} dw = 1 - \gamma(1/2, (1/2)\lambda^2) = e^{-(1/2)\lambda^2}((\sqrt{2}/\sqrt{\pi}\lambda) + O(\lambda^{-3}))$.

$$R_{\mathbf{0}}^{(a)}(\lambda) = 1 - \int_{w_1^2 + w_2^2 \leq \lambda^2} w_1^2 \frac{1}{2\pi}e^{-\frac{1}{2}(w_1 + w_2)^2}dw_1 dw_2$$
$$= e^{-\frac{1}{2}\lambda^2}\left(1 + \frac{1}{2}\lambda^2\right).$$

For $s = r$ and $u = 1, 4$, denoted by $s = r_1$ in Fig. 1(b)–(d), it follows

$$
\begin{aligned}
R_0^{(r_1)}(\lambda) &= 1 - \int\limits_{w_1^2 + w_2^2 + w_3^2 \le \lambda^2} \frac{2 e^{-\frac{1}{2}\left(w_1^2 + 2\left(w_2^2 + w_3^2\right)\right)}}{(\sqrt{2\pi})^3} \\
&\quad \times w_1^2 \, dw_1 \, dw_2 \, dw_3 \\
&= 1 + \gamma\left(1/2, \frac{1}{2}\lambda^2\right) + 4\lambda^2 \left(\Phi(\sqrt{2}\lambda) - \Phi(\lambda)\right) \\
&\quad - 2\gamma(1/2, \lambda^2) \\
&= e^{-\frac{1}{2}\lambda^2} \left(\frac{4\sqrt{2}}{\sqrt{\pi}\lambda} + O(\lambda^{-3})\right)
\end{aligned}
\tag{38}
$$

while for $u = 2, 3$, denoted by $s = r_2$

$$
\begin{aligned}
R_0^{(r_2)}(\lambda) &= 1 - \int\limits_{\sum_l w_l^2 \le \lambda^2} w_1^2 \frac{1}{(\sqrt{2\pi})^3 \sqrt{a^{(r,u)}\left(1 - a^{(r,u)}\right)}} \\
&\quad \times e^{-\frac{1}{2}\left(w_1^2 + w_2^2/a^{(r,u)} + w_3^2/(1-a^{(r,u)})\right)} dw_1 \, dw_2 \, dw_3 \\
R_0^{(h)}(\lambda) &= 1 - \int\limits_{w_1^2 + w_2^2 + w_3^2 + w_4^2 \le \lambda^2} w_1^2 \frac{1}{(\sqrt{2\pi})^4} e^{-\frac{1}{2}\sum w_l^2} dw \\
&= 1 - \frac{\left(\frac{\pi}{4}\right)(2\pi)}{4\pi^2} \int_0^\lambda r^5 e^{-\frac{1}{2}r^2} dr \\
&= 1 - \frac{1}{8} \int_0^{\frac{1}{2}\lambda^2} 4s^2 e^{-s} ds \\
&= e^{-\frac{1}{2}\lambda^2} \left(1 + \frac{1}{2}\lambda^2 + \frac{1}{8}\lambda^4\right).
\end{aligned}
\tag{39}
$$

ACKNOWLEDGMENT

REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–55, 1994.

[2] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

[3] J. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[4] T. Cai and B. Silverman, "Incorporating information on neighbouring coefficients into wavelet estimation," *Sankhyā B*, vol. 63, pp. 127–148, 2001.

[5] P. Dragotti and M. Vetterli, "Wavelet footprints: Theory, algorithms and applications," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1306–1323, May 2003.

[6] A. Pižurica, W. I. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 5, pp. 545–557, May 2002.

[7] M. Crouse, R. Nowak, and Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[8] P. Fryzlewicz, "Bivariate hard thresholding in wavelet function estimation," *Statist. Sin.*, to be published.

[9] S. C. Olhede and A. T. Walden, "'Analytic' wavelet thresholding," *Biometrika*, vol. 91, pp. 955–973, 2004.

[10] S. L. Hahn, *Hilbert Transforms in Signal Processing*. Boston, MA: Artech House, 1996.

[11] B. Boashash, "Note on the use of the Wigner distribution for time-frequency signal analysis," *IEEE Trans. Accoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1518–1521, Sep. 1988.

[12] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[13] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[14] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[15] S. Barber and G. P. Nason, "Denoising real images using complex-valued wavelets," Stochastic Geometry, Biological Structure and Images, Dept. Statist., Univ. Leeds, Leeds, U.K., 2003, pp. 91–95.

[16] G. Metikas and S. C. Olhede, "Multiple multidimensional Morse wavelets," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 921–936, Mar. 2007.

[17] S. C. Olhede and G. Metikas, "The hyperanalytic wavelet transform," arXiv math.ST/0605623, Tech. Rep., Statist. Sect., Imperial College London, London, U.K., 2006.

[18] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol. 19, pp. 553–572, 2003.

[19] M. Nielsen, "On the construction and frequency localisation of finite orthogonal quadrature filters," *J. Approx. Theory*, vol. 108, pp. 36–52, 2001.

[20] R. A. Gopinath, "Phaselets of framelets," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1794–1806, May 2005.

[21] T. R. Downie and B. W. Silverman, "The discrete multiple wavelet transform and threshold methods," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2558–2561, Sep. 1998.

[22] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc. B*, vol. 59, pp. 319–351, 1997.

[23] R. L. Dykstra, "Product inequalities involving the multivariate normal distribution," *J. Amer. Stat. Assoc.*, vol. 75, pp. 646–650, 1980.

[24] T. S. Ferguson, *A Course in Large Sample Theory*. London, U.K.: Chapman & Hall/CRC, 1996.

[25] A. Grad and A. Solomon, "Distribution of quadratic forms and some applications," *Ann. Math. Statist.*, vol. 26, pp. 464–477, 1955.

[26] J. S. Marron and S. Adak, "Exact risk analysis of wavelet regression," *J. Comput. Graph. Statist.*, vol. 7, pp. 278–309, 1998.

[27] S. C. Olhede, "Hyperanalytic denoising," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2006, vol. 1, pp. 1421–1424.

[28] I. I. Gikhman and A. V. Skorokhod, *The Theory of Stochastic Processes*. London, U.K.: Springer, 1980.

Sofia C. Olhede received the M.Sci. degree in mathematics from Imperial College London, London, U.K., in August 2000, graduating with the Sir John William Lubbock Prize, awarded to the most meritorious candidate obtaining First Class Honours in Mathematics throughout the University of London. She was awarded the Beit Scientific Research Fellowship in October 2001 and received the Ph.D. degree in mathematics from Imperial College London in January 2003.

She joined the staff at the Department of Mathematics, Imperial College London, in October 2002, and was promoted to Senior Lecturer of statistics in October of 2006. She currently serves on the research committee of the Royal Statistical Society. Her research interests are nonstationary time series and random field analysis.