The Estimation of Laplace Random Vectors in Additive White Gaussian Noise

Ivan W. Selesnick, *Senior Member, IEEE*

***Abstract*—This paper develops and compares the maximum *a posteriori* (MAP) and minimum mean-square error (MMSE) estimators for spherically contoured multivariate Laplace random vectors in additive white Gaussian noise. The MMSE estimator is expressed in closed-form using the generalized incomplete gamma function. We also find a computationally efficient yet accurate approximation for the MMSE estimator. In addition, this paper develops an expression for the MSE for any estimator of spherically contoured multivariate Laplace random vectors in additive white Gaussian noise (AWGN), the development of which again depends on the generalized incomplete gamma function. The estimators are motivated and tested on the problem of wavelet-based image denoising.**

***Index Terms*—Denoising, estimation, Laplace distribution.**

I. Introduction

THE zero-mean Laplace probability distribution function (pdf) is useful in several signal processing applications. It can be used in wavelet-based signal processing to model the distribution of wavelet coefficients [63] and in speech processing to model the distribution of discrete Fourier transform (DFT) coefficients of short speech segments [5], [22], [26], [40], [43]. Similarly, the zero-mean elliptically contoured multivariate Laplace pdf can be used to model *groups* (or 'blocks') of such coefficients. A zero-mean elliptically contoured Laplace pdf is a multivariate pdf for which all the univariate marginals are zero-mean Laplace. We recall that a zero-mean Laplace random variable u with variance σ^2 has the density

$$p_u(u) = \frac{1}{\sqrt{2}\,\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}\,|u|\right). \tag{1}$$

The zero-mean elliptically contoured Laplace pdf (or simply, multivariate Laplace pdf) is described in detail in [37]. In [37], it is derived, for example, that the multivariate Laplace pdf can be expressed in terms of a Bessel function. The estimation of the covariance matrix of a multivariate Laplace distributed vector, and its application to speech processing, is addressed in [22].

Random vectors distributed according to elliptically contoured densities are also known as spherically invariant random processes (SIRPs) [8], [67]. The use of SIRPs in detection problems for radar has been developed in, for example, [4], [17], [53].

Manuscript received March 29, 2007; revised November 20, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrice Abry. This work was supported in part by ONR under grant N00014-03-1-0217.

The author is with Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: selesi@poly.edu).

Digital Object Identifier 10.1109/TSP.2008.920488

In this paper we are interested in the problem of estimating a d-component Laplace vector, $\mathbf{s}$, in additive white Gaussian noise (AWGN), $\mathbf{n}$

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \quad \mathbf{y}, \mathbf{s}, \mathbf{n} \in \mathbb{R}^d. \tag{2}$$

The estimation of $\mathbf{s}$ is based on the observed vector $\mathbf{y}$. In the scalar case ($d = 1$), the maximum *a posteriori* (MAP) and minimum mean-square error (MMSE) estimators $\hat{s}(y)$ are already known. The MAP estimator is given by the soft-threshold rule [29], [32], [46], [47], [66], and the MMSE estimator can be expressed in terms of the erf function [26], [29], [43]. The estimators given below generalize these estimators to the vector case.

In Sections IV-A and B, we find the MAP and MMSE estimators when $\mathbf{s}$ is a spherically contoured Laplace random vector (when its components s_i are uncorrelated and have equal variance). We also provide computationally efficient approximations for the MAP and MMSE estimators, and in Section V we compare the estimators' performance in terms of MSE. We have reported some of these results in a conference paper [55]. For the more general case, where $\mathbf{s}$ has an elliptically contoured (nonspherical) distribution, we have not yet obtained a simple expression for the MMSE estimator; we report some preliminary results for that case in [56].

In the derivation of the MMSE estimator $\hat{\mathbf{s}}(\mathbf{y})$ given below, we use the *generalized incomplete Gamma function*, $\Gamma(\alpha, x; b)$, a special function introduced in 1994 by Chaudhry and Zubair [12], [13]. (Their motivation for defining and studying this function was the role it plays in the closed-form solution to several problems in heat conduction.) We need the generalized incomplete Gamma function to write also the observational density $p_{\mathbf{y}}(\mathbf{y})$. For the general case, where $\mathbf{s}$ has an elliptically contoured Laplace distribution, the function $\Gamma(\alpha, x; b)$ can no longer be used [56]. For that reason, our results below are limited to the spherically contoured case.

We are motivated to study this estimation problem because of our interest in wavelet-based image denoising, or more generally transform-domain denoising [48]. Statistical modeling of images in the wavelet domain calls for peaked, heavy-tailed symmetric densities; and the problem of fitting suitable models (especially from noisy data) has several proposed solutions in the literature [54]. A simple model is the Laplace distribution. However, the distribution of wavelet coefficients of images can usually be modeled more accurately using densities with two or more parameters; examples include: the generalized Gaussian [10], [42], [46], [63], Gaussian mixtures [14], [16], [18], scale mixtures models [52], mixtures of a point mass and Laplace or other density [35], Bessel K densities [24], symmetric alpha-stable densities [1], [2], [6], [65], and the

normal inverse Gaussian density [34]. Methods for handling a variety of densities have also been developed [3], [20], [31]. In addition, we also note that some nonlinear processing rules can be derived without using a Bayesian approach or without explicitly specifying a prior distribution [21], [25], [39], [41].

While accurately modeling the univariate marginal distribution of coefficients is useful, it is usually more important to model groups of coefficients together [38]. If adjacent wavelet coefficients (in location and/or scale) are modeled together, then substantial improvement can be realized in image denoising. Indeed, even though adjacent wavelet coefficients are approximately uncorrelated (assuming the transform is orthonormal) they are strongly dependent [9], [30], [62], [64]. Several methods have been developed to model groups (or "blocks") of wavelet coefficients, for example, the local variance, context, or related concepts can be used [11], [18], [23], [33], [36], [45], [50], [51], [61]. Some wavelet-domain denoising algorithms are based, either implicitly or explicitly, on multivariate elliptically contoured probability densities [15], [27], [44], [57], [58].

In this paper we develop the MAP and MMSE nonlinear threshold/shrinkage rules that follow from the spherically contoured multivariate Laplace density. Although it is only a one-parameter distribution and therefore limited in its effectiveness to accurately fit the univariate marginal distribution of coefficients, the multivariate Laplace pdf does at least model the dependencies among adjacent coefficients while being consistent with the basic and frequently used univariate Laplace pdf. Moreover, to the best our knowledge, the basic problem of estimating a Laplace vector in independent white Gaussian noise does not appear to be previously addressed.

In our previous work, we developed in [57] and [59] a multivariate spherically contoured radially exponential density that is similar to the Laplace density in its functional form, but the marginals of the density in [57] and [59] are not Laplace. Although the multivariate density in [57] and [59] specializes to the Laplace pdf in the scalar case ($d = 1$), for $d > 1$ it is less kurtotic (more like a Gaussian pdf). In Section III-B we compare the univariate marginals of that distribution with the true multivariate Laplace density to show the differences between them. It is useful to employ the true multivariate Laplace density because then Laplace models (or Laplace mixture models [49]) for the univariate marginal can be directly extrapolated to obtain multivariate probability models for groups of coefficients.

A. Empirical Histograms

Fig. 1 illustrates the histogram of the coefficients in one subband of the wavelet transform of a photographic image. The histogram conforms to the well-known behavior of such histograms — namely, compared to the Gaussian pdf, the histogram has a different behavior both at the center (it is more peaked) and in the tails (they are heavier). Wavelet coefficient histograms are more kurtotic than Gaussian distributions.

Fig. 1 also illustrates two scatter plots. Both scatter plots were generated using the same set of wavelet coefficients used to to generate the histogram in Fig. 1. Scatter plot A illustrates the joint distribution of two wavelet coefficients that are horizontally adjacent $(w(n_1, n_2), w(n_1, n_2 + 1))$. Scatter plot B illustrates the joint distribution of two horizontally distant coefficients $(w(n_1, n_2), w(n_1, n_2 + 10))$. The marginal histogram of both scatter plots are the same and is given by the histogram in Fig. 1 (because both scatter plots are generated from the same data). Scatter plot A suggests that the joint distribution of adjacent coefficients is approximately elliptically contoured. Also a slight positive correlation is visible — the contours of

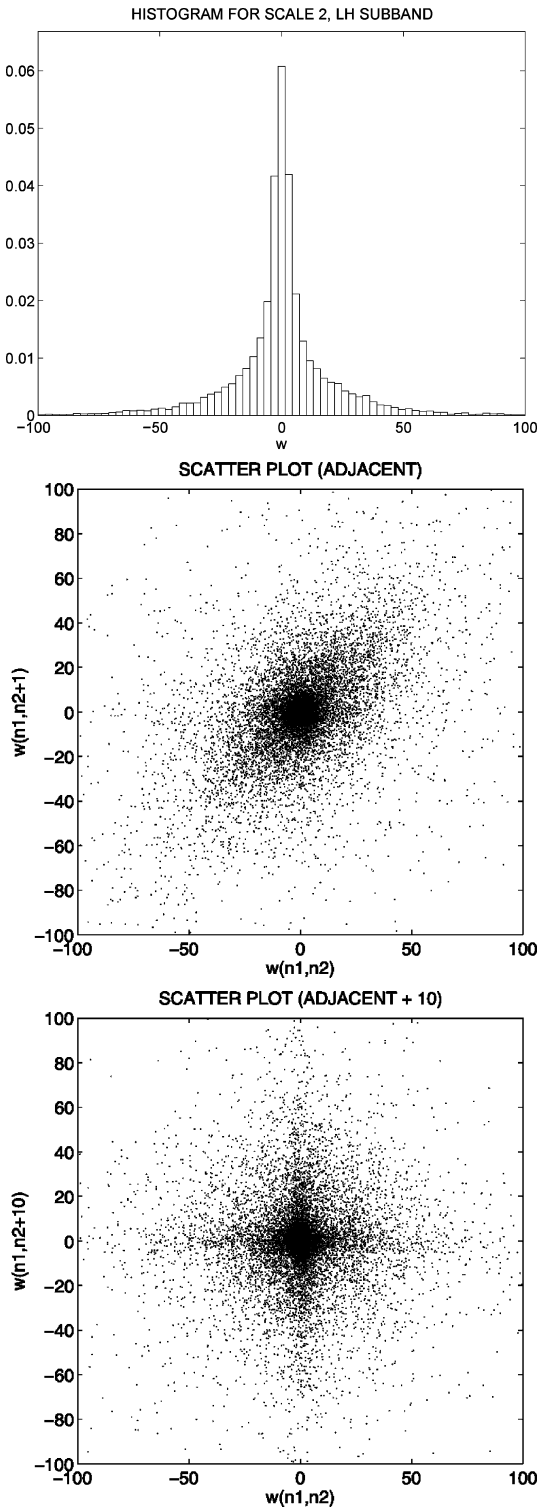

Fig. 1. Histogram and scatter plots of wavelet coefficients in the LH subband at scale-2 of the 512×512 pixel *boat* image. Each plot represents 256^2 values. The scatter plots illustrate pairs $[w(n_1, n_2), w(n_1, n_2 + i)]$ for $i = 1, 10$.

the density function will be elliptic, not circular (however, this paper will examine only spherically contoured density functions). (Note that if scatter plot A were displayed as a bivariate conditional histogram where each column of this histogram is normalized by its sum, then we will see the "bow-tie" shape as discussed in earlier work [62]). Scatter plot B shows that the joint distribution of distant coefficients is not elliptically contoured, but roughly start-shaped. The distribution of points in scatter plot B is concentrated along the axes — that is because 1) distant coefficients are essentially independent so the joint pdf is therefore given by the product, $p(w_1, w_2) = p(w_1)\,p(w_2)$, and 2) the marginal distribution is peaked at the center. Indeed, the star-shaped distribution illustrated in scatter plot B is expected of joint pdfs of independent random variables where each follows a distribution that is peaked at the origin.

Both scatter plots in Fig. 1 have a high concentration of points at the origin because a high proportion of wavelet coefficients are small in absolute value. The scatter plots in Fig. 1 confirm that an elliptically contoured pdf is a reasonable model for pairs of adjacent coefficients, and also that such a model is not valid for pairs of coefficients that are distant from one another.

II. Generalized Incomplete Gamma Function

In 1994, Chaudhry and Zubair introduced the the *generalized incomplete gamma function* [12], [13], defined as

$$\Gamma(\alpha,\, x;\, b) := \int_x^\infty t^{\alpha-1}\, \exp\left(-t - \frac{b}{t}\right)\, dt. \tag{3}$$

When $b = 0$, the incomplete Gamma function is obtained as a special case, $\Gamma(\alpha, x) = \Gamma(\alpha,\, x;\, 0)$. For $\alpha = \mathbb{Z} + 1/2$ there is a closed form expression for the generalized incomplete gamma function. For $\alpha = 1/2$, for example, there is the formula [12], [13]

$$\Gamma(\frac{1}{2},\, x;\, b) = 0.5\sqrt{\pi}\, \left[\exp\left(-2\sqrt{b}\right)\operatorname{erfc}\left(\sqrt{x} - \sqrt{\frac{b}{x}}\right) + \exp\left(2\sqrt{b}\right)\operatorname{erfc}\left(\sqrt{x} + \sqrt{\frac{b}{x}}\right)\right] \tag{4}$$

where $\operatorname{erfc}(x) := 1 - \operatorname{erf}(x)$. For $\alpha = -1/2$ there is the formula

$$\Gamma\left(-\frac{1}{2},\, x;\, b\right) = 0.5\sqrt{\frac{\pi}{b}} \times \left[\exp\left(-2\sqrt{b}\right)\operatorname{erfc}\left(\sqrt{x} - \sqrt{\frac{b}{x}}\right) - \exp\left(2\sqrt{b}\right)\operatorname{erfc}\left(\sqrt{x} + \sqrt{\frac{b}{x}}\right)\right]. \tag{5}$$

The generalized incomplete Gamma function satisfies a recurrence relation that is useful for computing its values for other orders α, from [12], [13]

$$\Gamma(\alpha - 1, x; b) = \frac{1}{b}\left[\Gamma(\alpha + 1, x; b) - \alpha\,\Gamma(\alpha, x; b) - x^\alpha\, \exp\left(-x - \frac{b}{x}\right)\right]. \tag{6}$$

As stated in [7], the recursive formula together with (4) and (5) can be used to numerically compute $\Gamma(\alpha,\, x;\, b)$ for $\alpha = -1.5, -2.5, \ldots$.

For α not of the form $\mathbb{Z} + 1/2$, no closed form expression is available for $\Gamma(\alpha,\, x;\, b)$. However, it can be computed accurately and efficiently.[1]

III. Multivariate Densities

This section develops the multivariate densities required to obtain the sought MAP and MMSE estimators.

A. Laplace Random Vectors

A Gaussian scale mixture (GSM) representation of a d-component random vector having a spherically contoured density is

$$\mathbf{s} = \sqrt{z} \cdot \mathbf{x}, \qquad \mathbf{s}, \mathbf{x} \in \mathbb{R}^d, \quad z \in \mathbb{R} \text{ with } z \geq 0 \tag{7}$$

where $\mathbf{x}$ is a d-component zero-mean white Gaussian with variance σ^2 random vector, $\mathcal{N}(0, \sigma^2\, I_d)$

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\,\pi\,\sigma^2)^{d/2}}\, \exp\left(-\frac{\|\mathbf{x}\|^2}{2\,\sigma^2}\right). \tag{8}$$

Setting $a = \sqrt{z}$, then $\mathbf{s} = a\,\mathbf{x}$ and the pdf of the random vector $\mathbf{s}$ is given by

$$p_{\mathbf{s}}(\mathbf{s}) = \int_0^\infty p_a(a)\, \frac{1}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right)\, da. \tag{9}$$

Note, because $a = \sqrt{z}$, the density of a is

$$p_a(a) = 2\,a\,p_z(a^2). \tag{10}$$

The GSM representation will be used as an intermediate step in the derivation of several expressions below. Namely, it will be used to derive the density (26) of a Laplace random vector in AWGN in Section III-C, the MMSE estimator (41) in Section IV-B, and the MSE (45) in Section V. In the estimation rules described below it will not be necessary to estimate the value of the variable z in (7).

As described in [37], a spherically contoured d-component random vector with Laplace marginals (with variance σ^2) can be generated by (7) where z is a unit mean exponential (scalar) random variable, $p_z(z) = \exp(-z)$ for $z \geq 0$. Therefore, (10) gives

$$p_a(a) = 2\,a\, \exp(-a^2). \tag{11}$$

Using (8) and (11) in (9) gives

$$p_{\mathbf{s}}(\mathbf{s}) = \frac{1}{(2\,\pi\,\sigma^2)^{d/2}} \times \int_0^\infty a^{-d}\, \exp\left(-a^2 - \frac{1}{2\,a^2\,\sigma^2}\, \|\mathbf{s}\|^2\right)\, 2\,a\, da. \tag{12}$$

Changing the variable of integration, using $a = \sqrt{t}$, and converting to the modified Bessel function of the second kind,

[1] G. Chudnovsky, M. Leung. Private communication.

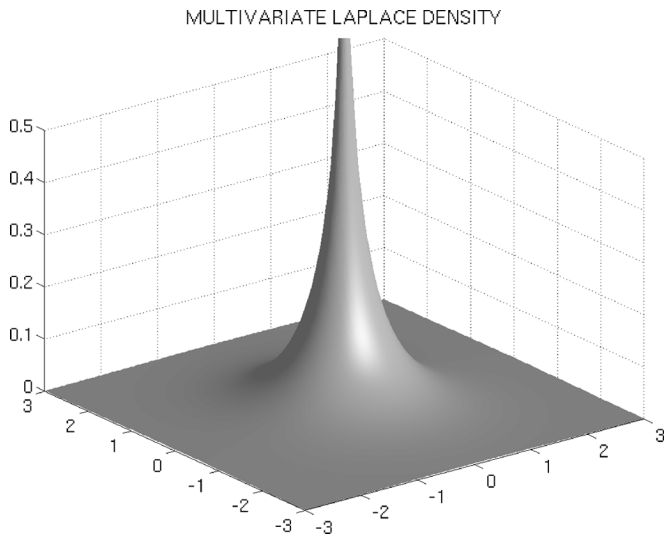

Fig. 2. Multivariate spherically contoured Laplace distribution (13) with $\sigma = 1$, $d = 2$.

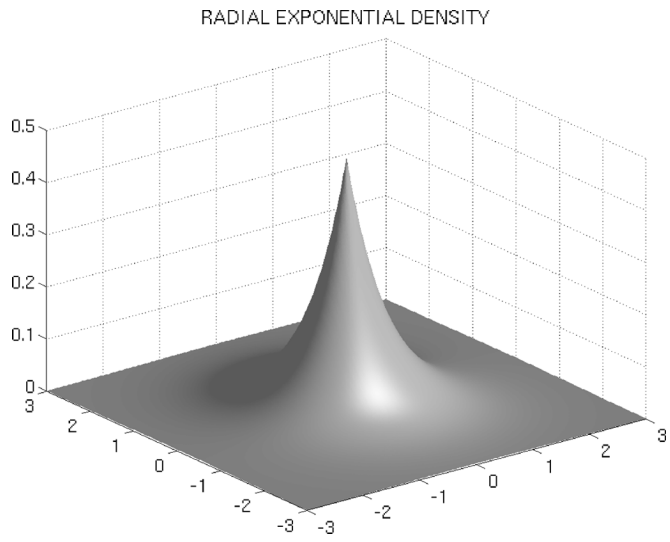

Fig. 3. Multivariate radial exponential distribution (14) with $\sigma = 1$, $d = 2$.

$K_\lambda(u)$ in (50), the density of the Laplace random vector is given by

$$p_{\mathbf{s}}(\mathbf{s}) = \frac{1}{\pi\,\sigma^2}\left(\frac{1}{\sqrt{2}\,\pi\,\sigma\;\|\mathbf{s}\|}\right)^{d/2-1} K_{d/2-1}\left(\frac{\sqrt{2}}{\sigma}\;\|\mathbf{s}\|\right). \tag{13}$$

The spherically contoured multivariate Laplace distribution is illustrated in Fig. 2 for $d = 2$. Note that the pdf (13) has a singularity at zero unless $d = 1$. For $d = 2$, the singularity is evident in Fig. 2. The singularity for $d > 1$ explains in part, why the MMSE and MAP estimators of Laplace vectors in AWGN, behave more differently from each other around zero for $d > 1$ than they do for $d = 1$, as will be illustrated below.

B. Radial Exponential Distribution

In our previous work [57], [59], [60], we developed MAP estimators based on the d-component spherically contoured multivariate density

$$p_{\mathbf{s}}(\mathbf{s}) = C\,\frac{1}{\sigma^d}\,\exp\left(-\frac{\sqrt{d+1}}{\sigma}\;\|\mathbf{s}\|\right), \qquad \mathbf{s}\in\mathbb{R}^d \tag{14}$$

where the normalization constant C is

$$C = \frac{\sqrt{\pi}}{\Gamma\left(\frac{d+1}{2}\right)}\left(\frac{d+1}{4\,\pi}\right)^{d/2}.$$

We call (14) the spherically contoured radial exponential distribution, see also [37, p. 270]. The radial profile of this spherically contoured density is exponential in $\|\mathbf{s}\|$. The constant in the exponent is chosen so that $\mathbb{E}[s_i^2] = \sigma^2$. The radial exponential distribution is illustrated in Fig. 3.

It turns out that the radial exponential density in (14) can be represented as a GSM as in (7) where the random variable z is gamma distributed

$$p_z(z) = \frac{\beta^\alpha}{\Gamma(\alpha)}\,z^{\alpha-1}\,\exp(-\beta\,z) \tag{15}$$

with $\alpha = \beta = 0.5\,(d+1)$ and $\mathbf{x}$ is $\mathcal{N}(0, \sigma^2\,I_d)$ as in (8). To confirm that this GSM yields the density (14), use these values of α and β in (15), with (8) and (10) in (9), and simplify using (51).

This GSM representation of (14) can be used to obtain the marginal density

$$p_{s_i}(s_i) = \frac{1}{\sqrt{\pi}\,\Gamma(\frac{d+1}{2})}\left(\frac{d+1}{\sigma^2}\right)^{d/4+1/2} \times\left|\frac{s_i}{2}\right|^{d/2} K_{d/2}\left(\frac{\sqrt{d+1}}{\sigma}\,|s_i|\right) \tag{16}$$

for $1 \le i \le d$.

The density (16) is a special case of the Bessel K form density [28, Eq. (6)] used for wavelet-based denoising in [24], that has the form

$$f(s;c,p) = \frac{1}{\sqrt{\pi}\,\Gamma(p)}\left(\frac{c}{2}\right)^{-p/2-1/4} \times\left|\frac{s}{2}\right|^{p-1/2} K_{p-1/2}\left(\sqrt{\frac{2}{c}}\,|s|\right). \tag{17}$$

A random variable v with density (17) is denoted $v \sim \mathrm{BKF}(p,c)$. If the d-component random vector $\mathbf{s}$ is distributed according to (14), then the marginal of s_i in (16) is

$$s_i \sim \mathrm{BKF}\left(\frac{d+1}{2}, \frac{2\,\sigma^2}{d+1}\right). \tag{18}$$

Fig. 4 illustrates the the density (16) for $d = 1$ and $d = 9$. For $d = 1$, the marginal is the Laplace density; however, as d increases, the marginal becomes more similar to the Gaussian density.

The kurtosis of a random variable is a combined measure of how heavily tailed the distribution is, and how peaked the distribution is around its mean. The kurtosis of a Gaussian random variable is 3. A distribution whose tails are heavier than the Gaussian and that is more peaked around its mean, has a kurtosis that is higher than that of the Gaussian. Such a random variable

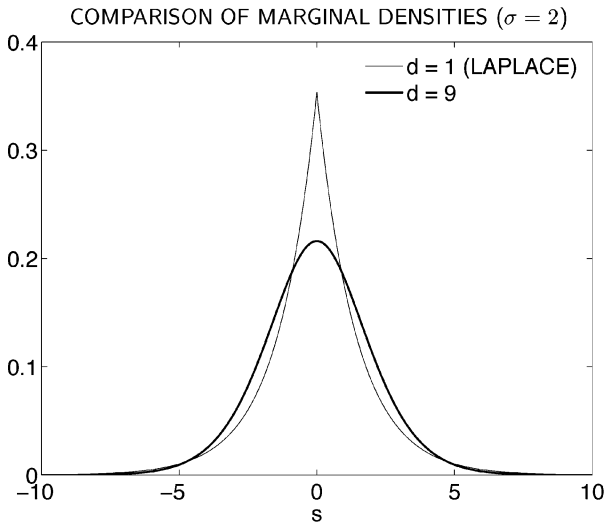

Fig. 4. Univariate marginal distribution of the d-component spherically contoured radial-exponential distribution (14) for $d = 1$ and $d = 9$. When $d = 1$ the distribution is the Laplace distribution. For $d > 1$ the marginal distribution is less kurtotic than the Laplace distribution. As d increases, the marginal distribution becomes more Gaussian.

is "more kurtotic" than the Gaussian. The kurtosis of a Laplace random variable is 6. From [24], the kurtosis of the random variable in (16), equivalently (18), is $= 3(d+3)/(d+1)$. As the number of components, d, increases, the kurtosis of s_i decreases to the kurtosis of the Gaussian. That means, that for higher d, the spherically contoured radial-exponential probability model is more Gaussian, and therefore, less appropriate for modeling the wavelet coefficients of natural images. This is why we are motivated in this paper to revisit the approach taken in [57] and [60], but to this time use the multivariate Laplace density (13) rather than (14). The multivariate Laplace density (13) does not behave that way — the marginal is Laplace and therefore the kurtosis of s_i is 6 for any d.

C. Laplace Random Vectors in AWGN

To develop algorithms for processing signals in noise, it is often necessary to know the pdf of the noisy signal. If a d-component signal $\mathbf{s}$ is modeled as a Laplace random vector (with variance σ^2 and uncorrelated components) and if the noise signal $\mathbf{n}$ is independent additive zero-mean white Gaussian noise (with variance σ_n^2), as in (2), then the pdf of $\mathbf{y}$ is given by the multivariate convolution

$$p_{\mathbf{y}}(\mathbf{y}) = (p_{\mathbf{s}} * p_{\mathbf{n}})(\mathbf{y}) = \int_{\mathbb{R}^d} p_{\mathbf{s}}(\mathbf{s})\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s} \qquad (19)$$

where the pdf of $\mathbf{s}$ is given by (8), (9), (11), and $\mathbf{n}$ is $\mathcal{N}(0, \sigma_n^2\, I_d)$

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{(2\,\pi\,\sigma_n^2)^{d/2}} \exp\left(-\frac{||\mathbf{n}||^2}{2\,\sigma_n^2}\right). \qquad (20)$$

Using (9), we write $p_{\mathbf{y}}(\mathbf{y})$ in (19) as

$$p_{\mathbf{y}}(\mathbf{y}) = \int_{\mathbb{R}^d} \left[\int_0^\infty p_a(a)\, \frac{1}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right) da\right] p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s} \qquad (21)$$

$$= \int_0^\infty p_a(a) \left[\int_{\mathbb{R}^d} \frac{1}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right) p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s}\right] da. \qquad (22)$$

The inside term is the d-dimensional convolution of two spherical multivariate Gaussian pdfs, one with variance $a^2\,\sigma^2$, $[\mathcal{N}(0, a^2\,\sigma^2\, I_d)]$, the other with variance σ_n^2, $[\mathcal{N}(0, \sigma_n^2\, I_d)]$. The result is a multivariate Gaussian pdf with variance $a^2\,\sigma^2 + \sigma_n^2$, $[\mathcal{N}(0, (a^2\,\sigma^2 + \sigma_n^2)\, I_d)]$, so the inner integral in (22) is given by

$$\int_{\mathbb{R}^d} \frac{1}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right) p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s} = \frac{1}{(2\,\pi)^{d/2}} \times \frac{1}{(a^2\,\sigma^2 + \sigma_n^2)^{d/2}} \exp\left(-\frac{||\mathbf{y}||^2}{2\,(a^2\,\sigma^2 + \sigma_n^2)}\right). \qquad (23)$$

Using (11) and (23) in (22) gives

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\,\pi)^{d/2}} \cdot \int_0^\infty \left[\frac{1}{(a^2\,\sigma^2 + \sigma_n^2)^{d/2}} \exp\left(-a^2 - \frac{||\mathbf{y}||^2}{2\,(a^2\,\sigma^2 + \sigma_n^2)}\right)\right] 2\,a\,da. \qquad (24)$$

Changing the variable of integration, using

$$t = a^2 + \frac{\sigma_n^2}{\sigma^2}, \qquad dt = 2\,a\,da \qquad (25)$$

gives

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\,\pi\,\sigma^2)^{d/2}} \exp\left(\frac{\sigma_n^2}{\sigma^2}\right) \times \int_{\sigma_n^2/\sigma^2}^\infty \left[\frac{1}{t^{d/2}} \exp\left(-t - \frac{\frac{||\mathbf{y}||^2}{2\,\sigma^2}}{t}\right)\right] dt.$$

Using the generalized incomplete gamma function, $\Gamma(\alpha,\, x;\, b)$ in (3), we get

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\,\pi\,\sigma^2)^{d/2}} \exp\left(\frac{\sigma_n^2}{\sigma^2}\right) \Gamma\left(1 - \frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2};\, \frac{||\mathbf{y}||^2}{2\,\sigma^2}\right). \qquad (26)$$

When $d = 1$ in (26) we recover the scalar case which has been given in [29] and [35].

For large $||\mathbf{y}||$ an approximation of (26) is

$$p_{\mathbf{y}}(\mathbf{y}) \approx \frac{1}{\sqrt{2}\,\sigma} \left(\frac{1}{\sqrt{2}\pi\sigma\,||\mathbf{y}||}\right)^{d/2-1/2} \exp\left(\frac{\sigma_n^2}{\sigma^2} - \frac{\sqrt{2}}{\sigma}\,||\mathbf{y}||\right).$$

IV. Estimators

In Section IV-A the MAP estimator of a spherically contoured Laplace random vector in AWGN is derived. The MAP estimator is simpler to derive than the MMSE estimator, to be derived in Section IV-B, because the derivation of the MAP estimator requires no integration. In addition, for the derivation of the MAP estimator it is unnecessary to know explicitly the density $p_{\mathbf{y}}(\mathbf{y})$ of the observed data. For these reasons, the generalized incomplete gamma function $\Gamma(\alpha,\, x;\, b)$ will not be used in the derivation of the MAP estimator, although it will be used in

the derivation of the MMSE estimator. However, it will be illustrated in Section V that for the vector case ($d > 1$) the MAP estimator is a poor substitute for the MMSE estimator, at least for the estimation of Laplace random vectors in AWGN.

A. MAP Estimator

The MAP estimate of the d-component vector $\mathbf{s} \in \mathbb{R}^d$ from the noisy observation $\mathbf{y} \in \mathbb{R}^d$ is given by

$$\hat{\mathbf{s}}(\mathbf{y}) = \arg\max_{\mathbf{s}} p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y}). \tag{27}$$

When the noise is additive and independent as in (2), the MAP estimator is given by

$$\hat{\mathbf{s}}(\mathbf{y}) = \arg\max_{s} \left[\log(p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})) + \log(p_{\mathbf{s}}(\mathbf{s}))\right]. \tag{28}$$

The pdf $p_{\mathbf{s}}(\mathbf{s})$ is given by (13); the pdf $p_{\mathbf{n}}(\mathbf{n})$ is given by (20).

Setting derivatives to zero in (28) so as to maximize over $\mathbf{s}$, gives

$$y_i = \hat{s}_i - \sigma_n^2 \frac{\partial}{\partial \hat{s}_i} \log p_{\mathbf{s}}(\hat{\mathbf{s}}) \qquad 1 \le i \le d. \tag{29}$$

From (52) we have

$$\frac{\partial}{\partial u} \log K_\lambda(u) = \frac{\lambda}{u} - \frac{K_{\lambda+1}(u)}{K_\lambda(u)}. \tag{30}$$

Using (30) with (13) gives

$$\frac{\partial}{\partial \hat{s}_i} \log p_{\mathbf{s}}(\hat{\mathbf{s}}) = -\frac{\sqrt{2}\,\hat{s}_i}{\sigma\,\|\hat{\mathbf{s}}\|} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}$$

therefore, from (29)

$$y_i = \hat{s}_i \left[1 + \frac{\sqrt{2}\,\sigma_n^2}{\sigma\,\|\hat{\mathbf{s}}\|} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}\right]. \tag{31}$$

Taking the square root of the sum of the squares over $1 \le i \le d$ gives

$$\|\mathbf{y}\| = \|\hat{\mathbf{s}}\| \left[1 + \frac{\sqrt{2}\,\sigma_n^2}{\sigma\,\|\hat{\mathbf{s}}\|} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}\right]$$

or

$$\|\hat{\mathbf{s}}\| = \left(\|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|}{\sigma}\right)}\right)_+ \tag{32}$$

where $(x)_+$ represents $\max(x, 0)$. This rule can be implemented by successive substitution. That is a simple iteration in which $\hat{\mathbf{s}}$ is initialized (for example to $\mathbf{y}$) and then the left-hand-side of (32) is repeatedly computed

$$\|\hat{\mathbf{s}}\|^{(0)} \leftarrow \|\mathbf{y}\|$$

$$\|\hat{\mathbf{s}}\|^{(k)} \leftarrow \left(\|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|^{(k-1)}}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\hat{\mathbf{s}}\|^{(k-1)}}{\sigma}\right)}\right)_+.$$

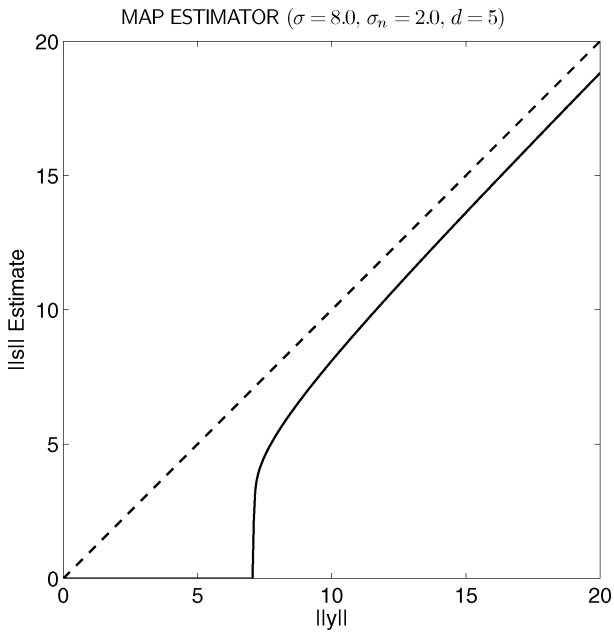

Fig. 5. Nonlinear function of the MAP estimator (32).

for $k \ge 1$. The convergence can be verified by noting that the successive estimates are nonincreasing and are lower bounded by zero. Once $\|\hat{\mathbf{s}}\|$ is found, the same attenuation ratio is applied to each component $\hat{s}_i$

$$\hat{s}_i = \frac{\|\hat{\mathbf{s}}\|}{\|\mathbf{y}\|}\, y_i, \qquad 1 \le i \le d. \tag{33}$$

Each component is attenuated by the same multiplier because the signal and noise distributions are spherically contoured. In the more general case where the distributions are elliptically contoured each component y_i would be attenuated by a different multiplier.

The shrinkage rule is illustrated in Fig. 5 for a 5-component vector ($d = 5$) and specific values of σ and σ_n. When $d = 1$, we can use the property $K_\lambda(u) = K_{-\lambda}(u)$ to recover from (32) the scalar soft-threshold rule with threshold $\sqrt{2}\,\sigma_n^2/\sigma$; that is, $|\hat{s}| = \left(|y| - \sqrt{2}\,\sigma_n^2/\sigma\right)_+$.

For large $\|\mathbf{y}\|$, using (54) the shrinkage rule (32) is approximately

$$\|\hat{\mathbf{s}}\| \approx \|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma}, \quad \|\mathbf{s}\| \gg 0 \tag{34}$$

which is the soft-threshold rule with the MAP threshold.

To reduce computation, it may be desirable to avoid the iterative successive substitution algorithm for the computation of the MAP estimator. In that case, a noniterative approximate form of (32) simply replaces $\|\hat{\mathbf{s}}\|$ on the right-hand-side of (32) by $\|\mathbf{y}\|$

$$\|\hat{\mathbf{s}}\| \approx \left(\|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma} \frac{K_{d/2}\left(\frac{\sqrt{2}\|\mathbf{y}\|}{\sigma}\right)}{K_{d/2-1}\left(\frac{\sqrt{2}\|\mathbf{y}\|}{\sigma}\right)}\right)_+. \tag{35}$$

The nonlinear threshold function (35) is illustrated in Fig. 6. A further computational simplification of (32) replaces the Bessel function ratio with an approximation. Using (54) in (35) gives

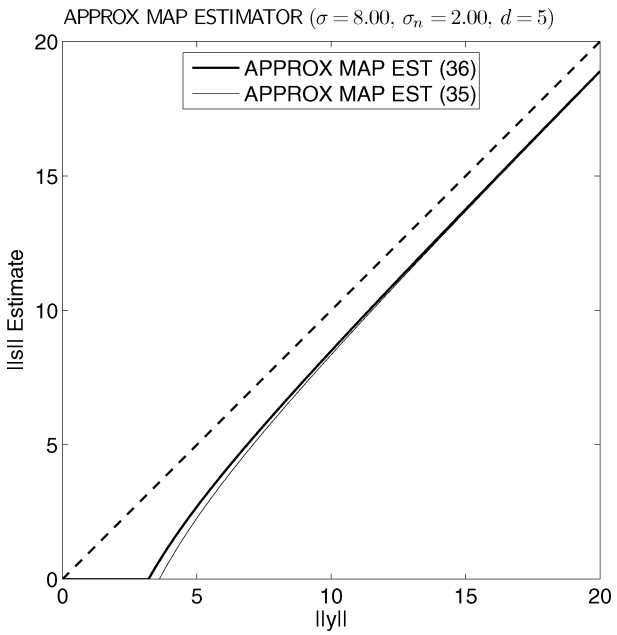

Fig. 6. Noniterative approximate MAP estimators (35) and (36).

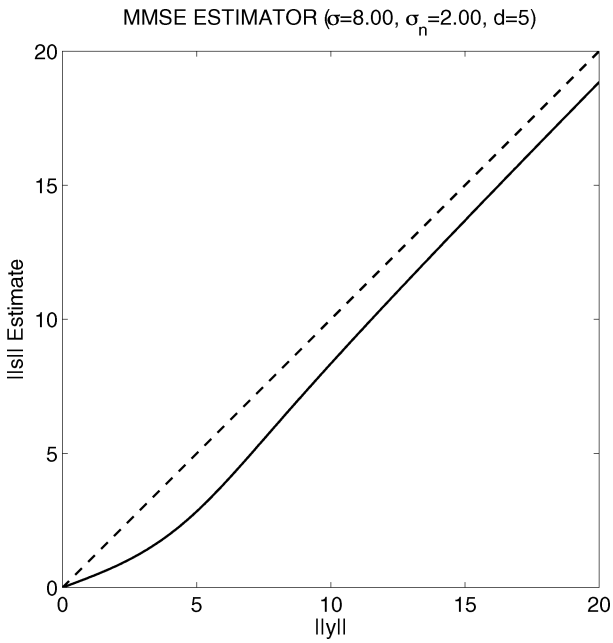

Fig. 7. Nonlinear shrinkage function of the MMSE estimator (41).

the computationally simple noniterative approximate MAP estimator

$$\|\hat{\mathbf{s}}\| \approx \left(\|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma} - \frac{d-1}{2}\,\frac{\sigma_n^2}{\|\mathbf{y}\|} \right)_+ . \tag{36}$$

When $\|\mathbf{y}\|$ is large, this will be approximately the soft-threshold rule because the last term will be small. Note that when $d = 1$, then (36) gives the scalar MAP soft-threshold rule, with a threshold of $\sqrt{2}\,\sigma_n^2/\sigma$. Fig. 6 illustrates the noniterative approximate MAP estimators (35) and (36). Using a more accurate approximation of the Bessel function ratio in (35) gives the estimator

$$\|\hat{\mathbf{s}}\| \approx \left(\|\mathbf{y}\| - \sqrt{2}\frac{\sigma_n^2}{\sigma} - \frac{d-1}{2}\frac{\sigma_n^2}{\|\mathbf{y}\|} - \frac{(d-1)(d-3)}{8\sqrt{2}}\frac{\sigma_n^2 \sigma}{\|\mathbf{y}\|^2} \right)_+ . \tag{37}$$

The threshold value associated with the MAP estimator (32) can not easily be expressed in closed-form. However, note that for the noniterative approximate MAP estimator (36) we can explicitly find the threshold value

$$T_{\mathrm{LAP}} = \frac{\sigma_n}{\sqrt{2}} \left(\frac{\sigma_n}{\sigma} + \sqrt{\frac{\sigma_n^2}{\sigma^2} + (d-1)} \right). \tag{38}$$

For $\|\mathbf{y}\|$ below this value, the estimated norm $\|\hat{\mathbf{s}}\|$ in (36) is zero. The threshold depends on the noise standard deviation σ_n, the ratio σ_n/σ, and the number of components d.

B. MMSE Estimator

The MMSE estimator of a Laplace scalar random variable in additive independent Gaussian noise has been derived in [29]. In this section we derive the MMSE estimator of a spherically contoured multivariate Laplace random vector in AWGN. The MMSE estimator of $\mathbf{s}$ given $\mathbf{y}$ is given by

$$\hat{s}_i = \int_{\mathbb{R}^d} s_i\, p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y})\, d\mathbf{s} = \frac{1}{p_{\mathbf{y}}(\mathbf{y})} \int_{\mathbb{R}^d} s_i\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})\, d\mathbf{s}. \tag{39}$$

The pdf $p_{\mathbf{y}}(\mathbf{y})$ is given by (26); the pdf $p_{\mathbf{s}}(\mathbf{s})$ is given by (9) and (11) [equivalently (13)]; the pdf $p_{\mathbf{n}}(\mathbf{n})$ is given by (20).

The MMSE estimator, derived in Appendix C, is given by

$$\hat{s}_i = y_i \left[1 - \frac{\sigma_n^2}{\sigma^2} \frac{\Gamma\left(\frac{-d}{2},\, \frac{\sigma_n^2}{\sigma^2};\, \frac{\|\mathbf{y}\|^2}{(2\sigma^2)}\right)}{\Gamma\left(1-\frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2};\, \frac{\|\mathbf{y}\|^2}{(2\sigma^2)}\right)} \right]. \tag{40}$$

Writing (40) in the same form as (32), (35), and (36) facilitates comparing the estimators to each other. Expressing the estimator as an attenuation of the norm, we have

$$\|\hat{\mathbf{s}}\| = \|\mathbf{y}\| \left[1 - \frac{\sigma_n^2}{\sigma^2} \frac{\Gamma\left(-\frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2};\, \frac{\|\mathbf{y}\|^2}{(2\sigma^2)}\right)}{\Gamma\left(1-\frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2};\, \frac{\|\mathbf{y}\|^2}{(2\sigma^2)}\right)} \right]. \tag{41}$$

When $d = 1$ in (41) we recover the scalar case which has been derived in [29], [35]. (We use (4) and (5) to confirm that (41) agrees with [29, p. 1787] in the scalar case.)

The MMSE estimator is illustrated in Fig. 7. Note that the nonlinearity of the MMSE estimator illustrated in Fig. 7 is quite different from the MAP estimator illustrated in Fig. 5.

As $\|\mathbf{y}\|$ goes to zero in (41), we get

$$\|\hat{\mathbf{s}}\| \approx \|\mathbf{y}\| \left[1 - \frac{\sigma_n^2}{\sigma^2} \frac{\Gamma\left(-\frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2}\right)}{\Gamma\left(1-\frac{d}{2},\, \frac{\sigma_n^2}{\sigma^2}\right)} \right], \qquad \mathbf{y} \approx 0$$

where $\Gamma(\alpha, x)$ is the ordinary incomplete Gamma function.

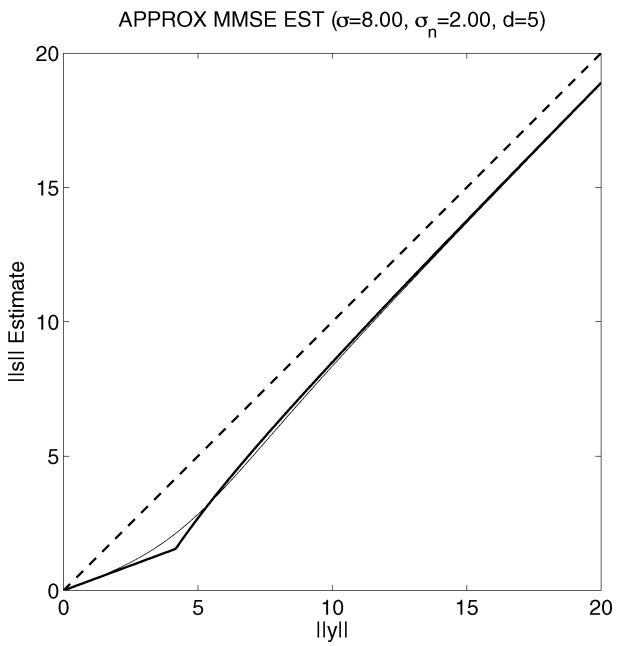

Fig. 8. Computationally efficient approximate MMSE estimator (42).

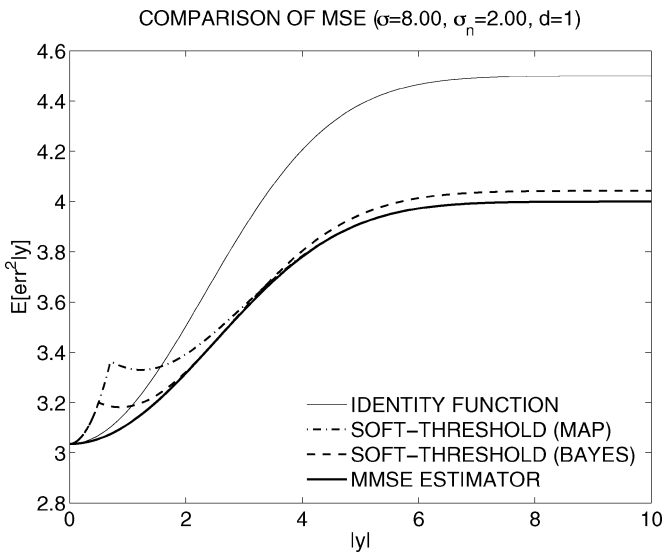

Fig. 9. MSE as a function of $\mathbf{y}$, $\mathbb{E}[e^2|y]$, for several estimators, with $d = 1$. For the soft-threshold function and for small $|y|$ the threshold $T = \sigma_n^2/\sigma$ gives the lower MSE, while for larger $|y|$ the threshold $T = \sqrt{2}\,\sigma_n^2/\sigma$ gives the lower MSE. The soft-threshold function with threshold $T = \sqrt{2}\,\sigma_n^2/\sigma$ gives the same MSE as the MMSE estimator for large $|y|$.

For a real-time image processing algorithm that calls the nonlinear MMSE estimator in (41) millions of times for a single image, an efficient implementation is desirable. One way to create a computationally simple approximation to this MMSE estimator is to approximate it for small $\mathbf{y}$ as a linear function the slope of which matches the slope of the exact MMSE estimator at $\mathbf{y} = 0$; and to use the noniterative approximate MAP estimator (36) for large $\mathbf{y}$. That gives,

$$\|\hat{\mathbf{s}}\| \approx \max\left(\|\mathbf{y}\|\left(1 - \frac{\sigma_n^2}{\sigma^2}\frac{\Gamma\left(-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2}\right)}{\Gamma\left(1-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2}\right)}\right),\right.$$
$$\left.\left(\|\mathbf{y}\| - \sqrt{2}\,\frac{\sigma_n^2}{\sigma} - \frac{d-1}{2}\,\frac{\sigma_n^2}{\|\mathbf{y}\|}\right)_+\right). \quad (42)$$

This approximation does not require the computation of the generalized incomplete Gamma function, and is therefore more computationally efficient. Fig. 8 illustrates that the computationally simple approximation (42) to the exact MMSE nonlinearity is reasonably good. For a slightly more accurate approximation we can use (37) in place of (36) in (42).

V. MSE

To compare the performance of several nonlinear estimation rules we can analyze the MSE. Let us examine the MSE as a function of the observed vector $\mathbf{y}$. The signal $\mathbf{s}$ is observed in noise $\mathbf{n}$, as in (2), where $\mathbf{s}$ and $\mathbf{n}$ are independent with pdfs $p_{\mathbf{s}}(\mathbf{s})$ and $p_{\mathbf{n}}(\mathbf{n})$. The estimate of $\mathbf{s}$, denoted $\hat{\mathbf{s}}$, is obtained by some function of $\mathbf{y}$. That is $\hat{\mathbf{s}} = f(\mathbf{y})$. The error of the estimate is defined as

$$\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}} = \mathbf{s} - f(\mathbf{y}). \quad (43)$$

We can find the average value of the error e_i^2 as a function of the observed noisy data $\mathbf{y}$. Namely, an expression for $\mathbb{E}[e_i^2\,|\,\mathbf{y}]$ is given by

$$\mathbb{E}[e_i^2\,|\,y] = \mathbb{E}[(s_i - \hat{s}_i)^2\,|\,\mathbf{y}] = \mathbb{E}[s_i^2\,|\,\mathbf{y}] - 2\,\hat{s}_i\,\mathbb{E}[s_i\,|\,\mathbf{y}] + \hat{s}_i^2. \quad (44)$$

The formula (44) requires $\mathbb{E}[s_i^2\,|\,\mathbf{y}]$ and $\mathbb{E}[s_i\,|\,\mathbf{y}]$. Note that $\mathbb{E}[s_i\,|\,\mathbf{y}]$ is the MMSE estimate of s_i, and therefore, a formula for $\mathbb{E}[s_i\,|\,\mathbf{y}]$ is given already by (40). A formula for $\mathbb{E}[s_i^2\,|\,\mathbf{y}]$, derived in Appendix D, is given by

$$\mathbb{E}[s_i^2\,|\,\mathbf{y}] = (y_i^2 + \sigma_n^2)$$
$$- (2\,y_i^2 + \sigma_n^2)\,\frac{\sigma_n^2}{\sigma^2}\,\frac{\Gamma\left(-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2};\,\frac{\|\mathbf{y}\|^2}{(2\,\sigma^2)}\right)}{\Gamma\left(1-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2};\,\frac{\|\mathbf{y}\|^2}{(2\,\sigma^2)}\right)}$$
$$+ y_i^2\,\frac{\sigma_n^4}{\sigma^4}\,\frac{\Gamma\left(-1-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2};\,\frac{\|\mathbf{y}\|^2}{(2\,\sigma^2)}\right)}{\Gamma\left(1-\frac{d}{2},\,\frac{\sigma_n^2}{\sigma^2};\,\frac{\|\mathbf{y}\|^2}{(2\,\sigma^2)}\right)}. \quad (45)$$

The explicit formula (45) for $\mathbb{E}[e_i^2|\mathbf{y}]$ can be used to compare various estimators, at least in terms of the MSE that results from their use. We start with a comparison of estimators for the scalar case, $d = 1$. In this case we can refer to e rather than to its i-th component, as e is a scalar. Fig. 9 illustrates the MSE as a function of y (namely $\mathbb{E}[e^2|y]$) in the scalar case for different estimation functions $f(y)$. First, Fig. 9 illustrates $\mathbb{E}[e^2|y]$ for the identity function $f(y) = y$. In this case, the observed noisy signal value y is unprocessed. We expect that this "estimator" provides the worst performance among the various estimators; indeed it does for large $|y|$. Next, Fig. 9 illustrates $\mathbb{E}[e^2|y]$ for $f(y)$ being the soft-threshold rule; two different thresholds T are used to produce two graphs. With the threshold value, $T = \sqrt{2}\,\sigma_n^2/\sigma_s$, the soft-threshold rule is the MAP estimator. The threshold value, $T = \sigma_n^2/\sigma_s$, is suggested in [29]. Note from the figure, that for small $|y|$, the soft-threshold

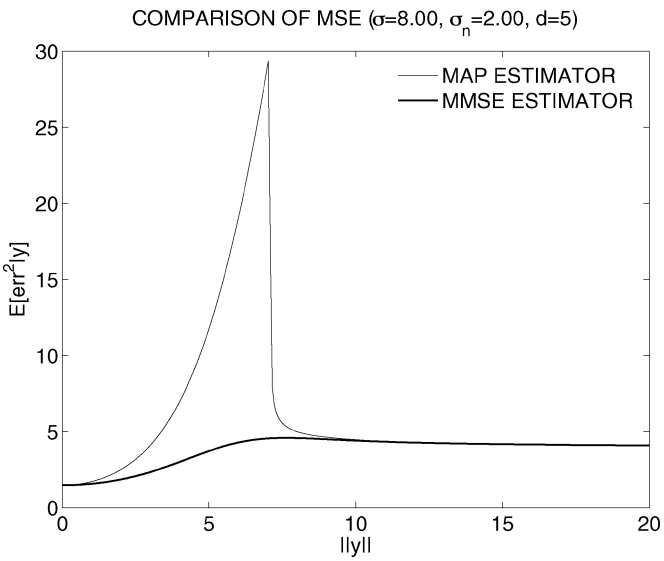

Fig. 10. MSE as a function of $\mathbf{y}$, $\mathbb{E}[e_i^2|\mathbf{y}]$, for the MAP and MMSE estimators, with $d = 5$. The MAP estimator gives a substantially higher MSE than the MMSE estimator for a range of $||\mathbf{y}||$. (For $d = 1$ the MAP estimator, namely the soft-threshold function, does not deviate so far from the MMSE estimator, see Fig. 9.)

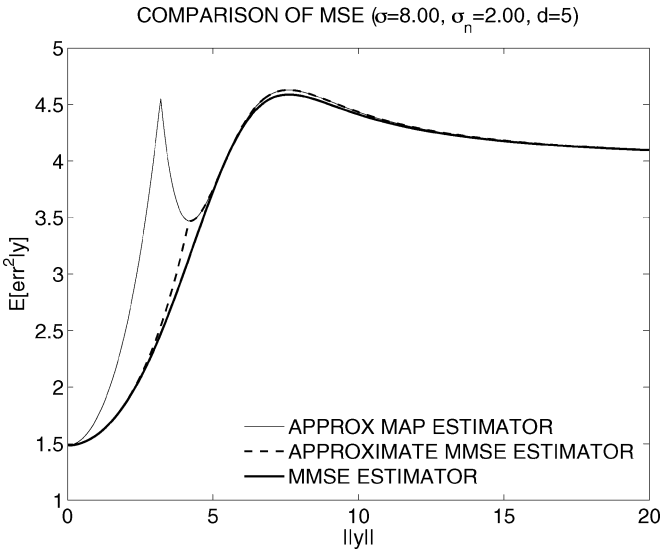

Fig. 11. $\mathbb{E}[e_i^2|\mathbf{y}]$ for several estimators, with $d = 5$. The approximate MAP estimator (36) provides a lower MSE than the exact MAP estimator illustrated in Fig. 10. The approximate MMSE estimator (42) closely follows the exact MMSE estimator.

rule performs worse than doing no processing at all (with either threshold value). Therefore, for small $|y|$, it is better to do no processing than to do soft-thresholding. For the smaller of the two thresholds values, the soft-threshold rule has the better performance for small $|y|$ but has the inferior performance for large $|y|$. Finally, Fig. 9 illustrates $\mathbb{E}[e^2|y]$ for $f(y)$ being the MMSE estimator, which, by definition, will have the lowest $\mathbb{E}[e^2|y]$ among all estimators. Note that the soft-threshold rule with the MAP threshold agrees with the MMSE estimator for large $|y|$. For the values of σ and σ_n used here, the MMSE estimator reduces the MSE of the noisy signal from 4.5 down to 4.0 for large $|y|$.

Despite its worse performance for small $|y|$, an advantage of the soft-threshold rule in comparison with the MMSE estimator is the computational simplicity of the soft-threshold rule. However, the approximate MMSE estimator (42) is also computationally simple (the nonlinear rule (42) requires no evaluation of the generalized incomplete Gamma function). When $\mathbb{E}[e^2|y]$ is graphed for the approximate and exact MMSE estimators, the graphs are almost indistinguishable. Even though it is very simple, the performance of the approximate MMSE estimator (42) closely follows that of the exact MMSE estimator (41). Therefore, the approximate MMSE estimator (42) is an effective computationally simple substitute for the exact MMSE estimator.

We now use the explicit formula for $\mathbb{E}[e_i^2|\mathbf{y}]$ to compare several estimators for the multivariate case, $d > 1$. Fig. 10 illustrates $\mathbb{E}[e_i^2|\mathbf{y}]$ with $d = 5$ for two estimation rules: the MAP and MMSE estimators. (The MAP estimator (32) is computed by successive substitution.) The figure reveals that the performance of the MAP estimator is very poor (in terms of MSE) for a range of $||\mathbf{y}||$. For some $||\mathbf{y}||$ the MSE of the MAP estimator is five times that of the MMSE estimator. Of course, the MAP estimator is not designed to minimize the MSE; however, it is simpler to derive and compute so it is sometimes used as a substitute for the MMSE estimator. If the MAP estimator is similar to the MMSE estimator (as it is in the case $d = 1$, see Fig. 9), then the use of the MAP estimator does not significantly increase the MSE. However, for $d > 1$ the MAP estimator has a significantly different shape than the MMSE estimator for small observation values. Specifically, the "dead-zone" of the MAP estimator (see Fig. 5) is especially large, due in part to the singularity of the multivariate Laplace pdf at the origin. Therefore, for $d > 1$ the MAP estimator is a poor substitute for the MMSE estimator (much more so than in the scalar case). For large $||\mathbf{y}||$ that problem is absent — the performance of the MAP estimator converges to that of the MMSE, as illustrated in Fig. 10.

Fig. 11 illustrates $\mathbb{E}[e_i^2|\mathbf{y}]$ with $d = 5$ for three estimation rules: the noniterative approximate MAP estimator (36), the approximate MMSE estimator (42), and the exact MMSE estimator (41). First, note that the noniterative approximate MAP estimator improves upon the exact MAP estimator (compare to Fig. 10) in terms of MSE. In this case, the computationally simpler of the two nonlinearities performs better. Next, note that the approximate MMSE estimator (42) provides further improvement — the MSE generated by it follows that of the exact MMSE estimator quite closely. Also notice that for $||\mathbf{y}|| > 5$ the performance of the approximate MMSE estimator (42) is exactly that of the noniterative approximate MAP estimator (36), because the latter is used in the definition of the former. In conclusion, the approximate MMSE estimator (42) is a computationally simple, yet reliable, approximation of the exact MMSE estimator (41).

VI. Comparison of Estimators

In this section we compare the functional form of the estimators derived in this paper with two other estimators that also work on groups of coefficients.

A. MAP Estimator for Radial Exponential Vector in AWGN

For a spherically contoured radial-exponential random vector (14) in AWGN, the MAP estimator, derived in [60], is given by

$$\|\hat{\mathbf{s}}\| = \left(\|\mathbf{y}\| - \sqrt{d+1}\,\frac{\sigma_n^2}{\sigma} \right)_+ \tag{46}$$

which we will call REXMAP (for 'Radial EXponential'). This is a soft-threshold rule applied to the vector norm. Let us compare the MAP estimate (46) derived from the multivariate radial-exponential model with the MAP estimate derived from multivariate Laplace model. For a large observation $\|\mathbf{y}\|$ under the Laplace model, from (34) we find that the estimator subtracts $\sqrt{2}\,\sigma_n^2/\sigma$ from $\|\mathbf{y}\|$ to obtain $\|\hat{\mathbf{s}}\|$, while under the radial-exponential model, the estimator subtracts $\sqrt{d+1}\cdot\sigma_n^2/\sigma$, a larger value (for $d > 1$). Therefore, the radial-exponential model leads to more attenuation of large observations than does the Laplace model.

We can also make a comparison for small observation $\|\mathbf{y}\|$ by comparing the threshold values. For the Laplace model, we will use the approximate MAP estimator (36). Comparing the threshold of the nonlinearity (46), $T_{\mathrm{EXP}} = \sqrt{d+1}\cdot\sigma_n^2/\sigma$ with T_{LAP} in (38) we find that $T_{\mathrm{LAP}} = T_{\mathrm{EXP}}$ when $d = 1$ (that is expected because for $d = 1$ the two models coincide). For small σ_n/σ, the approximate MAP estimator under the Laplace model has a greater threshold than the MAP estimator under the radial-exponential model; while for large σ_n/σ the reverse is true.

Unlike the Laplace case, there does not appear to be any simple form for the MMSE estimate of a radial-exponential random vector in AWGN. An approximate MMSE estimator is presented in [60], however, the difference between the obtained approximate MMSE estimator and the MAP estimator (46) was found to be neglible.

B. LAWMAP Algorithm

The nonlinear processing rules derived in this paper can be compared with the LAWMAP processing rule introduced in [45]. The processing rules developed in this paper are based on a stationary nonGaussian multivariate model, while on the other hand, the LAWMAP processing rule developed in [45] is based on a locally adaptive doubly stochastic model. It turns out that both approaches ultimately process the noisy observed data in similar ways.

In the LAWMAP algorithm, a group ('or block') of d noise-free signal values $\mathbf{s}$ is modeled as independent zero-mean Gaussian random variables conditioned on the *local variance*, σ^2. The estimate of each component $\hat{s}_i$ is then given by the Wiener estimate

$$\hat{s}_i = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \sigma_n^2}\, y_i, \qquad 1 \le i \le d \tag{47}$$

where $\hat{\sigma}^2$ is an estimate of the local variance estimated from the group of d coefficients $\mathbf{y}$.

The LAWMAP algorithm takes into account statistical dependencies among neighboring coefficients by modeling the local variance of each coefficient as a random variable. Specifically, the local variance (the variance in a 3×3 square window for example) is modeled as having an exponential distribution with mean $1/\lambda$. The first step of the LAWMAP algorithm is the determination of λ. The second step of the LAWMAP algorithm is the estimation of the local variance $\hat{\sigma}^2$, an estimate that depends on $\|\mathbf{y}\|$. The estimate $\hat{\sigma}^2$ given in [45] is

$$\hat{\sigma}^2 = \left[\frac{d}{4\lambda} \left(\sqrt{1 + \frac{8\lambda\,\|\mathbf{y}\|^2}{d^2}} - 1 \right) - \sigma_n^2 \right]_+ . \tag{48}$$

The LAWMAP algorithm then comprises of (48) and (47). To compare the LAWMAP processing rule with the MAP and MMSE estimators developed in Sections IV-A and IV-B, we can write the LAWMAP rule concisely as

$$\|\hat{\mathbf{s}}\| = \left(\|\mathbf{y}\| - \frac{4\lambda\,\|\mathbf{y}\|\,\sigma_n^2}{\sqrt{d^2 + 8\lambda\,\|\mathbf{y}\|^2} - d)} \right)_+ . \tag{49}$$

This rule states how the norm of a group of signal values is modified if the rule were applied to each value in the coefficient group. In practice, only the center signal value is modified — specifically, the center value is multiplied by the ratio $\|\hat{\mathbf{s}}\| \,/\, \|\mathbf{y}\|$ (which will always be less than 1). Writing the LAWMAP estimation rule in the form (49) makes clear how to compare it with the estimation rules developed in Sections IV-A and B.

Fig. 12 illustrates the LAWMAP nonlinearity (49) for particular values of λ, d and σ_n. Fig. 12 also illustrates the noniterative approximate MAP estimation rule (36) derived from the stationary multivariate Laplace model. As evident in the figure, the nonlinear estimation rules of the two methods are similar. Therefore, even though one method is based on a nonstationary signal model, and the other method is based on a stationary signal model, the estimation rules derived in each case exhibit

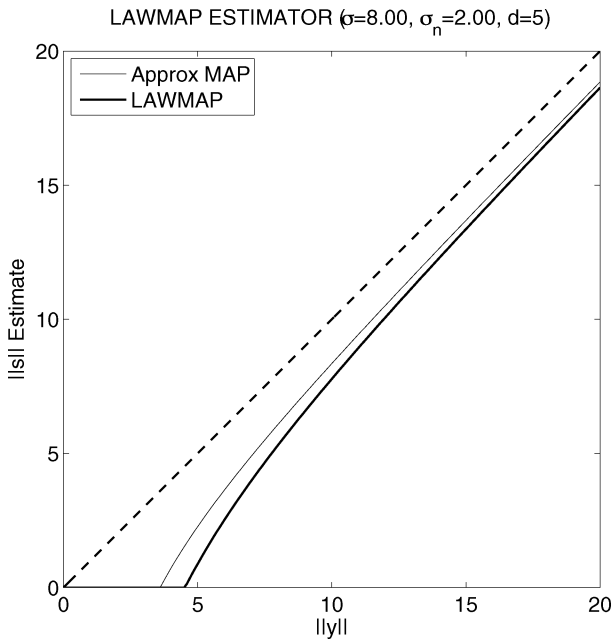

Fig. 12. Comparison of LAWMAP estimator (49) and the Laplace-based estimator (36).

TABLE I
RESULTS OF WAVELET-BASED IMAGE DENOISING USING THE APPROXIMATE MMSE (37), LAWMAP (49), AND REXMAP (46) ESTIMATORS

		MLAP-MMSE			LAWMAP		REXMAP	
Image	σ_n	1×1	3×3	5×5	3×3	5×5	3×3	5×5
Lena	10	33.51	34.35	34.36	34.26	34.32	33.50	33.34
	15	31.57	32.35	32.39	32.28	32.35	31.55	31.42
	20	30.28	30.97	31.02	30.87	30.97	30.29	30.14
	25	29.30	29.93	29.98	29.76	29.91	29.30	29.17
	30	28.51	29.08	29.14	28.84	29.04	28.50	28.37
Barbara	10	31.21	32.57	32.67	32.54	32.64	31.59	31.50
	15	28.81	30.06	30.30	30.11	30.26	29.25	29.19
	20	27.27	28.40	28.66	28.47	28.66	27.90	27.89
	25	26.13	27.23	27.45	27.24	27.47	26.69	26.67
	30	25.24	26.31	26.52	26.27	26.53	25.52	25.48
Boat	10	32.16	32.77	32.65	32.66	32.66	32.19	32.10
	15	30.03	30.70	30.61	30.58	30.60	30.09	29.92
	20	28.59	29.27	29.22	29.14	29.18	28.48	28.26
	25	27.55	28.15	28.15	28.03	28.10	27.35	27.18
	30	26.75	27.27	27.29	27.14	27.24	26.55	26.43

similar behavior. However, note that the LAWMAP estimator is different than the MMSE estimator (41) around the origin. In particular, the MMSE estimator is not a threshold function (it does not have a dead-zone) like the LAWMAP estimator.

This section has illustrated that the type of nonlinear processing that can be derived from a nonstationary Gaussian model can also be obtained from a stationary nonGaussian model. Thus, we can avoid the need of a nonstationary model, provided that a suitable nonGaussian model can be utilized (for which noisy data fitting and estimator derivation are both tractable). This provides an example that it is not always necessary to use a nonstationary signal model in order to obtain context-based or locally adaptive processing rules.

VII. RESULTS

In this section, we use the estimators described in this paper for wavelet-based reduction of additive white Gaussian noise in images. In this case, the nonlinear estimation rule is applied to every neighborhood in the wavelet domain, the neighborhoods being square, for example 3×3. In the scheme implemented here, the resulting shrinkage is applied to estimate only the middle coefficient of the neighborhood, and the noise variance σ_n^2 is assumed to be known. The signal variance σ^2 is estimated for each subband of the wavelet transform by subtracting the noise variance from the subband sample variance

$$\hat{\sigma} = \sqrt{(\sigma_w^2 - \sigma_n^2)_+}$$

where σ_w^2 is the mean of the squares of the noisy wavelet coefficients in the subband. Table I provides the PSNR results for several images, noise levels, and neighborhood sizes, using the noniterative approximate MMSE estimator (42) with (37), denoted by MLAP-MMSE in the table. The table illustrates the well-recognized gain achieved by processing coefficients as a group rather than individually (using a 1×1 neighborhood). Table I also tabulates the PSNR values for the LAWMAP algorithm [45] for comparison, with which the proposed estimator performs comparably. Finally, Table I also provides results for the REXMAP rule (46) which gives inferior performance. There are several other wavelet-based image denoising algorithms to which comparisons can be made; however, our intention is the comparison with a similarly simple one-pass algorithm designed to exploit intrascale dependencies. The LAWMAP algorithm is similar to the new estimator in that it estimates each coefficient using the coefficient's neighborhood. The estimators derived in this paper also accomplish that but are derived using a stationary model rather than a nonstationary one.

In our implementation of the MLAP-MMSE, LAWMAP, and REXMAP denoising algorithms, we used a length-10 Daubechies symlet filter to implement a 3-level orthonormal discrete wavelet transform with periodic boundary extensions [19]. The PSNR values in Table I are obtained by averaging over five realizations. The images are standard 8-bit test images of size 512×512.

Because the Laplace density is only a one-parameter model it can fit the histogram of the wavelet coefficients only approximately. Thefore, we expect it to have limited effectiveness. However, its performance in the image denoising test here does surpass that of the one-parameter radial exponential density, which supports the use of the multivariate Laplace density over that one.

VIII. CONCLUSION

In this paper we have developed and compared the MAP and MMSE estimators for spherically contoured multivariate Laplace random vectors in additive white Gaussian noise. The MMSE estimator is expressed in closed-form using the generalized incomplete gamma function, $\Gamma(\alpha, x; b)$. We also find a computationally efficient yet accurate approximation for the MMSE estimator. In addition, we develop an expression for the MSE for any estimator of spherically contoured multivariate Laplace random vectors in AWGN, which again depends on $\Gamma(\alpha, x; b)$.

NonGaussian multivariate probability models generally lead to nonlinear estimators, and therefore such an approach provides a way to develop nonlinear signal processing rules. However, often either the MAP estimator is employed because it is simpler to derive, or numerical integration is required to implement an MMSE estimator. This paper illustrates that for spherically contoured Laplace random vectors in AWGN, the MMSE estimator is available in closed form and can be implemented efficiently.

Note that in practice, the signal $\mathbf{s}$ of interest will likely not follow a multivariate Laplace distribution exactly, in which case the MAP estimator may yield a lower MSE than the MMSE estimator. The results and figures in Section V are based on the assumption that $\mathbf{s}$ follows a multivariate Laplace distribution, which, being a one-parameter distribution, is a simple probability model. The extension of this work to a two-parameter family of multivariate densities, for example the multivariate Bessel K form density or the multivariate normal inverse Gaussian (NIG), would provide a more flexible family of nonlinear estimators.

In wavelet-based denoising, it is common to exploit statistical dependencies between scales. In that case it is desirable to

take into account the different variances of the wavelet coefficients in different scales. (Typically, wavelet coefficients at fine scales have smaller variance than wavelet coefficients at coarse scales.) To apply the multivariate Laplace model and MMSE estimator to that problem, it is desirable that a nonspherically contoured form of the density be used. In wavelet-based denoising, it is also common to use redundant (overcomplete, or expansive) transforms — they provide superior results compared to nonredundant transforms. In that case, the noise in the wavelet domain will be correlated, as will be the signal components of the wavelet coefficients. For this case, it will again be desirable that a nonspherically contoured Laplace density, and corresponding estimators, be used. Therefore, it will be of interest to extend the results in this paper by developing an exact or approximate MMSE estimator for general elliptically contoured Laplace densities in additive correlated noise.

However, for the general elliptically contoured multivariate Laplace density (where the components have different variances, or where there is nonzero correlation between components) the derivations in this paper break down. The simplification of the MMSE estimator using the generalized incomplete gamma function does not appear to be possible. We can however, obtain an implicit expression for the MAP estimator as in (32), which can likewise be solved by successive substitution; preliminary results are given in [56]. However, recalling Fig. 10 and the related discussion, it can be predicted that the MAP estimator will be a poor substitute for the MMSE estimator. The development of an effective computationally efficient MMSE estimator for this case remains an open question.

Appendix A
Bessel Functions

The modified Bessel function of the second kind is defined as

$$K_\lambda(u) = \frac{1}{2}\left(\frac{u}{2}\right)^\lambda \int_0^\infty t^{-\lambda-1}\exp\left(-t-\frac{u^2}{4\,t}\right)dt. \quad (50)$$

For $\lambda = 1/2$, there is a simple form for $K_\lambda(u)$

$$K_{1/2}(u) = \sqrt{\frac{\pi}{2\,u}}\exp(-u). \quad (51)$$

The derivative of the Bessel function is given by

$$\frac{\partial}{\partial u}K_\lambda(u) = \frac{\lambda}{u}K_\lambda(u) - K_{\lambda+1}(u). \quad (52)$$

For large u there is the approximation

$$K_\lambda(u) \approx \sqrt{\frac{\pi}{2\,u}}\exp(-u), \quad u \gg 1. \quad (53)$$

For large u there is also the approximation

$$\frac{K_\lambda(u)}{K_{\lambda-1}(u)} \approx 1 + \left(\lambda - \frac{1}{2}\right)\frac{1}{u}, \qquad u \gg 1. \quad (54)$$

Appendix B
Gaussian PDFs and Convolution

The following identities are used in the derivations. If the random variables x and n are zero-mean Gaussian with variance σ^2 and σ_n^2 respectively, then

$$\int \frac{1}{a}p_x\left(\frac{s}{a}\right)p_n(y-s)\,ds = \frac{1}{\sqrt{2\pi\,(a^2\,\sigma^2+\sigma_n^2)}} \times\exp\left(-\frac{y^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right) \quad (55)$$

$$\int \frac{s}{a}p_x\left(\frac{s}{a}\right)p_n(y-s)ds = \frac{y\,a^2\,\sigma^2}{a^2\,\sigma^2+\sigma_n^2} \times \frac{1}{\sqrt{2\pi\,(a^2\,\sigma^2+\sigma_n^2)}} \exp\left(-\frac{y^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right) \quad (56)$$

$$\int \frac{s^2}{a}p_x\left(\frac{s}{a}\right)p_n(y-s)\,ds = \frac{1}{\sqrt{2\pi\,(a^2\,\sigma^2+\sigma_n^2)}} \left(\frac{y^2\,a^4\,\sigma^4}{(a^2\,\sigma^2+\sigma_n^2)^2}+\frac{a^2\,\sigma^2\,\sigma_n^2}{(a^2\,\sigma^2+\sigma_n^2)}\right) \times\exp\left(-\frac{y^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right). \quad (57)$$

Appendix C
Derivation of MMSE Estimator

Using (9), the numerator of (39) is given by

$$\begin{aligned}&\int_{\mathbb{R}^d} s_i\,p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\,p_{\mathbf{s}}(\mathbf{s})\,d\mathbf{s}\\ &\quad= \int_{\mathbb{R}^d} s_i\,p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\left[\int_0^\infty p_a(a)\,\frac{1}{a^d}\,p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right)da\right]d\mathbf{s}\\ &\quad= \int_0^\infty p_a(a)\left[\int_{\mathbb{R}^d}\frac{s_i}{a^d}\,p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right)\,p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\,d\mathbf{s}\right]da \quad (58)\end{aligned}$$

where $p_a(a)$ is given by (11) and $p_{\mathbf{x}}(\mathbf{x})$ is given by (8). The inner integral can be manipulated similar to (23) — because $p_{\mathbf{n}}$ and $p_{\mathbf{x}}$ are white Gaussian the multivariate integral is separable and the inner integral can be obtained. It is the convolution of $p_{\mathbf{n}}(\mathbf{y})$ and $s_i\,p_{\mathbf{x}}(\mathbf{y}/a)/a^d$ where both are Gaussian. Using (56)

$$\int_{\mathbb{R}^d}\frac{s_i}{a^d}\,p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right)\,p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\,d\mathbf{s} = \frac{y_i\,a^2\,\sigma^2}{a^2\,\sigma^2+\sigma_n^2} \times\frac{1}{(2\,\pi\,(a^2\,\sigma^2+\sigma_n^2))^{d/2}}\,\exp\left(-\frac{y_1^2+\cdots+y_d^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right). \quad (59)$$

Using (11) and (59) in (58) gives

$$\int_{\mathbb{R}^d} s_i\,p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\,p_{\mathbf{s}}(\mathbf{s})\,d\mathbf{s} = \frac{1}{(2\,\pi)^{d/2}} \times\int_0^\infty \frac{y_i\,a^2\,\sigma^2}{(a^2\,\sigma^2+\sigma_n^2)^{d/2+1}}\exp\left(-a^2-\frac{\|\mathbf{y}\|^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right)2\,a\,da.$$

Changing the of variable of integration, using $t = a^2 + \sigma_n^2/\sigma^2$, gives

$$\int_{\mathbb{R}^d} s_i\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})\, d\mathbf{s} = \frac{1}{(2\,\pi)^{d/2}} \times \int_{\sigma_n^2/\sigma^2}^{\infty} \frac{y_i \left(\frac{t-\sigma_n^2}{\sigma^2}\right) \sigma^2}{(t\,\sigma^2)^{d/2+1}} \exp\left(-t + \frac{\sigma_n^2}{\sigma^2} - \frac{\|\mathbf{y}\|^2}{2\,(t\,\sigma^2)}\right) dt.$$

Rearranging gives

$$\int_{\mathbb{R}^d} s_i\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})\, d\mathbf{s} = \frac{y_i}{(2\,\pi\,\sigma^2)^{d/2}\,\sigma^2} \exp\left(\frac{\sigma_n^2}{\sigma^2}\right) \times \int_{\sigma_n^2/\sigma^2}^{\infty} \left(\frac{1}{t^{d/2}} - \frac{\sigma_n^2}{\sigma^2}\frac{1}{t^{d/2+1}}\right) \exp\left(-t - \frac{\frac{\|\mathbf{y}\|^2}{(2\,\sigma^2)}}{t}\right) dt.$$

Using the generalized incomplete gamma function, we have

$$\int_{\mathbb{R}^d} s_i p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}) p_{\mathbf{s}}(\mathbf{s}) d\mathbf{s} = \frac{y_i}{(2\pi\sigma^2)^{d/2}} \exp\left(\frac{\sigma_n^2}{\sigma^2}\right) \times \left[\Gamma\left(1-\frac{d}{2}, \frac{\sigma_n^2}{\sigma^2}; \frac{\|\mathbf{y}\|^2}{2\sigma^2}\right) - \frac{\sigma_n^2}{\sigma^2}\Gamma\left(-\frac{d}{2}, \frac{\sigma_n^2}{\sigma^2}; \frac{\|\mathbf{y}\|^2}{2\sigma^2}\right)\right]. \quad (60)$$

Using (26) and (60) in (39) gives the MMSE estimator of s_i in (40).

Appendix D
Derivation of MSE

In the following, we find a formula for $\mathbb{E}[s_i^2\,|\,\mathbf{y}]$

$$\mathbb{E}[s_i^2\,|\,\mathbf{y}] = \frac{1}{p_{\mathbf{y}}(\mathbf{y})} \int_{\mathbb{R}^d} s_i^2\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})\, d\mathbf{s}. \quad (61)$$

The pdf $p_{\mathbf{y}}(\mathbf{y})$ is given by (26); the pdf $p_{\mathbf{s}}(\mathbf{s})$ is given by (9) and (11) [equivalently (13)]; the pdf $p_{\mathbf{n}}(\mathbf{n})$ is given by (20).

Using (9), the numerator of (61) is given by

$$\int_{\mathbb{R}^d} s_i^2\, p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, p_{\mathbf{s}}(\mathbf{s})\, d\mathbf{s} = \int_0^{\infty} p_a(a) \left[\int_{\mathbb{R}^d} \frac{s_i^2}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right) p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s}\right] da \quad (62)$$

where $p_a(a)$ is given by (11) and $p_{\mathbf{x}}(\mathbf{x})$ is given by (8). The inner integral can be manipulated similar to (23) — because $p_{\mathbf{n}}$ and $p_{\mathbf{x}}$ are white Gaussian the multivariate integral is separable and the inner integral can be obtained. It is the convolution of $p_{\mathbf{n}}(\mathbf{y})$ and $s_i^2\, p_{\mathbf{x}}(\mathbf{y}/a)/a^d$ where both are Gaussian. Using (57)

$$\int_{\mathbb{R}^d} \frac{s_i^2}{a^d}\, p_{\mathbf{x}}\left(\frac{\mathbf{s}}{a}\right) p_{\mathbf{n}}(\mathbf{y}-\mathbf{s})\, d\mathbf{s} = \left(\frac{y_i^2\, a^4\, \sigma^4}{(a^2\,\sigma^2+\sigma_n^2)^2} + \frac{a^2\,\sigma^2\,\sigma_n^2}{(a^2\,\sigma^2+\sigma_n^2)}\right) \times \frac{1}{(2\,\pi\,(a^2\,\sigma^2+\sigma_n^2))^{d/2}} \exp\left(-\frac{y_1^2+\cdots+y_d^2}{2\,(a^2\,\sigma^2+\sigma_n^2)}\right). \quad (63)$$

Using (11) and (63) in (62) gives

$$\int_{\mathbb{R}^d} s_i^2 p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}) p_{\mathbf{s}}(\mathbf{s}) d\mathbf{s} = \frac{1}{(2\pi)^{d/2}} \times \int_0^{\infty} \left(\frac{y_i^2 a^4 \sigma^4}{(a^2\sigma^2+\sigma_n^2)^2} + \frac{a^2\sigma^2\sigma_n^2}{(a^2\sigma^2+\sigma_n^2)}\right) \times \frac{1}{(a^2\sigma^2+\sigma_n^2)^{d/2}} \exp\left(-a^2 - \frac{\|\mathbf{y}\|^2}{2(a^2\sigma^2+\sigma_n^2)}\right) 2a da. \quad (64)$$

Changing the variable of integration, using $t = a^2 + \sigma_n^2/\sigma^2$, rearranging, and using the generalized incomplete gamma function, gives

$$\int_{\mathbb{R}^d} s_i^2 p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}) p_{\mathbf{s}}(\mathbf{s}) d\mathbf{s} = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(\frac{\sigma_n^2}{\sigma^2}\right) \times \left[(y_i^2+\sigma_n^2)\Gamma\left(1-\frac{d}{2}, \frac{\sigma_n^2}{\sigma^2}; \frac{\|\mathbf{y}\|^2}{2\sigma^2}\right) - (2y_i^2+\sigma_n^2)\frac{\sigma_n^2}{\sigma^2}\Gamma\left(-\frac{d}{2}, \frac{\sigma_n^2}{\sigma^2}; \frac{\|\mathbf{y}\|^2}{2\sigma^2}\right) + y_i^2\frac{\sigma_n^4}{\sigma^4}\Gamma\left(-1-\frac{d}{2}, \frac{\sigma_n^2}{\sigma^2}; \frac{\|\mathbf{y}\|^2}{2\sigma^2}\right)\right]. \quad (65)$$

Using (26) and (65) in (61) gives (45).

References

[1] A. Achim and E. E. Kuruoğlu, "Image denoising using bivariate α-stable distributions in the complex wavelet domain," *IEEE Signal Process. Lett.*, vol. 12, pp. 17–20, Jan. 2005.

[2] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 8, pp. 1773–1784, Aug. 2003.

[3] T. I. Alecu, S. Voloshynovskiy, and T. Pun, "Denoising with infinite mixture of Gaussians," presented at the Eur. Signal Process. Conf. (EUSIPCO), Antalya, Turkey, Sep. 2005.

[4] T. J. Barnard and D. D. Weiner, "NonGaussian clutter modeling with generalized spherically invariant random vectors," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2384–2390, 1996.

[5] K. A. Birney and T. R. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Trans. Image Process.*, vol. 4, no. 2, pp. 186–193, Feb. 1995.

[6] L. Boubchir and M. J. Fadili, "A closed-form nonparametric Bayesian estimator in the wavelet-domain of images using an approximate α-stable prior," *Pattern Recognit.Letters*, vol. 27, no. 12, pp. 1370–1382, Sep. 2006.

[7] M. T. Boudjelkha and M. A. Chaudhry, "On the approximation of a generalized incomplete Gamma function arising in heat conduction problems," *J. Math. Anal. Appl.*, vol. 248, no. 2, pp. 509–519, Aug. 2000.

[8] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Process.*, vol. 12, no. 2, pp. 119–141, Mar. 1987.

[9] T. Cai and B. W. Silverman, "Incorporating information on neighboring coefficients into wavelet estimation," *Sankhya: Ind. J. Stat. B*, vol. 63, pp. 127–148, 2001.

[10] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[11] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modelling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

[12] M. A. Chaudhry and S. M. Zubair, "Generalized incomplete Gamma functions with applications," *J. Comput. Appl. Math.*, vol. 55, no. 1, pp. 99–124, 1994.

[13] M. A. Chaudhry and S. M. Zubair, *On a Class of Incomplete Gamma Functions with Applications*. New York: Chapman & Hall, 2001.
[14] H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Am. Stat. Assoc.*, vol. 92, no. 439, pp. 1413–1421, Dec. 1997.
[15] D. Cho and T. D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Process.: Image Commun.*, vol. 20, no. 1, pp. 77–89, Jan. 2005.
[16] M. Clyde and E. I. George, "Empirical bayes estimation in wavelet nonparametric regression," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, Eds. New York: Springer-Verlag, 1999, pp. 309–322.
[17] E. Conte, M. Lops, and G. Ricci, "Radar detection in K-distributed clutter," *IEE Proceedings – Radar, Sonar and Navigation*, vol. 141, no. 2, pp. 116–118, 1994.
[18] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.
[19] I. Daubechies, *Ten Lectures on Wavelets*. Singapore: SIAM, 1992.
[20] G. Deng, "Generalized Wiener estimation algorithms based on a family of heavy-tailed distributions," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2005, pp. 457–460.
[21] S. Durand and J. Froment, "Reconstruction of wavelet coefficients using total variation minimization," *SIAM J. Sci. Comput.*, vol. 24, no. 5, pp. 1754–1767, 2003.
[22] T. Eltoft, T. Kim, and T.-W. Lee, "On the multivariate Laplace distribution," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 300–303, May 2006.
[23] I. L. Eom and Y. S. Kim, "Wavelet-based denoising with nearly arbitrarily shaped windows," *IEEE Signal Process. Lett.*, vol. 11, no. 12, pp. 937–940, Dec. 2003.
[24] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 231–240, Feb. 2005.
[25] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1322–1331, Nov. 2001.
[26] S. Gazor and W. Zhang, "Speech enhancement employing Laplacian-Gaussian mixture," *IEEE Trans. Signal Process.*, vol. 13, no. 5, pp. 896–904, Sep. 2005.
[27] P. V. Gehler and M. Welling, "Product of "edge-perts"," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 419–426, 2005.
[28] U. Grenander and A. Srivastava, "Probability models for clutter in natural images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 424–429, Apr. 2001.
[29] M. Hansen and B. Yu, "Wavelet thresholding via MDl for natural images," *IEEE Trans. Inf. Theory*, vol. 46, no. 8, pp. 1778–1788, Aug. 2000.
[30] J. Huang, "Statistics of Natural Images and Models," Ph.D. dissertation, Brown Univ., Providence, RI, 2000.
[31] A. Hyvärinen, "Sparse code shrinkage: Denoising of nonGaussian data by maximum likelihood estimation," *Neural Comput.*, vol. 11, pp. 1739–1768, 1999.
[32] A. Hyvärinen, E. Oja, and P. Hoyer, "Image denoising by sparse code shrinkage," in *Intelligent Signal Process.*, S. Haykin and B. Kosko, Eds. Piscataway, NJ: IEEE Press, 2001.
[33] M. Jansen and A. Bultheel, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *J. Amer. Statist. Assoc. (JASA)*, vol. 96, no. 454, pp. 629–639, Jun. 2001.
[34] R. Jenssen, T. A. Øigård, T. Eltoft, and A. Hanssen, "Sparse code shrinkage using the normal inverse Gaussian density mode," in *Proc. Int. Workshop Independent Component Analysis and Blind Signal Separation (ICA2001)*, San Diego, CA, Dec. 2001, pp. 212–217.
[35] I. M. Johnstone and B. W. Silverman, "Empirical Bayes selection of wavelet thresholds," *Ann. Statist.*, vol. 33, no. 4, pp. 1700–1752, 2005.
[36] M. Kazubek, "Wavelet domain image denoising by thresholding and Wiener filtering," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 324–326, Nov. 2003.
[37] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations*. New York: Birkhauser, 2001.
[38] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, Nov. 2001.
[39] D. Lorenz, "Wavelet shrinkage in signal and image processiing – An investigation of relations and equivalences," Ph.D. dissertation, Univ. Bremen, Bremen, Germany, 2005.
[40] T. Lotter and P. Vary, "Noise reduction by joint maximum *a posteriori* spectral amplitude and phase estimation with super-Gaussian speech model," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 1457–1460.
[41] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.
[42] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[43] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.
[44] B. Matalon, M. Elad, and M. Zibulevsky, "Improved denoising of images using modelling of a redundant contourlet transform," in *Proc. SPIE, Wavelets XI*, San Diego, CA, Jul. 2005, vol. 5914.
[45] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.
[46] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 909–919, Apr. 1999.
[47] M. Nikolova, "Estimées locales fortement homogénes," *Comptes Rendus Acad. Sci. Paris—Ser. I*, vol. 325, pp. 665–670, 1997.
[48] R. Öktem, L. Yaroslavsky, K. Egiazarian, and J. Astola, "Transform domain approaches for image denoising," *J. Electron. Imag.*, vol. 11, no. 2, pp. 149–156, Apr. 2002.
[49] H.-J. Park and T.-W. Lee, "Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution," in *Adv. Neural Information Processing Systems*, L. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005, vol. 17, pp. 1041–1048.
[50] A. Pizurica and W. Philips, "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654–665, Mar. 2006.
[51] A. Pizurica, W. 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.
[52] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
[53] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for nonGaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2101–2111, Jun. 2005.
[54] F. Ruggeri and B. Vidakovic, "Bayesian modeling in the wavelet domain," in *Handbook of Statistics 25: Bayesian Thinking, Modeling and Computation*, D. Dey and C. R. Rao, Eds. The Netherlands: North Holland, 2005.
[55] I. W. Selesnick, "Laplace random vectors, Gaussian noise, and the generalized incomplete Gamma function," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Atlanta, Oct. 2006, pp. 2097–2100.
[56] I. W. Selesnick, "Modeling and estimation of wavelet coefficients using elliptically contoured multivariate Laplace vectors," in *Proc. SPIE Wavelets XII*, Aug. 26–29, 2007, vol. 6701.
[57] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.
[58] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Process. Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.
[59] L. Sendur and I. W. Selesnick, "Multivariate shrinkage functions for wavelet-based denoising," in *Proc. Asilomar Conf. Sig., Sys. Comp.*, Nov. 2002, vol. 1, pp. 953–957.
[60] F. Shi and I. W. Selesnick, "An elliptically contoured exponential mixture model for wavelet based image denoising," *Appl. Comput. Harmon. Anal.*, vol. 23, no. 1, pp. 131–151, Jul. 2007.
[61] P.-L. Shui and Y.-B. Zhao, "Image denoising algorithm using doubly local Wiener filtering with block-adaptive windows in wavelet domain," *Signal Process.*, vol. 87, no. 7, pp. 1721–1734, Jul. 2007.
[62] E. P. Simoncelli, "Modelling the joint statistics of images in the wavelet domain," *Proc. SPIE*, vol. 313, no. 1, pp. 188–195, 1999.
[63] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Sep. 16–19, 1996, vol. 1, pp. 379–382.

[64] A. Srivastava, "Stochastic models for capturing image variability," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 63–76, Sep. 2002.
[65] P. Tsakalides, P. Reveliotis, and C. L. Nikias, "Scalar quantization of heavy-tailed signals," *IEE Vis. Image Signal Process.*, vol. 147, no. 5, pp. 475–484, Oct. 2000.
[66] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: Wiley, 1999.
[67] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 5, pp. 600–608, May 1973.

Ivan W. Selesnick (SM'08) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Rice University, Houston, TX, in 1990, 1991, and 1996, respectively.

In 1997, he was a Visiting Professor at the University of Erlangen-Nurnberg, Germany. He then joined the Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY, where he is an Associate Professor. His current research interests are in the area of digital signal processing, wavelet-based signal processing, and non-Gaussian probability models.

Dr. Selesnick received a DARPA-NDSEG fellowship in 1991. In 1996, his Ph.D. dissertation received the Budd Award for Best Engineering Thesis at Rice University and an award from the Rice-TMC chapter of Sigma Xi. He received an Alexander von Humboldt Award (1997) and a National Science Foundation Career award (1999). He has been a member of the IEEE SPTM Technical Committee, an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and is currently an Associate Editor of IEEE SIGNAL PROCESSING LETTERS.