AN UNBIASED RISK ESTIMATOR FOR GAUSSIAN MIXTURE NOISE DISTRIBUTIONS — APPLICATION TO SPEECH DENOISING

Jishnu Sadasivan[1] *and Chandra Sekhar Seelamantula*[2]

[1]Department of Electrical Communication Engineering, [2]Department of Electrical Engineering
Indian Institute of Science, Bangalore 560012, India
Emails: jishnus@ece.iisc.ernet.in, chandra.sekhar@ieee.org

ABSTRACT

We develop an unbiased estimate of mean-squared error (MSE), where the observations are assumed to be drawn from a Gaussian mixture (GM) distribution. Stein's unbiased risk estimate (SURE) is an unbiased estimate of the MSE, and was originally proposed for independent and identically distributed (i.i.d.) multivariate Gaussian observations. Subsequently, it was extended to the exponential family of distributions. In this paper, we extend the idea of SURE to observations drawn from a Gaussian mixture distribution (GMD). Since Gaussian mixture models (GMM) can model any given distribution sufficiently accurately, this generalized framework allows us to apply the SURE technique to the observations drawn from an arbitrary distribution. As an application, we consider the problem of denoising speech corrupted by a GM distributed noise. It is observed that the denoising performance of the algorithm developed using SURE based on GMD is superior in terms of the signal-to-noise ratio (SNR) and average segmental SNR (ASSNR), compared with that obtained using SURE based on the single Gaussian assumption.

Index Terms— Stein's unbiased risk estimate (SURE), Mean-squared error, Gaussian mixture model, Stein's lemma, Speech enhancement.

1. INTRODUCTION

In a statistical estimation framework, the parameter to be estimated is obtained by minimizing a cost function (risk), which measures the proximity between the actual parameter and its estimate. The actual risk such as mean-squared error (MSE) (which is the expected value of squared error between the actual parameter and its estimate) is a function of the unknown parameter or its statistics, which is difficult to get in real-world problems. An alternative approach is to assume that the parameter to be estimated is deterministic and use a risk estimation framework. In a risk estimation framework, instead of minimizing the original risk, an unbiased estimate of the risk, which is a function of the observation is minimized to obtain the unknown parameter. Stein's unbiased risk estimate (SURE) is an unbiased estimate of the MSE originally proposed in [1]; it assumes that the observations are independent and identically distributed (i.i.d.) (Gaussian). Stein further showed that the shrinkage type estimator of the mean of an i.i.d. multivariate Gaussian, that minimizes SURE dominates the classical least-squares estimate, if the observation dimension is greater than or equal to three [2]. The main idea in SURE theory is to replace the original risk by an unbiased estimate that depends only on the observations. Recently, many image and speech denoising applications optimizing SURE have been proposed [3-25]. The original formulation of SURE based on Gaussian i.i.d. assumption was extended in [26] to certain distributions of the continuous exponential family. The discrete exponential case is discussed in [27]. Both [26] and [27] are based on the assumption of i.i.d. observations. SURE for any non-i.i.d multivariate distribution belonging to the exponential family was recently proposed in [19]. Even though the SURE formulation allows us to directly approximate the MSE based on an estimate obtained from the observations, existing formulations are specific to a particular class of probability density function (p.d.f.) of the observations. This limits the applicability of the SURE technique to the case where the observations come from a distribution other than a member of the exponential family. This precludes application of the SURE theory in many practical cases. Since Gaussian mixture model (GMM) can model any p.d.f. with finite number of discontinuities sufficiently closely [31] [32], we can circumvent the problem of the observations following arbitrary p.d.f.s by modeling them using a GMM and develop an unbiased estimate of the MSE based on the Gaussian mixture distribution (GMD). In this paper, we derive an unbiased estimate of MSE assuming that the observations are drawn from a GMD. This generalization of SURE framework for GMD (GMD-SURE) allows us to apply SURE-based techniques to any given distribution. Moreover, the proposed formulation does not assume any particular structure on the estimator, and enables us to handle a wide class of estimators. To illustrate its usefulness, we compare the performance of speech denoising, based on SURE and GMD-SURE, where noise follows a Gaussian mixture (GM) distribution. A point-wise shrinkage estimator in discrete cosine transform (DCT) domain is used as the denoising function. In Section 2, SURE for GMD is formulated. In Section 3, we show how to estimate the parameters of the point-wise denoising function based on the theory developed in Section 2.

2. SURE FORMULATION FOR GMD

Let observation $\mathbf{x} \in \mathbf{R}^n$ be distributed according to a GMD

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m \mathcal{N}(\mathbf{x}; \theta_m + \mathbf{s}, \mathbf{C}_m), \tag{1}$$

where M is the number of Gaussian components in the mixture, α_m s are the mixture weights and $\mathcal{N}(\mathbf{w}; \theta_m + \mathbf{s}, \mathbf{C}_m)$ represents multivariate Gaussian distribution with covariance matrix $\mathbf{C}_m$ and mean $\theta_m + \mathbf{s}$. The goal is to obtain an estimate $\mathbf{h}(\mathbf{x})$, of the (non-random) parameter $\mathbf{s} \in \mathbf{R}^n$, from the observation $\mathbf{x}$ that minimizes the MSE:

$$\begin{aligned} \mathcal{R} &= \mathcal{E}\left\{\|\mathbf{s} - \mathbf{h}(\mathbf{x})\|^2\right\} \\ &= \mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{s}\right\} - 2\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\} + \mathcal{E}\left\{\mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\}. \end{aligned} \tag{2}$$

We observe that, in the preceding cost function, the first term is not a function of $\mathbf{h}(\mathbf{x})$, and hence does not affect the minimization of $\mathcal{R}$. Instead of minimizing $\mathcal{R}$, one could minimize the function

$$\mathcal{J}(\mathbf{s}, \mathbf{h}(\mathbf{x})) = -2\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\} + \mathcal{E}\left\{\mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\}. \tag{3}$$

Direct minimization of the above function results in an unrealizable estimator, which is a function of the unknown parameter to be estimated. An alternative to this approach is the SURE-based approach. The key idea of SURE-based estimation is that instead of minimizing the risk $\mathcal{J}$ directly, one minimizes an unbiased estimate $\widehat{\mathcal{J}}$, of $\mathcal{J}$, to find the optimum estimator $\mathbf{h}$. In order to find an unbiased estimate of $\mathcal{J}$, one requirement is an unbiased estimate of the term $\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\}$. Let $\mathcal{V}(\mathbf{x})$ be a function of the observation $\mathbf{x}$ such that it is an unbiased estimate of $\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\}$, that is, $\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\} = \mathcal{E}\left\{\mathcal{V}(\mathbf{h}(\mathbf{x}))\right\}$, then the unbiased estimate of $\mathcal{J}$ becomes

$$\widehat{\mathcal{J}} = -2\mathcal{V}(\mathbf{h}(\mathbf{x})) + \mathbf{h}(\mathbf{x})^{\mathsf{T}}\mathbf{h}(\mathbf{x}). \tag{4}$$

The unbiased estimate $\mathcal{V}(\mathbf{h}(\mathbf{x}))$ depends on the p.d.f. of the observations. In this paper, we develop a method to determine $\widehat{\mathcal{J}}$.

Theorem: Let $\mathbf{x}$ denote a random vector with probability density function given by (1), and let $\mathbf{u}_m = \mathbf{C}_m^{-1}\mathbf{x}$. Let $\mathbf{h}(\mathbf{u}_m)$ be an arbitrary function of $\mathbf{u}_m$ that is weakly differentiable in $\mathbf{u}_m$ and such that $\mathcal{E}\left\{|h_i(\mathbf{u}_m)|\right\}$ is bounded, then $\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\} = \mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\left\{\mathcal{T}(\mathbf{u}_1, \ldots, \mathbf{u}_M)\right\}$, where,

$$\mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\left\{\mathcal{T}(\mathbf{u}_1, \ldots, \mathbf{u}_M)\right\} = -\sum_{m=1}^{M}\alpha_m\left(\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^{\mathsf{T}}\theta_m\right\}\right.$$
$$\left.+\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^{\mathsf{T}}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\} + \mathcal{E}_{\mathbf{u_m}}\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C}_m\mathbf{u}_m)}{\partial \mathbf{u}_m}\right)\right\}\right),$$

and an unbiased estimate of $\mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\left\{\mathcal{T}(\mathbf{u}_1, \ldots, \mathbf{u}_M)\right\}$ becomes

$$\mathcal{T}(\mathbf{u}_1, \ldots, \mathbf{u}_M) = -\sum_{m=1}^{M}\alpha_m\left(\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^{\mathsf{T}}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}\right.$$
$$\left.+\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C}_m\mathbf{u}_m)}{\partial \mathbf{u}_m}\right)\right\} + \left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^{\mathsf{T}}\theta_m\right\}\right),$$

where $q(\mathbf{u}_m) = \dfrac{1}{(2\pi)^{n/2}|\mathbf{C}_m^{-1}|^{1/2}}\exp\left\{-\dfrac{1}{2}\mathbf{u}_\mathbf{m}^{\mathsf{T}}\mathbf{C}_\mathbf{m}\mathbf{u}_\mathbf{m}\right\}$, and $h_i(\mathbf{u}_m)$ is i^{th} component of $\mathbf{h}(\mathbf{u}_m)$.

Proof: Using the p.d.f. of $\mathbf{x}$ given in (1)

$$\begin{aligned}\mathcal{E}\left\{\mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\right\} &= \sum_{m=1}^{M}\int \mathbf{s}^{\mathsf{T}}\mathbf{h}(\mathbf{x})\alpha_m\mathcal{N}(\mathbf{x}; \theta_m + \mathbf{s}, \mathbf{C}_m)\,\mathrm{d}\mathbf{x}, \qquad (5)\\ &= \sum_{m=1}^{M}\sum_{i=1}^{n}\int h_i(\mathbf{x})s_i\alpha_m\mathcal{N}(\mathbf{x}; \theta_m + \mathbf{s}, \mathbf{C}_m)\,\mathrm{d}\mathbf{x},\end{aligned}$$

where $h_i(\mathbf{x})$ and s_i are the i^{th}components in the vector $\mathbf{h}(\mathbf{x}) \in \mathbf{R}^n$ and $\mathbf{s} \in \mathbf{R}^n$, respectively. The expectation of the i^{th} component is given by

$$\mathcal{E}\left\{h_i(\mathbf{x})s_i\right\} = \sum_{m=1}^{M}\int h_i(\mathbf{x})s_i\alpha_m\mathcal{N}(\mathbf{x}; \theta_m + \mathbf{s}, \mathbf{C}_m)\,\mathrm{d}\mathbf{x}. \tag{6}$$

Considering the contribution of the m^{th} Gaussian component separately (in the mixture model (1)), we get that

$$\mathcal{E}_{\mathbf{x},\mathbf{m}}\left\{h_i(\mathbf{x})s_i\right\} = \int h_i(\mathbf{x})s_i\mathcal{N}(\mathbf{x}; \theta_m + \mathbf{s}, \mathbf{C}_m)\,\mathrm{d}\mathbf{x}. \tag{7}$$

Substituting $\mathbf{u}_m = \mathbf{C}_m^{-1}\mathbf{x}$ in (7), yields

$$\begin{aligned}\mathcal{E}_{\mathbf{x},\mathbf{m}}\left\{h_i(\mathbf{x})s_i\right\} = \int s_i h_i(\mathbf{C}_m\mathbf{u}_m)\,q(\mathbf{u}_m) \qquad (8)\\ \exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}\mathrm{d}\mathbf{u}_m,\end{aligned}$$

where $q(\mathbf{u}_m) = \dfrac{1}{(2\pi)^{n/2}|\mathbf{C}_m^{-1}|^{1/2}}\exp\left\{-\dfrac{1}{2}\mathbf{u}_\mathbf{m}^{\mathsf{T}}\mathbf{C}_\mathbf{m}\mathbf{u}_\mathbf{m}\right\}$, and $\mathbf{g}(\theta_m + \mathbf{s}) = \left\{\dfrac{1}{2}(\theta_m + \mathbf{s})^{\mathsf{T}}\mathbf{C}_\mathbf{m}^{-1}(\theta_m + \mathbf{s})\right\}$. Note that we have

$$\begin{aligned}s_i\exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\} \qquad (9)\\ = \frac{\partial \exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}}{\partial u_{m,i}}\\ -\theta_{m,i}\exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\},\end{aligned}$$

where $u_{m,i}$ and $\theta_{m,i}$ are the i^{th} components of the vectors $\mathbf{u}_m$ and θ_m, respectively. We have

$$\begin{aligned}\mathcal{E}_{\mathbf{u_m}}\left\{h_i(\mathbf{C}_m\mathbf{u}_m)\theta_{m,i}\right\} \triangleq \int h_i(\mathbf{C}_m\mathbf{u}_m)\theta_{m,i}\mathbf{q}(\mathbf{u}_m) \qquad (10)\\ \exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}\mathrm{d}\mathbf{u}_m,\end{aligned}$$

and using integration by parts, we arrive at

$$\begin{aligned}\int_{-\infty}^{+\infty} h_i(\mathbf{C}_m\mathbf{u}_m)q(\mathbf{u}_m)\frac{\partial \exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}}{\partial u_{m,i}}\mathrm{d}u_{m,i} \qquad (11)\\ = -\int_{-\infty}^{+\infty}\frac{\partial h_i(\mathbf{C}_m\mathbf{u}_m)q(\mathbf{u}_m)}{\partial u_{m,i}}\exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}\mathrm{d}u_{m,i},\end{aligned}$$

where

$$\frac{\partial h_i(\mathbf{C}_m\mathbf{u}_m)q(\mathbf{u}_m)}{\partial u_{m,i}} = h_i(\mathbf{C}_m\mathbf{u}_m)\frac{\partial q(\mathbf{u}_m)}{\partial u_{m,i}} + q(\mathbf{u}_m)\frac{\partial h_i(\mathbf{C}_m\mathbf{u}_m)}{\partial u_{m,i}}.$$

We have assumed that $h_i(\mathbf{C}_m\mathbf{u}_m)$ is weakly differentiable and $\mathcal{E}\{|h_i(\mathbf{C}_m\mathbf{u}_m)|\}$ is bounded. Since $\mathcal{E}\{|h_i(\mathbf{C}_m\mathbf{u}_m)|\}$ is bounded, in order to arrive at the RHS of (11), we have used the property that

$$\lim_{|u_{m,i}|\to\infty}\left|h_i(\mathbf{C}_m\mathbf{u}_m)q(\mathbf{u}_m)\exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}\right| = 0.$$

Substituting (9) in (8), and using (10) and (11), we arrive at

$$\begin{aligned}\int s_i h_i(\mathbf{C}_m\mathbf{u}_m)q(\mathbf{u}_m)\exp\left\{\mathbf{u}_m^{\mathsf{T}}(\theta_\mathbf{m} + \mathbf{s}) - \mathbf{g}(\theta_m + \mathbf{s})\right\}\mathrm{d}\mathbf{u}_m \qquad (12)\\ = -\left(\mathcal{E}_{\mathbf{u_m}}\left\{h_i(\mathbf{C}_m\mathbf{u}_m)\theta_{m,i}\right\} + \mathcal{E}_{\mathbf{u_m}}\left\{\frac{\partial h_i(\mathbf{C}_m\mathbf{u}_m)}{\partial u_{m,i}}\right\}\right.\\ \left.+\mathcal{E}_{\mathbf{u_m}}\left\{h_i(\mathbf{C}_m\mathbf{u}_m)\frac{\partial \ln q(\mathbf{u}_m)}{\partial u_{m,i}}\right\}\right).\end{aligned}$$

Using (12), we express

$$\mathcal{E}_{\mathbf{x},\mathbf{m}}\left\{\mathbf{s}^\mathsf{T}\mathbf{h}(\mathbf{x})\right\} = -\sum_{i=1}^{n}\left(\mathcal{E}_{\mathbf{u_m}}\left\{h_i(\mathbf{C}_m\mathbf{u}_m)\theta_{m,i}\right\}\right. \tag{13}$$
$$\left. +\mathcal{E}_{\mathbf{u_m}}\left\{\frac{\partial h_i(\mathbf{C}_m\mathbf{u}_m)}{\partial u_{m,i}}\right\} + \mathcal{E}_{\mathbf{u_m}}\left\{h_i(\mathbf{C}_m\mathbf{u}_m)\frac{\partial \ln(\mathbf{u}_m)}{\partial u_{m,i}}\right\}\right)$$
$$= -\left(\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}\right.$$
$$\left. +\mathcal{E}_{\mathbf{u_m}}\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C}_m\mathbf{u}_m)}{\partial \mathbf{u}_m}\right)\right\} + \mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\theta_m\right\}\right).$$

Substituting (13) in (5), we get

$$\mathcal{E}\left\{\mathbf{s}^\mathsf{T}\mathbf{h}(\mathbf{x})\right\} = -\sum_{m=1}^{M}\alpha_m\left(\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}\right. \tag{14}$$
$$\left. +\mathcal{E}_{\mathbf{u_m}}\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C}_m\mathbf{u}_m)}{\partial \mathbf{u}_m}\right)\right\} + \mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\theta_m\right\}\right).$$

We can write, $\mathcal{E}\left\{\mathbf{s}^\mathsf{T}\mathbf{h}(\mathbf{x})\right\} = \mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\left\{\mathcal{T}(\mathbf{u}_1,\ldots,\mathbf{u}_M)\right\}$, where

$$\mathcal{T}(\mathbf{u}_1,\ldots,\mathbf{u}_M) = -\sum_{m=1}^{M}\alpha_m\left(\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}\right. \tag{15}$$
$$\left. +\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C}_m\mathbf{u}_m)}{\partial \mathbf{u}_m}\right)\right\} + \left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\theta_m\right\}\right).$$

One can observe that $\mathcal{T}(\mathbf{u}_1,\ldots,\mathbf{u}_M)$ is an unbiased estimate of $\mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\left\{\mathcal{T}(\mathbf{u}_1,\ldots,\mathbf{u}_M)\right\}$, where the expectation is with respect to the random variables $\mathbf{u}_1,\ldots,\mathbf{u}_M$, which are functions of $\mathbf{x}$. Thus, an unbiased estimate of $\mathcal{E}\left\{\mathbf{s}^\mathsf{T}\mathbf{h}(\mathbf{x})\right\}$, after the variable transformation is obtained as $\mathcal{T}(\mathbf{u}_1,\ldots,\mathbf{u}_M)$. Thus, the proof is complete. Using the substitution $\mathbf{u}_m = \mathbf{C}_m^{-1}\mathbf{x}$, we obtain

$$\mathcal{E}\left\{\mathbf{h}(\mathbf{x})^\mathsf{T}\mathbf{h}(\mathbf{x})\right\} = \sum_{m=1}^{M}\alpha_m\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)\right\}. \tag{16}$$

The risk (3) is simplified as

$$\mathcal{J}(\mathbf{s},\mathbf{h}(\mathbf{x})) = \sum_{m=1}^{M}\alpha_m\left(2\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}\right. \tag{17}$$
$$+2\mathcal{E}_{\mathbf{u_m}}\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C_m}\mathbf{u_m})}{\partial \mathbf{u}_m}\right)\right\} + 2\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\theta_m\right\}$$
$$\left. +\mathcal{E}_{\mathbf{u_m}}\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)\right\}\right).$$

From (17), an unbiased estimate of $\mathcal{J}(\mathbf{s},\mathbf{h}(\mathbf{x}))$ is obtained as

$$\widehat{\mathcal{J}}(\mathbf{h}(\mathbf{x})) = \sum_{m=1}^{M}\alpha_m\left(2\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\theta_m\right\}\right. \tag{18}$$
$$+2\left\{\mathrm{Tr}\left(\frac{\partial \mathbf{h}(\mathbf{C_m}\mathbf{u_m})}{\partial \mathbf{u}_m}\right)\right\} + 2\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\frac{\partial \ln q(\mathbf{u}_m)}{\partial \mathbf{u}_m}\right\}$$
$$\left. +\left\{\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)^\mathsf{T}\mathbf{h}(\mathbf{C}_m\mathbf{u}_m)\right\}\right),$$

that is, $\mathcal{J} = \mathcal{E}_{\mathbf{u_1}\ldots\mathbf{u_M}}\{\widehat{\mathcal{J}}\}$. When the observation distribution is a GMD as given in (1), $\widehat{\mathcal{J}}$ is an unbiased estimate of $\mathcal{J}$, and to obtain the optimum $\mathbf{h}$, we minimize (18) with respect to $\mathbf{h}$. The derivation of $\widehat{\mathcal{J}}$ does not assume any specific form on $\mathbf{h}$. Hence, it is possible to use GMD-SURE to estimate parameters for different classes of $\mathbf{h}$.

3. APPLICATION TO SPEECH DENOISING

We assume an additive noise model: $\bar{\mathbf{x}} = \bar{\mathbf{s}} + \bar{\mathbf{w}}$ where $\bar{\mathbf{s}} \in \mathbf{R}^n$ is the clean signal parameter vector and $\bar{\mathbf{w}} \in \mathbf{R}^n$ is the noise vector with distribution $f_{\bar{\mathbf{w}}}(\bar{\mathbf{w}}) = \sum_{m=1}^{M}\alpha_m\mathcal{N}\left(\bar{\mathbf{w}};\bar{\theta}_m,\bar{\mathbf{C}}_m\right)$. As a consequence of the additive model, $\bar{\mathbf{x}}$ has the distribution $f_{\bar{\mathbf{x}}}(\bar{\mathbf{x}}) = \sum_{m=1}^{M}\alpha_m\mathcal{N}\left(\bar{\mathbf{x}};\bar{\theta}_m+\bar{\mathbf{s}},\bar{\mathbf{C}}_m\right)$. We perform denoising in the DCT domain [22]. Let $\mathbf{D}$ denote the DCT; $\mathbf{x} = \mathbf{D}\bar{\mathbf{x}}$ is the noisy speech DCT vector and follows the additive model $\mathbf{x} = \mathbf{s} + \mathbf{w}$, where $\mathbf{s} \in \mathbf{R}^n$ and $\mathbf{w} \in \mathbf{R}^n$ are clean signal and noise DCT vectors, respectively. The distribution of $\mathbf{x}$ is a GMD, $f_{\mathbf{x}}(\mathbf{x}) = \sum_{m=1}^{M}\alpha_m\mathcal{N}(\mathbf{x};\theta_m+\mathbf{s},\mathbf{C}_m)$ where $\theta_m = \mathbf{D}\bar{\theta}_m$, $\mathbf{s} = \mathbf{D}\bar{\mathbf{s}}$ and $\mathbf{C}_m = \mathbf{D}\bar{\mathbf{C}}_m\mathbf{D}^\mathsf{T}$. Our goal is to estimate $\mathbf{s} \in \mathbf{R}^n$ from the noisy observations $\mathbf{x} \in \mathbf{R}^n$ such that the estimate results in the minimum value of the MSE.

We consider a point-wise shrinkage estimator for denoising, that is $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = \mathbf{diag}(a_1,a_2,\cdots,a_n)$ [22]. The optimum shrinkage parameter $\mathbf{A}$ is obtained by the minimization of $\widehat{\mathcal{J}}$ (18), which is an unbiased estimate of $\mathcal{J}$ (17), ie $\tilde{\mathbf{A}} = \arg\min_{\mathbf{A}}\widehat{\mathcal{J}}(\mathbf{A}\mathbf{x})$. The clean signal DCT vector estimate is $\widehat{\mathbf{s}} = \tilde{\mathbf{A}}\mathbf{x}$, where $\tilde{\mathbf{A}} = \mathbf{diag}(\tilde{a}_1,\tilde{a}_2,\cdots,\tilde{a}_n)$. To obtain the optimum shrinkage parameter $\tilde{a}_i$, we solve $\frac{\partial\widehat{\mathcal{J}}}{\partial a_i} = 0$, which results in

$$\tilde{a}_i = \frac{\sum_{j=1}^{M}\alpha_j\left[(\mathbf{C}_{i,\mathbf{j}}\mathbf{u_j})^2 - \theta_{\mathbf{j},i}(\mathbf{C}_{i,\mathbf{j}}\mathbf{u_j}) - c_{i,i,\mathbf{j}}\right]}{\sum_{m=1}^{M}\alpha_m(\mathbf{C}_{i,\mathbf{m}}\mathbf{u_m})^2}, \tag{19}$$

where M is the number of mixture components, α_j is the mixture weight of j^{th} Gaussian, $\mathbf{C}_{i,\mathbf{j}}$ is the i^{th} row of the covariance matrix $\mathbf{C_j}$ of j^{th} Gaussian, $c_{i,i,\mathbf{j}}$ is the $(i,i)^{th}$ element of $\mathbf{C_j}$ and $\theta_{\mathbf{j},i}$ is the i^{th} component of the mean vector $\theta_\mathbf{j}$ of j^{th} Gaussian. It can be easily seen that if $f_{\bar{\mathbf{w}}}(\bar{\mathbf{w}})$ is zero mean i.i.d. Gaussian, then the estimator (19) reduces to the standard James-Stein estimator [22]. To illustrate the performance differences of SURE and GMD-SURE, we consider the application of speech denoising. A synthetic noise vector distributed according to GMD is generated and added to the clean speech signal vector to generate the noisy signal. In our simulations, we consider frame-wise processing of the speech signal, where each frame has length 20 ms (160 samples, at 8 kHz sampling frequency) and apply the optimum point-wise shrinkage in DCT domain to achieve denoising.

4. EXPERIMENTAL RESULTS

The GMD-SURE formulation proposed in this paper is applicable to real-world noise distributions that can be approximated by a GMM. Also, since no specific form for $\mathbf{h}$ is assumed, parameter estimation for a wide class of estimators is possible. To generate the noisy speech, add clean speech with a synthetic GM noise (with number of Gaussian components equal to 3, component weights [0.3 , 0.4 , 0.3], covariance matrices [$0.01 \times \mathbf{I_n}$, $0.02 \times \mathbf{I_n}$, $0.04 \times \mathbf{I_n}$] and

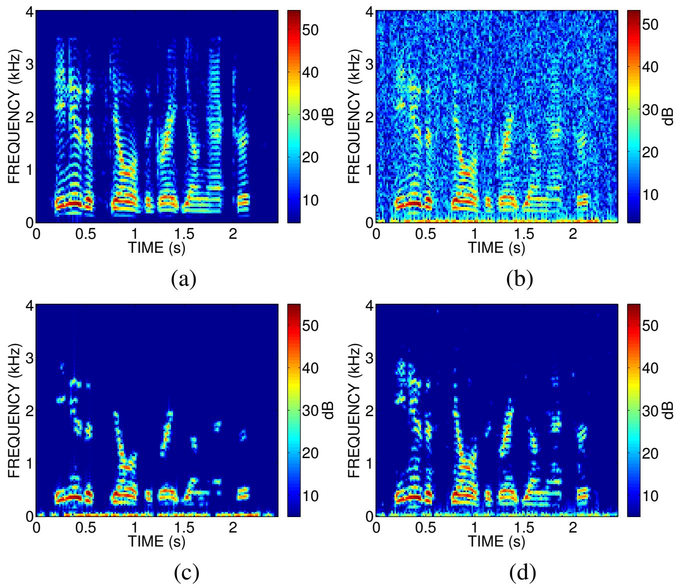

Fig. 1. (Color online) Spectrograms of (a) Clean speech, (b) Noisy speech corrupted by GMD noise, (c) Denoised speech using SURE (single Gaussian), and (d) Denoised speech using GMD-SURE.

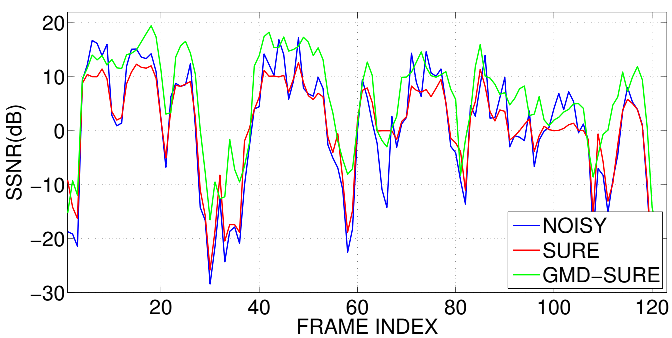

Fig. 2. (Color online) SSNR plot of enhanced speech using SURE and GMD-SURE.

component means $[0.1 \times \mathbf{1_n}, -0.1 \times \mathbf{1_n}, 0.4 \times \mathbf{1_n}]$, where $\mathbf{I_n}$ is $n \times n$ identity matrix, $\mathbf{1_n}$ is an $n \times 1$ vector with all entries equal to unity). The noisy speech, at input SNR of 5 dB, and segmental SNR (SSNR) of 0.57 dB, is denoised using the point-wise shrinkage estimator in the DCT domain. For input SNR of 5 dB, speech denoising based on SURE (single Gaussian assumption) yielded an output SNR=5.76 dB, and SSNR=0.76 dB, whereas GMD-SURE yielded an output SNR=$\mathbf{11.69}$ dB, and SSNR=$\mathbf{5.87}$ dB. Figures 1 and 2 contain the spectrogram plots and SSNR plots of noisy speech and denoised speech using SURE and GMD-SURE. We observe from the spectrograms that GMD-SURE based denoising has higher noise attenuation while maintaining the speech distortion is low compared with the standard SURE approach. In the spectrogram (color plot) of standard SURE-based denoising, high residual noise is present in the low-frequency region (near zero), but it is absent in GMD-SURE based denoising. SSNR plots show approximately 4 dB improvement for GMD-SURE based technique over the SURE-based one in all the frames. Figure 3 shows a comparison of the denoising performance of SURE and GMD-SURE for different input SNR values, where noise is GM distributed with the same parameters used in Figure 1. It is observed that SNR and ASSNR improvement is high for the case of GMD-SURE, compared to SURE. Since noise is distributed according to GMD, performance degradation in standard SURE-based approach reveals the importance of unbiased risk estimator based on the correct noise p.d.f.. Performance degradation in ordinary SURE-based approach is due to the model mismatch of the noise p.d.f.. Better modeling of the p.d.f. of the observations and unbiased estimator based on this p.d.f. leads to a better estimate of the MSE. Since we minimize the same cost function, and the form of the estimator is the same in both standard SURE and GMD-SURE, the SNR improvement obtained is due to a more accurate estimate of the MSE in the case of GMD-SURE. We considered speech denoising to illustrate the effectiveness of GMD-SURE over the standard SURE approach, when noise distribution is not Gaussian. The denoising framework proposed in this paper is also applicable to other signals. Even though other performance metrics are available to assess the quality of the denoised speech, we limited our evaluation to the widely used SNR-based metrics [28–30], because we are concerned about the estimate of MSE, which is an inverse function of SNR. Hence, SNR improvement is directly attributed to the improvement in the MSE estimate.

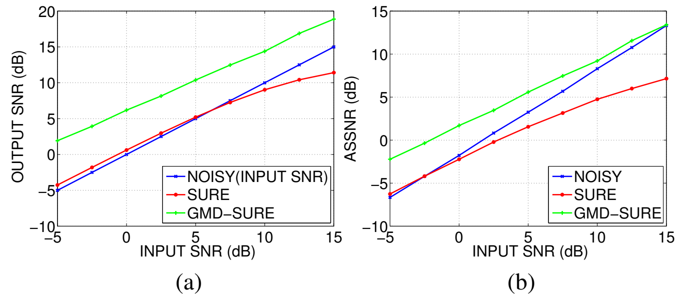

Fig. 3. (Color online) Comparison of denoising performance of SURE and GMD-SURE for different input SNR, where noise is GM distributed (a) Output SNR, (b) ASSNR averaged over 100 independent realizations of noise.

5. CONCLUSIONS

We developed an unbiased estimator of the MSE where the observations are distributed according to a GMD. SURE, which is an unbiased estimator originally proposed for i.i.d Gaussian noise and later extended to specific class of p.d.f., cannot be applied when the observations follow a different p.d.f.. Since a GMM can model any given p.d.f. sufficiently closely, an unbiased estimator of MSE for GMD gives generalizability to the SURE framework and allows us to apply it to any given p.d.f.. The proposed formulation does not assume any specific structure on the signal estimator. Hence, GMD-SURE can be used to obtain parameters for a wide class of estimators. As an application, we considered denoising of speech corrupted by an additive GM distributed synthetic noise. The simulations showed that denoising using point-wise linear weighting function obtained through GMD-SURE gave better denoising compared with ordinary SURE (single Gaussian) in terms of SNR and ASSNR. From the superior performance of GMD-SURE over standard SURE, we infer that the MSE estimate is more accurate in case of the former.

6. REFERENCES

[1] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135−1151, Nov. 1981.

[2] W. James and C. Stein, "Estimation with quadratic loss," *Berkeley Symp. Math. Statist. and Prob.*, Berkeley, CA, vol.1, pp. 361−379, 1961.

[3] 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.

[4] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200−1224, Dec. 1995.

[5] A. Benazza-Benyahia and J. C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814−1830, Nov. 2005.

[6] J. C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616−4632, Dec. 2009.

[7] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407−422, May 2011.

[8] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Information Theory*, vol. 45, no. 7, pp. 2225−2238, Nov. 1999.

[9] S. Ramani, T. Blu, and M. Unser, "Monte Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540−1544, Sep. 2008.

[10] D. V. De Ville and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Letters*, vol. 16, no. 11, pp. 973−976, Nov. 2009.

[11] D. V. De Ville and M. Kocher, "Nonlocal means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2683−2690, Sep. 2011.

[12] X. Zhang and M. D. Desai, "Adapting denoising based on SURE risk," *IEEE Signal Process. Letters*, vol. 5, no. 10, pp. 265−267, Oct. 1998.

[13] H. Huang and T. C. M. Lee, "Data adaptive median filters for signal and image denoising using a generalized SURE criterion," *IEEE Signal Process. Letters*, vol. 13, no. 9, pp. 561−564, Sep. 2006.

[14] T. Hsung, D. Lun and K. C. Ho, "Optimizing the multiwavelet shrinkage denoising," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 240−251, Jan. 2005.

[15] T. Qiu, A. Wang, N. Yu, and A. Song, "LLSURE: Local linear SURE-based edge-preserving image filtering," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 80−90, Jan. 2013.

[16] F. Luisier, T. Blu, and M. Unser, "SURE-LET for orthonormal wavelet-domain video denoising," *IEEE Trans. on Circ. and Syst. for Video Tech.*, vol. 20, no. 6, pp. 913−919, Jun. 2010.

[17] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778−2786, Nov. 2007.

[18] F. Luisier and T. Blu, "The SURE-LET multichannel image denoising: interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 482−492, Apr. 2008.

[19] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471−481, Feb. 2009.

[20] A. M. Atto, D. Pastor, and G. Mercier, "Smooth adaptation by sigmoid shrinkage," *EURASIP Journal on Image and Video Process.*, 2009 : 532312, pp. 1−16, 2009.

[21] A. M. Atto, D. Pastor, and G. Mercier, "Optimal sure parameter for sigmoidal wavelet shrinkage," *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Glasgow, Scotland, pp. 40−43, Aug. 2009.

[22] N. R. Muraka and C. S. Seelamantula, "A risk-estimation-based comparison of mean-square error and Itakura-Saito distortion measures for speech enhancement," in *Proc. Interspeech*, Florence, Italy, pp. 349−352, Aug. 2011.

[23] S. R. Krishnan and C. S. Seelamantula, "A generalized Stein's estimation approach for speech enhancement based on perceptual criteria," in *Proc. Workshop on Statistical and Perceptual Audition (SAPA)–Speech Communication with Adaptive Learning (SCALE)*, Portland, Sep. 7−8, 2012.

[24] S. R. Krishnan and C. S. Seelamantula, "On the selection of optimum Savitzky-Golay filters," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 380−391, Jan. 2013.

[25] N. Zheng, X. Li, T. Blu, and T. Lee, "SURE-MSE speech enhancement for robust speech recognition," in *Proc. 7^{th} Int. Symp. Chinese Spoken Language Process.*, pp. 271−274, Shenzhen, China, Nov. 2010.

[26] H. M. Hudson, "A natural identity for exponential families with applications in multiparameter estimation," *Ann. Stat.*, vol.6, no.3, pp. 4730−484, 1978.

[27] J. T. Hwang, "Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases," *Ann. Stat.*, vol.10, no.3, pp. 857−867, 1982.

[28] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 6, pp. 1741−1752, Aug. 2007.

[29] P. Mowlaee and R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement," *IEEE Signal Process. letters*, vol. 20, no. 12, pp. 1235−1239, Dec. 2013.

[30] P. C. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech, Audio Process.*, vol. 13, no.5, pp. 857−869, Sep. 2005.

[31] H. W. Sorenson and D. L. Alspach. "Recursive Bayesian estimation using Gaussian sums," *Automatica*, Vol 7, pp. 465−479, 1971.

[32] K. N. Plataniotis and D. Hatzinakos. *Gaussian Mixtures and Their applications to Signal Processing. Advanced Signal Processing: Theory and Implementation for Radar, Sonar and Medical Imaging Systems Handbook.* CRC Press, December 2000.