On the Multivariate Laplace Distribution

Torbjørn Eltoft, *Member, IEEE*, Taesu Kim, *Student Member, IEEE*, and Te-Won Lee, *Member, IEEE*

***Abstract*—In this letter, we discuss the multivariate Laplace probability model in the context of a normal variance mixture model. We briefly review the derivation of the probability density function (pdf) and discuss a few important properties. We then present two methods for estimating its parameters from data and include an example of usage, where we apply the model to represent the statistics of the discrete Fourier transform coefficients of a speech signal. Since the pdf is given in closed form, and the model parameters can be easily obtained, this distribution may be useful for representing multivariate, sparsely distributed data, with mutually dependent components.**

***Index Terms*—Multidimensional Laplace distribution, multivariate Laplace distribution, normal variance mixture model, scale mixture of Gaussians model, statistical modeling.**

I. Introduction

In many real-world data sets involving multivariate observations, the data have an empirical distribution that is highly peaked at zero (or the mean vector) and that asymptotically falls off more slowly than the Gaussian distribution as the distance from zero increases. We denote these distributions as *sparse distributions*. Sparse distributions are appropriate for representing the statistics of speech and image data, especially when observed in a transform domain like the wavelet or discrete Fourier transform (DFT) domain [1]. For example, the overcomplete wavelet transform coefficients of images are found to have sparse distributions, a property that has been extensively exploited in coding and denoising [2], [3]. Several authors have studied the statistics of speech signals in various transform domains and found the coefficients to have sparse, heavy-tailed distributions [4], [5]. Sparse distributions are also frequently encountered in various machine learning areas, e.g., independent component analysis (ICA) [6].

Multivariate observations, which are mutually correlated and have higher-order dependencies, have frequently been represented using a mixture of Gaussians models. These are convenient in many respects, they have a closed-form probability density function (pdf), and the parameters can easily be obtained using the EM algorithm. Recently, yet another class of mixture models, the so-called scale-mixture of Gaussian models, have emerged as a powerful set of distributions for modeling statistical dependencies in multivariate data [1]. The normal inverse Gaussian (NIG) distribution [7] is an example of this kind of model. The NIG pdf has been widely applied to model economical time series, and it has also been successfully applied in some engineering problems [8]. The parameters of this model are also estimated using an EM type of algorithm.

In [9], the multivariate Laplace (ML) distribution was presented as a specific multivariate Linnik distribution with Linnik parameter $\alpha = 2$. In this letter, we discuss the ML model as a multivariate scale mixture of Gaussian models (similar to the presentation of the multivariate NIG model in [10]), using an exponential prior for the scale factor. This formulation has advantages in the sense that the model becomes easy to analyze, and we can find efficient methods for estimating its parameters from data. We present two methods for estimating the model parameters. The first method is moment based and follows directly from the generative equation of the model. The second is an iterative EM-type approach. The latter algorithm may be favorable if *covariance selection* is applied, like, for instance, in speech processing, where the precision matrix could be sparse, or in radar polarimetry, where the covariance matrix could have a given structure.

We include an example of usage, where we apply the ML model to represent the fast Fourier transform (FFT) coefficients of a speech signal.

II. Multivariate Laplace Distribution

In [11], it was shown that if the pdf of some random variable Y, $p_Y(y)$ is symmetric about zero, and the derivatives of $p_Y(y)$ satisfy

$$\left(-\frac{d}{dy}\right)^k p_Y(y) \geq 0 \text{ for } y > 0 \tag{1}$$

then there exist independent variables X and Z, with X being a standard normal variable, such that

$$Y = \sqrt{Z}X. \tag{2}$$

The variable Z is allowed to take on only positive values. A random variable Y, which can be expressed as in (2), is referred to as *a normal variance mixture model* or *a scale mixture of Gaussians*. Of course, if the mean of Y should be nonzero, (2) may be modified by adding a scalar μ corresponding to the actual mean value. Now, let $p_Z(z)$ be the pdf of Z. Then, the marginal pdf of Y is obtained by averaging over Z, as in

$$p_Y(y) = \int_0^\infty \frac{1}{\sqrt{2\pi z}} \exp\left(-\frac{(y-\mu)^2}{2z}\right) p_Z(z) dz. \tag{3}$$

Many distributions satisfy the properties of (1), and among these, we find the Laplace distribution (also called the double

Manuscript received August 19, 2005; revised December 13, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian Sadler.

T. Eltoft is with the Department of Physics, University of Tromsø, Tromsø, Norway, and also with Norut IT, Tromsø, Norway (e-mail: torbjorn.eltoft@phys.uit.no).

T. Kim and T-W. Lee are with the Institute of Neural Computation, University of California San Diego, La Jolla, CA 92093 USA (e-mail: taesu@ucsd.edu; tewon@ucsd.edu).

Digital Object Identifier 10.1109/LSP.2006.870353

exponential distribution). In fact, it is easy to verify that if Z is an exponential stochastic variable with pdf

$$p_Z(z) = \frac{1}{\lambda}\exp\left(-\frac{z}{\lambda}\right) \tag{4}$$

and X is a standard normal variable, then Y, generated as $Y = \mu + \sqrt{Z}X$, will have pdf

$$p_Y(y) = \frac{1}{2}\sqrt{\frac{2}{\lambda}}\exp\left(-\sqrt{\frac{2}{\lambda}}|y-\mu|\right). \tag{5}$$

Equation (5) is recognized as the pdf of a Laplace distribution centered at μ.

The multidimensional extension of the generative model described above is straightforward. Let $\boldsymbol{X}$ be a d-dimensional, zero-mean Gaussian variable with covariance matrix equal to the identity matrix. Furthermore, let $\boldsymbol{\Gamma} \in \mathcal{R}^{d\times d}$ be a positive definite matrix with determinant $\det\boldsymbol{\Gamma} = 1$ and assume that Z is a scalar exponential random variable with pdf as in (4). We now generate a new variable $\boldsymbol{Y}$ as *a multivariate scale mixture of Gaussians* according to

$$\boldsymbol{Y} = \boldsymbol{\mu} + \sqrt{Z}\boldsymbol{\Gamma}^{(1/2)}\boldsymbol{X} \tag{6}$$

where $\boldsymbol{\mu}$ is the mean vector. It is now immediately observed that the conditional distribution of $\boldsymbol{Y}$, given Z, is multivariate Gaussian with pdf given as

$$p_{\boldsymbol{Y}|Z}(\boldsymbol{y}|Z=z) = \frac{1}{(2\pi z)^{(d/2)}} \times \exp\left[-\frac{1}{2z}(\boldsymbol{y}-\boldsymbol{\mu})^t\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]. \tag{7}$$

The matrix $\boldsymbol{\Gamma}$ defines the internal covariance structure of the variables of $\boldsymbol{Y}$. For this reason, we will refer to this matrix as the covariance structure matrix. Let us define

$$q(\boldsymbol{y}) = (\boldsymbol{y}-\boldsymbol{\mu})^t\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}). \tag{8}$$

To obtain the marginal pdf of $\boldsymbol{Y}$, we have to perform an integration similar to the one in (3) over the prior distribution $p_Z(z)$. The result of the integration (see the Appendix), using the prior in (4), turns out to be given as

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{(2\pi)^{(d/2)}}\frac{2}{\lambda}\frac{K_{(d/2)-1}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y})}\right)}{\left(\sqrt{\frac{\lambda}{2}q(\boldsymbol{y})}\right)^{(d/2)-1}} \tag{9}$$

where $K_m(x)$ denotes the modified Bessel function of the second kind and order m, evaluated at x. We denote the pdf in (9) as the *the ML distribution*, because its generative model is a multidimensional equivalent to the generative model of the one-dimensional Laplace distribution. It is also easily confirmed using the polynomial expansion of the Bessel K-function that when d is set equal to 1 in (9), we obtain the pdf in (5). We will use the notation $\boldsymbol{Y} \sim \mathrm{ML}\{\lambda, \boldsymbol{\mu}, \boldsymbol{\Gamma}\}$ to denote that $\boldsymbol{Y}$ is an ML distributed variable with parameters λ, $\boldsymbol{\mu}$, and $\boldsymbol{\Gamma}$.

A. *Some Properties of the ML Distribution*

From (6), we find that the mean and covariance matrix of $\boldsymbol{Y}$ is given as

$$\mathcal{E}\{\boldsymbol{Y}\} = \mu_{\boldsymbol{Y}} = \boldsymbol{\mu} \tag{10}$$

$$\mathcal{E}\{(\boldsymbol{Y}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})^t\} = \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \lambda\boldsymbol{\Gamma}. \tag{11}$$

We observe that $p_{\boldsymbol{Y}}(\boldsymbol{y})$ is a function of $\boldsymbol{y}$ through $q(\boldsymbol{y})$. If $\boldsymbol{\Gamma}$ is diagonal, it follows from (11) that two components Y_i and Y_j of $\boldsymbol{Y}$ will be uncorrelated, but $p_{\boldsymbol{Y}}(\boldsymbol{y})$ does not factorize into a product of the component pdfs. Hence, the components of $\boldsymbol{Y}$ are not statistically independent. However, the joint distribution conditioned on Z, which is Gaussian, will factorize, meaning that $Y_i|Z$ and $Y_j|Z$ are independent.

Let us for the moment assume that $\boldsymbol{\mu} = \boldsymbol{0}$, and let $||\boldsymbol{y}||_{\boldsymbol{\Gamma}^{-1}} = \sqrt{\boldsymbol{y}^t\boldsymbol{\Gamma}^{-1}\boldsymbol{y}}$ denote the weighted $\boldsymbol{\Gamma}^{-1}$-norm of $\boldsymbol{y}$. Noting that the Bessel function behaves as

$$K_d(x) \sim \sqrt{\frac{\pi}{2x}}\exp(-x), \text{ when } |x| \to \infty$$

we find that

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) \sim \frac{\exp\left(-\sqrt{\frac{2}{\lambda}}||\boldsymbol{y}||_{\boldsymbol{\Gamma}^{-1}}\right)}{||\boldsymbol{y}||_{\boldsymbol{\Gamma}^{-1}}^{((d/2)-(1/2))}} \text{ for large } ||\boldsymbol{y}||_{\boldsymbol{\Gamma}^{-1}}. \tag{12}$$

We observe that the ML distribution is more heavy-tailed than the normal distribution. We also remark that for $d > 1$, $p_{\boldsymbol{Y}}(\boldsymbol{y}) \to \infty$, as $\boldsymbol{y} \to \boldsymbol{\mu}$.

The moment-generating function of the ML distribution is found to be

$$M_Y(\boldsymbol{\omega}) = E\{e^{\boldsymbol{\omega}^t\boldsymbol{Y}}\} = \frac{e^{\boldsymbol{\mu}^t\boldsymbol{\omega}}}{1-\lambda\frac{\boldsymbol{\omega}^t\boldsymbol{\Gamma}\boldsymbol{\omega}}{2}}. \tag{13}$$

Now, let $\boldsymbol{V} = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{b}$ be an arbitrary linear transformation of a $\mathrm{ML}\{\lambda, \boldsymbol{\mu}, \boldsymbol{\Gamma}\}$ random vector $\boldsymbol{Y}$, where $\boldsymbol{A}$ is a $d \times d$ real-valued matrix. The transformed variable $\boldsymbol{V}$ can then be shown to be another ML distributed random vector, with parameters $\{\tilde{\lambda}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Gamma}}\}$, where

$$\tilde{\lambda} = \lambda|\det\boldsymbol{A}|^{(1/d)} \tag{14}$$

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} \tag{15}$$

$$\tilde{\boldsymbol{\Gamma}} = \boldsymbol{A}\boldsymbol{\Gamma}\boldsymbol{A}^t|\det\boldsymbol{A}|^{-(2/d)}. \tag{16}$$

III. Parameter Estimation

The model we have used to generate the ML distribution involves the latent variable Z, which means that the parameters of the pdf may be estimated using an iterative procedure. At the outset, we note that the *a posteriori* pdf $p_{Z|\boldsymbol{Y}}(z|\boldsymbol{y})$ is given as

$$p_{Z|\boldsymbol{Y}}(z|\boldsymbol{y}) = \frac{\left(\sqrt{\frac{\lambda}{2}q(\boldsymbol{y})}\right)^{-(d/2)+1}}{2K_{-(d/2)+1}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y})}\right)} \times z^{-(d/2)}\exp\left[-\frac{1}{2}\left(\frac{q(\boldsymbol{y})}{z}+\frac{2}{\lambda}z\right)\right]. \tag{17}$$

This pdf is, according to [7], a generalized inverse Gaussian (GIG) distribution, with parameters $\{(-(d/2)+1), \sqrt{q(\boldsymbol{y})}, \sqrt{(2/\lambda)}\}$. We denote this pdf as GIG$\{z; (-(d/2)+1), \sqrt{q(\boldsymbol{y})}, \sqrt{(2/\lambda)}\}$. The k-order moments of a GIG$\{z;\theta,\delta,\gamma\}$ distribution are given as [7]

$$\mu(k)_{\mathrm{GIG}} = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{\theta+k}(\delta\gamma)}{K_\theta(\delta\gamma)}. \quad (18)$$

Hence, it follows that for a given observation $\boldsymbol{y}_i$, we get

$$\eta_i = \mathcal{E}\{Z|\boldsymbol{Y}=\boldsymbol{y}_i\} = \sqrt{\frac{\lambda q(\boldsymbol{y}_i)}{2}} \frac{K_{-(d/2)+2}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y}_i)}\right)}{K_{-(d/2)+1}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y}_i)}\right)} \quad (19)$$

and

$$\xi_i = \mathcal{E}\{\frac{1}{Z}|\boldsymbol{Y}=\boldsymbol{y}_i\} = \sqrt{\frac{2}{\lambda q(\boldsymbol{y}_i)}} \frac{K_{-(d/2)}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y}_i)}\right)}{K_{-(d/2)+1}\left(\sqrt{\frac{2}{\lambda}q(\boldsymbol{y}_i)}\right)}. \quad (20)$$

A. Parameter Estimation. Method I

Equation (10) suggests that $\boldsymbol{\mu}$ can be estimated as the first-order moment of the sample set $\mathcal{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\}$, i.e., $\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{y}_i$, whereas according to (11), $\boldsymbol{\Gamma}$ can be estimated as the sample covariance matrix divided by an estimate of λ. Let $\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{i=1}^N (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}})^t$ denote the sample covariance matrix. Using the fact that $\det\boldsymbol{\Gamma} = 1$, an estimate of λ would be given as

$$\hat{\lambda} = \det\hat{\boldsymbol{R}}^{(1/d)} \quad (21)$$

and accordingly, an estimate of $\boldsymbol{\Gamma}$ is $\hat{\boldsymbol{\Gamma}} = \frac{1}{\hat{\lambda}}\hat{\boldsymbol{R}}$.

B. Parameter Estimation. Method II

The estimation of the parameters $\lambda, \boldsymbol{\mu}, \boldsymbol{\Gamma}$ of the ML distribution may also be done in a maximum-likelihood approach, using an EM-type algorithm. In this case, the E-step involves updating the first-order and inverse first-order moments of $Z|\boldsymbol{Y}$, and the M-step updates the parameters. This method is to be preferred if the data dimension d is large, and we know that the covariance matrix is sparse (has many zeros) and has a predefined structure.

Using the fact that $\boldsymbol{Y}|Z$ is a multivariate Gaussian variable, the log-likelihood function associated with the observation set $\mathcal{Y}$, given the Z_i's, is given as

$$LL(\boldsymbol{\mu},\boldsymbol{\Gamma}) \propto -\frac{N}{2}log(\det\Gamma) - \frac{1}{2}\sum_{i=1}^N \frac{1}{z_i}(\boldsymbol{y}_i - \boldsymbol{\mu})^t\Gamma^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}). \quad (22)$$

Maximizing the log-likelihood function with regard to $\boldsymbol{\mu}$, we obtain

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^N \frac{1}{z_i}\boldsymbol{y}_i}{\sum_{i=1}^N \frac{1}{z_i}} \quad (23)$$

i.e., a weighted sample mean.

Now, since $\boldsymbol{\Gamma}$, and hence also $\boldsymbol{\Gamma}^{-1}$, is a positive definite matrix, it may be factorized as $\boldsymbol{\Gamma}^{-1} = \boldsymbol{U}^t\boldsymbol{D}\boldsymbol{U}$, where $\boldsymbol{U}$ is a unit upper triangular matrix (i.e., it has only ones along its diagonal), and $\boldsymbol{D}$ is a positive diagonal matrix (i.e., all the diagonal elements are positive) [12]. We furthermore rewrite $\boldsymbol{U}$ as $\boldsymbol{U} = \boldsymbol{I} - \boldsymbol{B}$, where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{B}$ is an upper triangular matrix with zeros along its diagonal. It is noted that this factorization of $\boldsymbol{\Gamma}^{-1}$ helps us constrain the precision matrix, since if, for instance, $B_{ij} = 0$ for $i < k$, and $j > i$, then $\Gamma^{-1}_{ij} = \Gamma^{-1}_{ji} = 0$ for the same indexes.

The log-likelihood function now takes the form

$$LL(\boldsymbol{\mu},\boldsymbol{\Gamma}) \propto \frac{N}{2}log(\det\boldsymbol{D}) - \frac{1}{2}\sum_{i=1}^N \frac{1}{z_i}(\boldsymbol{y}_i - \boldsymbol{B}\boldsymbol{y}_i - \tilde{\boldsymbol{\mu}})^t\boldsymbol{D}(\boldsymbol{y}_i - \boldsymbol{B}\boldsymbol{y}_i - \tilde{\boldsymbol{\mu}}) \quad (24)$$

where $\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu} - \boldsymbol{B}\boldsymbol{\mu}$.

Now, maximizing $LL(\boldsymbol{\mu},\boldsymbol{\Gamma})$ with respect to $\boldsymbol{D}$, under the constraint that $\det\boldsymbol{D} = 1$, gives

$$\hat{\boldsymbol{D}}^{-1} = \mathrm{diag}\left\{\sum_{i=1}^N \frac{1}{z_i}(\boldsymbol{U}\boldsymbol{y}_i - \boldsymbol{\mu})(\boldsymbol{U}\boldsymbol{y}_i - \boldsymbol{\mu})^t\right\}. \quad (25)$$

Next, the maximization of $LL(\boldsymbol{\mu},\boldsymbol{\Gamma})$ with respect to $\boldsymbol{B}$ is computed as $d-1$ (one for each row of $\boldsymbol{B}$, except row d) linear regressions of the form

$$\hat{\boldsymbol{b}}_k = \arg\min_{B_{kj}}\left\{\sum_{i=1}^N \frac{1}{z_i}[y_{ik} - \mu_k - \sum_{j=k+1}^d (y_{ij} - \mu_j)B_{kj}]^2\right\} \quad (26)$$

where $\hat{\boldsymbol{b}}_k, k = 1, 2, \ldots, d-1$ denotes an estimate of the kth row of $\boldsymbol{B}$, and the indexes of y_{ij} refer to component j of observation vector $\boldsymbol{y}_i$. In the actual iteration, $\boldsymbol{\mu}$ is replaced by $\hat{\boldsymbol{\mu}}$, and $(1/z_i)$ is replaced by ξ_i.

Summary of parameter estimation:

1) Set l=0. Select some initial estimates for all parameters. We suggest to use $\hat{\lambda}_l = 1$, $\hat{\boldsymbol{\mu}}_l = (1/N)\sum_{i=1}\boldsymbol{y}_i$, and $\hat{\boldsymbol{\Gamma}}_l = (1/N)\sum_{i=1}^N(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}})^t$.
2) Calculate η_i and ξ_i using (19) and (20).
3) Set $l = l+1$. Estimate λ_l using (23) and $\boldsymbol{\mu}_l$ using (26). Estimate $\hat{\boldsymbol{B}}_l$ through linear regression using (29). From $\hat{\boldsymbol{B}}_l$, construct $\hat{\boldsymbol{U}}_l$. Estimate $\hat{\boldsymbol{D}}_l^{-1}$ according to (28) using $\hat{\boldsymbol{U}}_l$ and $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{U}$ and $\boldsymbol{\mu}$, respectively, and replacing $(1/z_i)$ with ξ_i. Invert $\hat{\boldsymbol{D}}_l^{-1}$ to get $\hat{\boldsymbol{D}}_l$.
4) Repeat *2)* and *3)* until convergence.

IV. Log-Likelihood Ratio Tests

In this section, we apply the ML distribution to model DFT coefficients of a speech signal. The signal is sampled at 8 kHz, and the 128-point DFT coefficients are calculated using a Hanning analysis. In order to be able to visualize the proposed model, we will here only discuss the 2-D model referring to the real and imaginary components of a single complex DFT component, i.e., $\boldsymbol{f}_k = (\mathrm{real}(F_k), \mathrm{imag}(F_k))$, where F_k is the kth component. We estimate the parameters of the ML distribution using method II above.

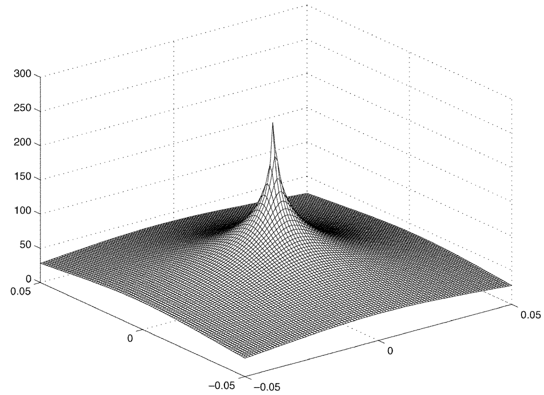

Fig. 1. Example of the 2-D Laplace fitted to the real and imaginary component of the DFT of a speech signal.

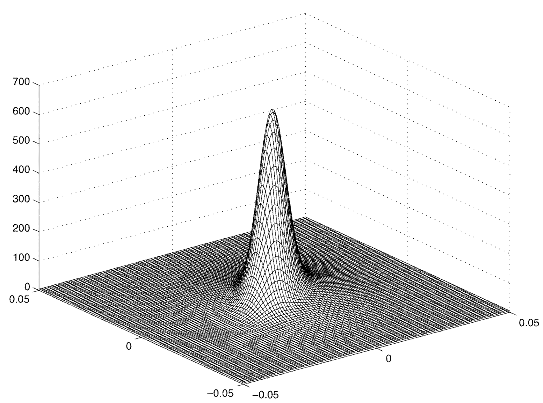

Fig. 2. Example of the ten-component discrete mixture of Gaussians fitted to the real and imaginary component of the DFT of a speech signal.

TABLE I
RESULTS OF LOG-LIKELIHOOD COMPARISONS

Model	ML	MG (1)	MG (3)	(MG (5)	MG (10)	MNIG
$\boldsymbol{f}_2$	887	451	621	622	631	476
$\boldsymbol{f}_{64}$	2242	1040	2167	2222	2233	2223

Furthermore, in order to relate the proposed model to alternative models, we include a tenfold cross-validation log-likelihood test, in which we compare the ML model with a discrete mixture of Gaussians pdf and a multivariate normal inverse Gaussian (MNIG) model [10]. The discrete mixture of Gaussians pdf is given as

$$p_{\boldsymbol{F}}(\boldsymbol{f}) = \sum_{m=1}^{M} \pi_m \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \qquad (27)$$

where $\mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$ is a 2-D Gaussian pdf, with mean $\boldsymbol{\mu}_m$ and $\boldsymbol{\Sigma}_m$. The parameters of this model are estimated from data using the EM algorithm. Figs. 1 and 2 show the resulting 2-D pdfs of frequency component $k = 2$ for the ML model and the mixture of Gaussians model, respectively. We observe that the former is much more peaky at the origin and has a slower decay, as the distance to origin increases. Table I displays the results of the log-likelihood cross-validation tests for one low-frequency and one high-frequency component, using 1, 3, 5, and 10 components in the discrete mixture model. We note that the ML model always has the highest log-likelihood scores, although for $\boldsymbol{f}_{64}$, the scores for the discrete mixture of Gaussian models and the MNIG model are almost equally high. Hence, in the log-likelihood sense, these tests, which by no means are very extensive, indicate that the ML model is a better model.

APPENDIX

To find the pdf of $\boldsymbol{Y}$, we have to solve the integral

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_0^\infty \frac{1}{\lambda(2\pi z)^{(d/2)}} \exp\left[-\frac{q(\boldsymbol{y})}{2z} - \frac{z}{\lambda}\right] dz. \qquad (28)$$

Using the substitutions $\delta^2 = q(\boldsymbol{y})$ and $\gamma^2 = (2/\lambda)$, the integral in (31) may be rewritten as

$$p_Y(\boldsymbol{y}) = \frac{1}{(2\pi)^{-(d/2)}} \frac{\gamma^2}{2} \frac{\sqrt{2\pi} e^{-\delta\gamma}}{\delta} \times \int_0^\infty z^{-((d/2)-(3/2))} \times \frac{\delta}{\sqrt{2\pi}} e^{\delta\gamma} z^{-(3/2)} \exp\left[-\frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right] dz. \qquad (29)$$

The integral is recognized as the $-((d/2)-(3/2))$-order moment of an IG distribution [7]. From [7], we have that the k-order moment of an IG distribution is given as

$$\mu(k)_{IG} = \left(\frac{\delta}{\gamma}\right)^k \frac{K_{-(1/2)+k}(\delta\gamma)}{K_{-(1/2)}(\delta\gamma)}. \qquad (30)$$

Using (33), the integral in (32) is easily solved in closed form, and when resubstituting for δ and γ, we obtain the closed-form expression for $p_{\boldsymbol{Y}}(\boldsymbol{y})$ given in (9).

REFERENCES

[1] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc. SPIE 44th Annu. Meeting*, Denver, CO, Jul. 1999.
[2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
[3] S. Solbø and T. Eltoft, "Homomorphic wavelet-based statistical despeckling of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 711–721, Apr. 2004.
[4] R. Martin and C. Breithaupt, "Speech enhancement in the DFT domain using Laplacian speech priors," in *Proc. IWAENC*, Kyoto, Japan, Sep. 2003.
[5] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.
[6] H. J. Park and T.-W. Lee, "Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA: MIT Press, 2005.
[7] O. E. Barndorff-Nielsen, "Normal inverse Gaussian distributions and stochastic volatility modeling," *Scand. J. Stat.*, vol. 24, pp. 1–13, 1997.
[8] R. Jenssen, T. A. Øigård, T. Eltoft, and A. Hanssen, "Sparse code shrinkage using the normal inverse Gaussian density model," in *Proc. ICA*, San Diego, CA, Dec. 2001.
[9] D. N. Anderson, "A multivariate Linnik distribution," *Stat. Prob. Lett.*, vol. 14, pp. 333–336, 1992.
[10] T. A. Øigård, A. Hanssen, R. E. Hansen, and F. Godtliebsen, "EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse Gaussian distribution," *Signal Process.*, vol. 85, no. 8, pp. 1655–1673, 2005.
[11] A. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. R. Stat. Soc.*, ser. B, vol. 36, no. 1, pp. 99–102, 1974.
[12] J. A. Bilmes, "Factored sparse inverse covariance matrices," in *Proc. ICASSP*, Istanbul, Turkey, Jun. 2000.