On MMSE estimation: A Linear Model under Gaussian Mixture Statistics

John T. Flåm, Saikat Chatterjee, Kimmo Kansanen, Torbjörn Ekman

Abstract—In a Bayesian linear model, suppose observation $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$ stems from independent inputs $\mathbf{x}$ and $\mathbf{n}$ which are Gaussian mixture (GM) distributed. With known matrix $\mathbf{H}$, the minimum mean square error (MMSE) estimator for $\mathbf{x}$, has analytical form. However, its performance measure, the MMSE itself, has no such closed form. Because existing Bayesian MMSE bounds prove to have limited practical value under these settings, we instead seek *analytical* bounds for the MMSE, both upper and lower. This paper provides such bounds, and relates them to the signal-to-noise-ratio (SNR).

I. INTRODUCTION

In estimation theory, an important instance is the linear model

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n}. \tag{1}$$

In this paper, $\mathbf{y}$ is a vector of observations, $\mathbf{H}$ is a known matrix, $\mathbf{x}$ is a vector to be estimated, and $\mathbf{n}$ is a vector of noise. We assume a Bayesian setting where the distributions on $\mathbf{x}$ and $\mathbf{n}$ are known a priori. Specifically, we posit $\mathbf{x}$ and $\mathbf{n}$ to be independent and Gaussian Mixture (GM) distributed. In order to estimate $\mathbf{x}$, we use the the minimum mean square error (MMSE) estimator

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}} \triangleq E\left\{\mathbf{x}|\mathbf{y}\right\}, \tag{2}$$

and inquire about its performance, measured by the MMSE:

$$E\left\{\|\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\|_2^2\right\}. \tag{3}$$

With GM inputs to (1), $\mathbf{u}_{\mathbf{x}|\mathbf{y}}$ becomes analytical, but in contrast to the familiar case of Gaussian inputs, the MMSE does not. Thus, while the optimal estimator is implementable, and used in practice, its performance is harder to assess. Moreover, in this setup, existing Bayesian bounds on the MMSE become no easier to estimate than the MMSE itself. Our objective is therefore to bound the MMSE *analytically*, both from above and below, and to relate these bounds to the signal-to-noise-ratio (SNR).

We find that the MMSE is *lower* bounded by the mean square error (MSE) of a 'genie-aided' estimator. This imaginary, non-implementable estimator knows precisely which source in the mixture is active, at any time, both for $\mathbf{x}$ and $\mathbf{n}$. It is therefore much better informed than the MMSE estimator. Yet, simulations indicate that the MMSE estimator is comparably accurate for SNRs above a quite modest threshold. The *upper* bound is provided by the MSE of the linear MMSE (LMMSE) estimator. In fact, we show that the LMMSE estimator becomes *the* MMSE estimator in two extreme cases: when the SNR is either zero or infinite. However, simulations indicate that the LMMSE estimator is nearly MSE optimal for a much larger range of SNR. The upper and lower bounds have closed form expressions, and are straightforward to calculate. Because they approach each other with increasing SNR, the MMSE can be determined rather accurately when the SNR is moderate to high.

John T. Flåm, Kimmo Kansanen and Torbjörn Ekman are all with the Department of Electronics and Telecommunications, NTNU-Norwegian University of Science and Technology, Trondheim, Norway. Emails: flam.john@gmail.com, flam@iet.ntnu.no, kimmo.kansanen@iet.ntnu.no and torbjorn.ekman@iet.ntnu.no. Saikat Chatterjee is with the Communication Theory Lab, School of Electrical Engineering, KTH-Royal Institute of Technology, Sweden. Email: saikatchatt@gmail.com, sach@kth.se.

The paper is organized as follows. Section II defines and motivates the input model and reviews parts of the related literature. Section III presents the main contribution: the upper and lower bounds of the MMSE. Section IV proves the MMSE bounds, and analyzes them in the asymptotic cases of zero and infinite SNR. Section V simulates how the MMSE evolves, between its bounds, as a function of SNR. Section VI concludes.

II. INPUT MODEL AND RELATED WORK

We assume $\mathbf{x}$ and $\mathbf{n}$ in (1) to be independent and GM distributed:

$$\mathbf{x} \sim \sum_{k \in \mathcal{K}} p_k \mathcal{N}\left(\mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)}\right), \qquad \mathbf{n} \sim \sum_{l \in \mathcal{L}} q_l \mathcal{N}\left(\mathbf{u}_{\mathbf{n}}^{(l)}, \mathbf{C}_{\mathbf{nn}}^{(l)}\right). \tag{4}$$

This notation should be read in the distributional sense. Thus, $\mathbf{x}$ results from a composite experiment. First, source $k \in \mathcal{K}$ is activated with probability $p_k \geq 0$, $\sum_{k \in \mathcal{K}} p_k = 1$. Second, that source generates a Gaussian signal with distribution law $\mathcal{N}(\mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)})$. The set $\left\{p_k, \mathbf{u}_{\mathbf{x}}^{(k)}, \mathbf{C}_{\mathbf{xx}}^{(k)}; k \in \mathcal{K}\right\}$ defines the GM *parameters* of $\mathbf{x}$. The noise $\mathbf{n}$ emerges in entirely similar, but independent manner. $\mathcal{K}$ and $\mathcal{L}$ are *finite* index sets. Their cardinalities determine the number of *components* in each mixture.

GM distributions can accommodate multi modality, asymmetry and heavy tails. In fact, with a sufficient number of components, and a judicious choice of parameters, a GM can approximate *any* distribution to desired accuracy [1], [2]. In practice, GM parameters are rarely given a priori. Most often they must be estimated; typically a non-trivial task [1], [3]. A common approach is to estimate GM parameters from training data. The expectation maximization (EM) algorithm [4]–[6] is well suited, and much used, for that purpose. Briefly, the algorithm relies on observations drawn from the distribution we wish to parametrize, and some initial estimate of the parameters. The observations are used to optimize the parameters, iteratively, until convergence to a local maximum of the likelihood function. Because the resulting GM parameters, and hence the GM distribution, depends on the initial estimates, the algorithm can alternatively be started from multiple initial estimates. This produces an ensemble of GM distributions, which can be averaged [7], [8]. For example, if there are $|\mathcal{M}|$ distributions in the ensemble, each with probability ξ_m, then $\mathbf{x}$ could be distributed as

$$\mathbf{x} \sim \sum_{m,k} \xi_m p_k \mathcal{N}\left(\mathbf{u}_{\mathbf{x}}^{(m,k)}, \mathbf{C}_{\mathbf{xx}}^{(m,k)}\right), \tag{5}$$

where $\mathbf{u}_{\mathbf{x}}^{(m,k)}$ and $\mathbf{C}_{\mathbf{xx}}^{(m,k)}$ are the k-th mean and k-th covariance, respectively, of GM distribution m. But apart from appropriate notational changes, (5) is a GM of the same format as (4). This emphasizes that notation (4) may also include uncertainty in the GM parameters. Our starting point is that (4) has resulted from such, or similar, model fitting. Above, we have compactly written $\sum_{m,k}$ for $\sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}}$, a convention we will continue to use.

Model (1), with (4) as input, closely matches the image restoration problem described in e.g. [9]. There, $\mathbf{x}$ represents an image patch to be estimated and $\mathbf{H}$ is typically a non-invertible degradation operator. Since the problem usually is underdetermined, good patch models become important. Modeling $\mathbf{x}$ by a GM and the noise $\mathbf{n}$ as Gaussian has proven to give good results. However, unlike [9] which proposes a maximum a posteriori EM algorithm to estimate both the GM parameters and $\mathbf{x}$, we assume that the GM parameters already have been estimated, and concentrate on the signal $\mathbf{x}$. Moreover, we focus on MMSE estimation, the primary objective being to provide analytical bounds for the non-analytical MMSE.

Mixture distributions are common, particularly within two groups of the engineering literature. The first one focuses on state sequence

estimation via noisy measurements, presuming statistics that fit the GM paradigm. Selected works include [2], [10]–[15]. These studies offer (approximate or exact) GM posterior state distributions, the mean always serving as the state estimate. For natural reasons, because the settings are not stationary, none of these works analyze the MMSE. The second group of studies uses GM distributions to simplify processing of speech, audio and images. Selected works include [16]–[27]. In these, one or more variates are modeled by finite GMs. This is often sufficiently accurate, and allows good practical estimators. However, none of these works investigate bounds on the MMSE.

The emphasis on *analytical* bounds separates this work from those offering general lower bounds on the MMSE. Included there, are the Bayesian bounds of Cramer-Rao [28], Bobrovsky-Zakai [29], Bhattacharyya [28], Weiss-Weinstein [30], and Reuven-Messer [31]. These bounds hold for most types of joint densities $f(\mathbf{x}, \mathbf{y})$ of practical interest but they rarely acquire analytical form. Yet, for most $f(\mathbf{x}, \mathbf{y})$ they are much simpler to evaluate numerically than the MMSE. That feature makes them both attractive and useful; for interested readers, [32] proposes a general class of Bayesian lower bounds. Like the MMSE, the above mentioned bounds all become non-analytical when (4) is input to (1), but in contrast to the usual case, they do *not* become simpler to evaluate numerically than the MMSE. For this reason, the mentioned bounds have limited practical value to the current problem. Such bounds, of analytical sort, are useful because they can provide a simple way to assess estimator performance. To our knowledge, the literature offers no explicit analysis on that account. This motivates us to explore and present both upper and lower bounds.

III. MAIN RESULT: BOUNDS ON THE MMSE

Proposition 1: Let observation $\mathbf{y}$ originate from (1), where $\mathbf{H}$ is a known matrix, and $\mathbf{x}$ and $\mathbf{n}$ are independent GM variates defined by (4). Let $\mathbf{u}_{\mathbf{x}|\mathbf{y}}$ in (2) denote the MMSE estimator for $\mathbf{x}$. Its performance, measured by $\epsilon^2 = E\left\{\|\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\|_2^2\right\}$, is bounded as follows:

$$\sum_{k,l} p_k q_l \text{Tr}\left(\mathbf{C}_{\mathbf{xx}}^{(k)} - \mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T \left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}^{(l)}\right)^{-1}\mathbf{H}\mathbf{C}_{\mathbf{xx}}^{(k)}\right)$$

$$\leq \epsilon^2 \leq \text{Tr}\left(\mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{xx}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1}\mathbf{H}\mathbf{C}_{\mathbf{xx}}\right).$$

Here $\text{Tr}(\cdot)$ denotes the trace operator, and

$$\mathbf{C}_{\mathbf{xx}} = \sum_k p_k \left(\mathbf{C}_{\mathbf{xx}}^{(k)} + \mathbf{u}_{\mathbf{x}}^{(k)}\mathbf{u}_{\mathbf{x}}^{(k)T}\right) - \mathbf{u}_{\mathbf{x}}\mathbf{u}_{\mathbf{x}}^T, \quad \mathbf{u}_{\mathbf{x}} = \sum_k p_k \mathbf{u}_{\mathbf{x}}^{(k)}, \quad (6)$$

are the covariance and mean of $\mathbf{x}$, respectively. Similarly

$$\mathbf{C}_{\mathbf{nn}} = \sum_l q_l \left(\mathbf{C}_{\mathbf{nn}}^{(l)} + \mathbf{u}_{\mathbf{n}}^{(l)}\mathbf{u}_{\mathbf{n}}^{(l)T}\right) - \mathbf{u}_{\mathbf{n}}\mathbf{u}_{\mathbf{n}}^T, \quad \mathbf{u}_{\mathbf{n}} = \sum_l q_l \mathbf{u}_{\mathbf{n}}^{(l)}, \quad (7)$$

are the covariance and mean of and $\mathbf{n}$, respectively. Proof is given in the next section.

IV. DERIVATION OF THE MMSE BOUNDS

Here, we first verify the non-analytical nature of the MMSE. Then we prove the bounds in Proposition 1. The MMSE is, by definition,

$$\epsilon^2 \triangleq E\left\{\|\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\|_2^2\right\}$$

$$= \iint \left(\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right)^T \left(\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right) f(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

$$= \iint \left(\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right)^T \left(\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right) f(\mathbf{x}|\mathbf{y}) d\mathbf{x} f(\mathbf{y}) d\mathbf{y}$$

$$= \int \text{Tr}\left(\mathbf{C}_{\mathbf{x}|\mathbf{y}}\right) f(\mathbf{y}) d\mathbf{y} \geq 0. \quad (8)$$

Under (1) and (4), it can be shown that $\mathbf{y}$ in (8) is a GM with $|\mathcal{K}| \cdot |\mathcal{L}|$ components. Its density has the form

$$f(\mathbf{y}) = \sum_{k,l} p_k q_l f^{(k,l)}(\mathbf{y}), \quad (9)$$

where $f^{(k,l)}(\mathbf{y})$ is a Gaussian density with mean and covariance

$$\mathbf{u}_{\mathbf{y}}^{(k,l)} = \mathbf{H}\mathbf{u}_{\mathbf{x}}^{(k)} + \mathbf{u}_{\mathbf{n}}^{(l)}, \qquad \mathbf{C}_{\mathbf{yy}}^{(k,l)} = \mathbf{H}\mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}^{(l)}, \quad (10)$$

respectively. Furthermore, it can be shown that the posterior covariance matrix in (8) is

$$\mathbf{C}_{\mathbf{x}|\mathbf{y}} = \sum_{k,l} \alpha^{(k,l)}(\mathbf{y})\left(\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)} + \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T}\right) - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^T, \quad (11)$$

where

$$\alpha^{(k,l)}(\mathbf{y}) = \frac{p_k q_l f^{(k,l)}(\mathbf{y})}{f(\mathbf{y})}, \quad (12)$$

$$\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)} = \mathbf{C}_{\mathbf{xx}}^{(k)} - \mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T \left(\mathbf{C}_{\mathbf{yy}}^{(k,l)}\right)^{-1}\mathbf{H}\mathbf{C}_{\mathbf{xx}}^{(k)}, \quad (13)$$

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} = \mathbf{u}_{\mathbf{x}}^{(k)} + \mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T \left(\mathbf{C}_{\mathbf{yy}}^{(k,l)}\right)^{-1} \left(\mathbf{y} - \mathbf{u}_{\mathbf{y}}^{(k,l)}\right), \quad (14)$$

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}} = \sum_{k,l} \alpha^{(k,l)}(\mathbf{y})\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}. \quad (15)$$

Detailed derivations on (9)-(15) are not included here. The interested reader is instead referred to e.g. [16], [20], [33], [34]. In (8), rather than dealing directly with ϵ^2, it is convenient to study the matrix

$$\mathbf{M} = \int \mathbf{C}_{\mathbf{x}|\mathbf{y}} f(\mathbf{y}) d\mathbf{y} =$$

$$\sum_{k,l} p_k q_l \int \left(\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)} + \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^T\right) f^{(k,l)}(\mathbf{y}) d\mathbf{y}. \quad (16)$$

Here, the last equality results by using (11) and (12). Observe that $\text{Tr}(\mathbf{M}) = \epsilon^2$. Since $\mathbf{M}$ is of crucial interest, we inspect integral (16) term-by-term. The first term is

$$\mathbf{M}_1 = \sum_{k,l} p_k q_l \int \mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)} f^{(k,l)}(\mathbf{y}) d\mathbf{y} = \sum_{k,l} p_k q_l \mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)}, \quad (17)$$

where the last equality holds because $\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$ is not a function of $\mathbf{y}$, as can be seen from (13) and (10). The second term of (16) is

$$\mathbf{M}_2 = \sum_{k,l} p_k q_l \int \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T} f^{(k,l)}(\mathbf{y}) d\mathbf{y} \quad (18)$$

$$= \sum_{k,l} p_k q_l \left(\mathbf{u}_{\mathbf{x}}^{(k)}\mathbf{u}_{\mathbf{x}}^{(k)T} + \mathbf{C}_{\mathbf{xx}}^{(k)} - \mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\right). \quad (19)$$

The last equality derives using (14) and (13). The third term of (16) is

$$\mathbf{M}_3 = -\int \mathbf{u}_{\mathbf{x}|\mathbf{y}}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^T \sum_{k,l} p_k q_l f^{(k,l)}(\mathbf{y}) d\mathbf{y} \quad (20)$$

$$= -\sum_{k,l,r,s} p_k q_l p_r q_s \int \frac{f^{(k,l)}(\mathbf{y})f^{(r,s)}(\mathbf{y})\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(r,s)T}}{\sum_{v,w} p_v q_w f^{(v,w)}(\mathbf{y})} d\mathbf{y}. \quad (21)$$

The last equality results by using (15), (12) and (9). As far as we can see, (21) cannot be solved due to the sum in the denominator of the integrand. The MMSE,

$$\epsilon^2 = \text{Tr}(\mathbf{M}) = \text{Tr}(\mathbf{M}_1) + \text{Tr}(\mathbf{M}_2) + \text{Tr}(\mathbf{M}_3) \geq 0, \quad (22)$$

can therefore not be determined exactly. Instead we turn to bounds.

It can be verified that the existing Bayesian bounds [28]–[31] all depend on integral expressions of the following form

$$\iint \frac{g(\mathbf{y}, \mathbf{x})}{f(\mathbf{y}, \mathbf{x})} d\mathbf{x} \, d\mathbf{y}. \quad (23)$$

With (4) as input to (1), it can be shown that $f(\mathbf{y}, \mathbf{x})$ is a GM with $|\mathcal{K}| \cdot |\mathcal{L}|$ components. The function $g(\mathbf{y}, \mathbf{x})$ depends on the bound in question, but (23) cannot be evaluated analytically for any of the above bounds. The main reason, is that the integrand of (23), like in (21), has a GM denominator which does not simplify by substitutions. In addition, for any of these bounds, (23) cannot be construed as an expectation. The latter complicates Monte Carlo approaches, and numerical estimation of the mentioned bounds is therefore not straightforward. In contrast, $\mathbf{u}_{\mathbf{x}|\mathbf{y}}$ in (15) is analytical, and because GMs are easy to sample from, the MMSE *can* be estimated by Monte Carlo methods. Clearly, in practice, the MMSE should only be bounded by quantities that are considerably simpler to obtain. Otherwise, one would rather estimate the MMSE directly. Therefore, we instead seek *analytical* bounds - both upper and lower. To that end, note from equations (17), (18) and (20), that

$$\mathrm{Tr}(\mathbf{M}_1) = \sum_{k,l} p_k q_l \mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)}\right) \geq 0, \tag{24}$$

$$\mathrm{Tr}(\mathbf{M}_2) = \sum_{k,l} p_k q_l \int \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T} \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} f^{(k,l)}(\mathbf{y}) d\mathbf{y} \geq 0, \tag{25}$$

$$\mathrm{Tr}(\mathbf{M}_3) = -\sum_{k,l} p_k q_l \int \mathbf{u}_{\mathbf{x}|\mathbf{y}}^T \mathbf{u}_{\mathbf{x}|\mathbf{y}} f^{(k,l)}(\mathbf{y}) d\mathbf{y} \leq 0. \tag{26}$$

These inequalities follow because $\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$ is a covariance matrix, and $\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T} \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$ and $\mathbf{u}_{\mathbf{x}|\mathbf{y}}^T \mathbf{u}_{\mathbf{x}|\mathbf{y}}$ are inner products. Observe also that $\mathrm{Tr}(\mathbf{M}_2) + \mathrm{Tr}(\mathbf{M}_3)$ is always non-negative. This can be seen by applying (25) and (26):

$$\begin{aligned}
&\mathrm{Tr}(\mathbf{M}_2) + \mathrm{Tr}(\mathbf{M}_3) \\
&= \int \sum_{k,l} p_k q_l f^{(k,l)}(\mathbf{y}) \left(\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)T} \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}^T \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right) d\mathbf{y} \\
&= \int \sum_{k,l} \alpha^{(k,l)}(\mathbf{y}) \left\|\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\right\|_2^2 f(\mathbf{y}) d\mathbf{y} \geq 0. \tag{27}
\end{aligned}$$

The last equality is obtained from (12) and (15). Because $\mathrm{Tr}(\mathbf{M}_3)$ is non-analytical, we shall inquire whether $\mathrm{Tr}(\mathbf{M}_1)$ and $\mathrm{Tr}(\mathbf{M}_2)$ may provide useful bounds.

A. Proving the lower bound for the MMSE

Inequalities (22) and (24)-(27), guarantee that

$$\mathrm{Tr}(\mathbf{M}_1) \leq \epsilon^2. \tag{28}$$

Now we ask whether $\mathrm{Tr}(\mathbf{M}_2)$ also could serve as a lower bound. If so, then $\mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_3) \geq 0$ must be satisfied. If not, then $\mathrm{Tr}(\mathbf{M}_2)$ would be an *upper* bound for the MMSE. It can be verified (numerically) that this inequality may hold for certain GM parameters, but not for all. A larger obstacle, however, is that $\mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_3)$ is almost as demanding to compute as the MMSE itself. Therefore, among $\mathrm{Tr}(\mathbf{M}_2)$ and $\mathrm{Tr}(\mathbf{M}_1)$, the latter is the only practical lower bound.

An alternative argument provides intuition for the lower bound in (28). Imagine that side information is available, such that for each observation $\mathbf{y}$, a genie tells precisely which Gaussian components k and l came into play. Then, for each $\mathbf{y}$, we face a Gaussian signal in Gaussian noise. Because of the genie, we no longer use the observation $\mathbf{y}$ to determine the weights in (12). Instead, $\alpha^{(k,l)}(\mathbf{y}) = 1$ for the correct index pair (k, l), whereas $\alpha^{(k',l')}(\mathbf{y}) = 0$ for all other index pairs. Consequently, the estimator (15) pinpoints the correct posterior component mean, $\mathbf{u}_{\mathbf{x}|\mathbf{y}} = \mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$, as the estimate for $\mathbf{x}$. From (27), it is then clear that $\mathrm{Tr}(\mathbf{M}_2) + \mathrm{Tr}(\mathbf{M}_3) = 0$, and the MSE of the genie-aided estimator is therefore $\epsilon_{genie}^2 = \mathrm{Tr}(\mathbf{M}_1)$. One can interpret this estimator as a perfect decision device, followed

by a decision dependent Gaussian signal and Gaussian noise MMSE estimator. It is not realizable in practice, but it provides a performance benchmark: Any practical estimator, has MSE of at least $\mathrm{Tr}(\mathbf{M}_1)$. Therefore $\mathrm{Tr}(\mathbf{M}_1)$ is a lower bound, also for the MMSE. Section IV-C, shows that this bound coincides with the MMSE at infinite SNR, but simulations indicate that, in practice, the lower bound may be tight already from moderate SNR levels. By appropriate substitutions, using (17), (13) and (10) one obtains the lower bound of Proposition 1.

B. Proving the upper bound for the MMSE

From inequalities (22) and (24)-(26), it is straightforward to conclude that

$$\epsilon^2 \leq \mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_2). \tag{29}$$

However, this upper bound is not really informative. In order to see this, apply (17) and (19) to have

$$\begin{aligned}
\mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_2) &= \sum_k p_k \left(\mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}^{(k)}\right) + \mathbf{u}_{\mathbf{x}}^{(k)T} \mathbf{u}_{\mathbf{x}}^{(k)}\right) \\
&\geq \sum_k p_k \left(\mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}^{(k)}\right) + \mathbf{u}_{\mathbf{x}}^{(k)T} \mathbf{u}_{\mathbf{x}}^{(k)}\right) - \mathbf{u}_x^T \mathbf{u}_x = \mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right).
\end{aligned}$$

The last equality follows from (6). $\mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right) = E\left\{\|\mathbf{x} - \mathbf{u}_{\mathbf{x}}\|_2^2\right\}$ is the MSE of the highly suboptimal estimator $\hat{\mathbf{x}} = \mathbf{u}_{\mathbf{x}}$. This estimator, which completely disregards the observation $\mathbf{y}$, has an MSE which is even smaller than $\mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_2)$!

Because $\mathrm{Tr}(\mathbf{M}_2)$, $\mathrm{Tr}(\mathbf{M}_1) + \mathrm{Tr}(\mathbf{M}_2)$ and $\mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right)$ all are inadequate as upper bounds, we invoke the LMMSE estimator[1]. This estimator is given by (see e.g. Theorem 12.1 of [4])

$$\hat{\mathbf{x}} = \mathbf{u}_{\mathbf{x}} + \mathbf{C}_{\mathbf{xx}} \mathbf{H}^T \left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{u}_{\mathbf{x}} - \mathbf{u}_{\mathbf{n}}), \tag{30}$$

where $\mathbf{C}_{\mathbf{xx}}$, $\mathbf{u}_{\mathbf{x}}$, $\mathbf{C}_{\mathbf{nn}}$ and $\mathbf{u}_{\mathbf{n}}$ are given in (6) and (7). Its error, $\mathbf{x} - \hat{\mathbf{x}}$, has zero mean and covariance matrix

$$\mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{xx}} \mathbf{H}^T \left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1} \mathbf{H}\mathbf{C}_{\mathbf{xx}}. \tag{31}$$

The (closed form) MSE of the LMMSE estimator is

$$\begin{aligned}
\epsilon_L^2 &= E\left\{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right\} \\
&= \mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{xx}}\mathbf{H}^T \left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1} \mathbf{H}\mathbf{C}_{\mathbf{xx}}\right) \\
&= \mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right) - \sum_j \mathbf{g}_j^T \left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1} \mathbf{g}_j, \tag{32}
\end{aligned}$$

where, in the last equality, $\mathbf{g}_j$ is the j-th column of $\mathbf{H}\mathbf{C}_{\mathbf{xx}}$. Recall that $\mathbf{C}_{\mathbf{xx}}$ and $\mathbf{C}_{\mathbf{nn}}$ are both positive definite and symmetric matrices. Thus, provided the LMMSE estimator in (30) exists, $\left(\mathbf{H}\mathbf{C}_{\mathbf{xx}}\mathbf{H}^T + \mathbf{C}_{\mathbf{nn}}\right)^{-1}$ has to be positive definite. This, and (32), shows that $\epsilon_L^2 \leq \mathrm{Tr}\left(\mathbf{C}_{\mathbf{xx}}\right)$. Because no practical estimator has MSE smaller than the MMSE, by definition, we can safely conclude that

$$\epsilon^2 \leq \epsilon_L^2.$$

Note that ϵ_L^2 in (32) corresponds to the upper bound of Proposition 1. Clearly, if another estimator with a closed form MSE less than ϵ_L^2 exists, it can tighten this upper bound. For the present problem, we are not aware of any.

---

[1] Among all estimators which are linear (affine) in the observations, the LMMSE estimator obtains the smallest MSE.

C. MMSE at zero and infinite SNR

Intuitively, one expects that the MMSE approaches its upper bound as

$$\text{SNR} = E\left\{\|\mathbf{x}\|_2^2\right\}/E\left\{\|\mathbf{n}\|_2^2\right\}, \quad (33)$$

tends to zero. Similarly, we expect the MMSE to approach its lower bound as the SNR tends to infinity. We will demonstrate that this intuition is true for a simple but instructive example. Throughout, we assume $\mathbf{H}$ to be square and full rank. In addition, we assume the noise to be distributed as

$$\mathbf{n} \sim \sum_{l \in \mathcal{L}} q_l \mathcal{N}(a\mathbf{u}_\mathbf{n}^{(l)}, a^2 \mathbf{C}_\mathbf{nn}^{(l)}), \quad (34)$$

where the scalar a can account for any SNR. Then (14) becomes

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$$
$$=\mathbf{u}_\mathbf{x}^{(k)} + \mathbf{C}_\mathbf{xx}^{(k)}\mathbf{H}^T\Big(\mathbf{H}\mathbf{C}_\mathbf{xx}^{(k)}\mathbf{H}^T + a^2\mathbf{C}_\mathbf{nn}^{(l)}\Big)^{-1}\Big(\mathbf{y} - \mathbf{H}\mathbf{u}_\mathbf{x}^{(k)} - a\mathbf{u}_\mathbf{n}^{(l)}\Big) \quad (35)$$

$$=\mathbf{u}_\mathbf{x}^{(k)} + \Big(\big(\mathbf{C}_\mathbf{xx}^{(k)}\big)^{-1} + \mathbf{H}^T\frac{1}{a^2}\big(\mathbf{C}_\mathbf{nn}^{(l)}\big)^{-1}\mathbf{H}\Big)^{-1}\mathbf{H}^T\frac{1}{a^2}\big(\mathbf{C}_\mathbf{nn}^{(l)}\big)^{-1}$$
$$\cdot\Big(\mathbf{y} - \mathbf{H}\mathbf{u}_\mathbf{x}^{(k)} - a\mathbf{u}_\mathbf{n}^{(l)}\Big), \quad (36)$$

where (36) is merely an equivalent but useful form (see e.g. equation (10.32) of [4]). Similarly, the LMMSE estimator in (30) becomes

$$\hat{\mathbf{x}} = \mathbf{u}_\mathbf{x} + \mathbf{C}_\mathbf{xx}\mathbf{H}^T\Big(\mathbf{H}\mathbf{C}_\mathbf{xx}\mathbf{H}^T + a^2\mathbf{C}_\mathbf{nn}\Big)^{-1}(\mathbf{y} - \mathbf{H}\mathbf{u}_\mathbf{x} - a\mathbf{u}_\mathbf{n}) \quad (37)$$

$$= \mathbf{u}_\mathbf{x} + \Big(\mathbf{C}_\mathbf{xx}^{-1} + \mathbf{H}^T\frac{\mathbf{C}_\mathbf{nn}^{-1}}{a^2}\mathbf{H}\Big)^{-1}\mathbf{H}^T\frac{\mathbf{C}_\mathbf{nn}^{-1}}{a^2}(\mathbf{y} - \mathbf{H}\mathbf{u}_\mathbf{x} - a\mathbf{u}_\mathbf{n}), \quad (38)$$

where (38) is also an equivalent form. Given (34), then (13) becomes

$$\mathbf{C}_{\mathbf{x}|\mathbf{y}}^{(k,l)} = \mathbf{C}_\mathbf{xx}^{(k)} - \mathbf{C}_\mathbf{xx}^{(k)}\mathbf{H}^T\Big(\mathbf{H}\mathbf{C}_\mathbf{xx}^{(k)}\mathbf{H}^T + a^2\mathbf{C}_\mathbf{nn}^{(l)}\Big)^{-1}\mathbf{H}\mathbf{C}_\mathbf{xx}^{(k)}. \quad (39)$$

1) Zero SNR: We drive the SNR to zero by $a \to \infty$. Then (36) reduces to $\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} = \mathbf{u}_\mathbf{x}^{(k)}$. Thus, the MMSE estimate for $\mathbf{x}$ is

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}} = \frac{\sum_{k,l} p_k q_l f^{(k,l)}(\mathbf{y})\mathbf{u}_\mathbf{x}^{(k)}}{\sum_{r,s} p_r q_s f^{(r,s)}(\mathbf{y})} = \sum_k p_k\mathbf{u}_\mathbf{x}^{(k)} = \mathbf{u}_\mathbf{x}. \quad (40)$$

Here, the second equality holds because $f^{(k,l)}(\mathbf{y})$ is a function of the covariance matrix, $\mathbf{H}\mathbf{C}_\mathbf{xx}^{(k)}\mathbf{H}^T + a^2\mathbf{C}_\mathbf{nn}^{(l)}$, and when $a \to \infty$, $f^{(k,l)}(\mathbf{y})$ approaches a uniform distribution with infinite support. Hence, $f^{(k,l)}(\mathbf{y})$ approaches a constant which is independent of $\mathbf{y}$, k and l. Note from (40), that the MMSE estimator discards the data and only uses prior information. This is expected at zero SNR.

Now we compare with the LMMSE estimator. When $a \to \infty$, (38) reduces to $\hat{\mathbf{x}} = \mathbf{u}_\mathbf{x} = \sum_k p_k\mathbf{u}_\mathbf{x}^{(k)}$. But this is equal to (40), which is the MMSE estimate. Thus, the LMMSE estimator becomes *the* MMSE estimator, at zero SNR. As argued in Section IV-B the MMSE becomes $\text{Tr}(\mathbf{C}_\mathbf{xx})$ in this case.

2) Infinite SNR: We drive the SNR to infinity by $a \to 0$. Then, because $\mathbf{H}$ is square and full rank, (35) reads $\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)} = \mathbf{H}^{-1}\mathbf{y}$. Using this in (15), we find that

$$\mathbf{u}_{\mathbf{x}|\mathbf{y}} = \mathbf{H}^{-1}\mathbf{y}. \quad (41)$$

From (27), it can then be seen that $\text{Tr}(\mathbf{M}_2) + \text{Tr}(\mathbf{M}_3) = 0$. Thus, at infinite SNR, the MMSE in (22) becomes $\text{Tr}(\mathbf{M}_1)$. Evaluating (39) when $a \to 0$, and plugging the result into (24), it is straight forward to verify that MMSE becomes zero at infinite SNR. Note from (41), that the estimator discards all prior knowledge and completely trusts the data. This is expected at infinitely high SNR.

To conclude this inquiry, we again compare with the LMMSE estimator. When $a \to 0$, (37) reduces to $\hat{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{y}$. But this is the
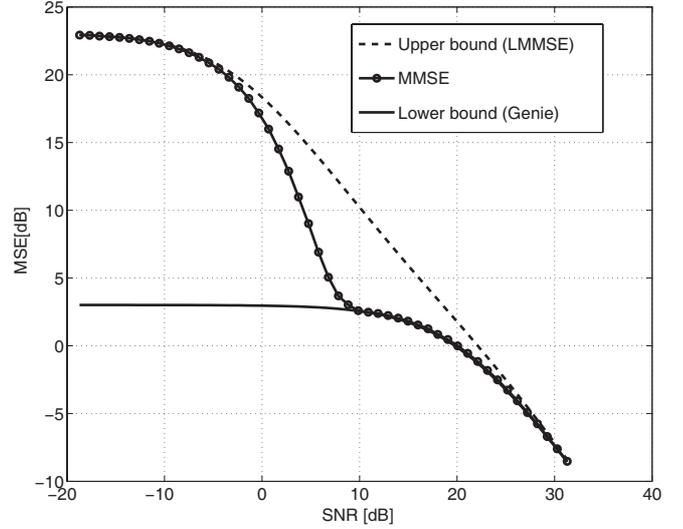

Fig. 1. The MMSE and its analytical upper and lower bounds.

same as the MMSE estimator in (41). Thus, the LMMSE estimator becomes *the* MMSE estimator, also at infinite SNR.

In summary, at both zero and infinite SNR, the MMSE becomes analytical and equals the MSE of the LMMSE estimator. At infinite SNR, the MMSE also equals the MSE of the genie-aided estimator. Therefore, the LMMSE upper bound and the genie lower bound must coincide as the SNR tends to infinity.

V. NUMERICAL RESULTS

Here we simulate the MMSE, and calculate the bounds, all as functions of SNR. The MMSE can be estimated as the sample mean of $\|\mathbf{x} - \mathbf{u}_{\mathbf{x}|\mathbf{y}}\|_2^2$ from a series of observations. However, the computational burden grows rapidly with the number of mixture components and the dimensions of $\mathbf{H}$. We have simulated several scenarios, all of which show the same trend. Therefore we only consider a low-dimensional system here, and impose inputs (4) such that $|\mathcal{K}| \cdot |\mathcal{L}|$ is moderate. Specifically, we assume

- $\mathbf{H} = \mathbf{I}$, with $\mathbf{I}$ being 2×2 identity matrix.
- $\mathbf{x}$ is GM distributed ($|\mathcal{K}| = 2$) with parameters

$$p_1 = p_2 = \frac{1}{2}, \quad \mathbf{u}_\mathbf{x}^{(1)} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad \mathbf{u}_\mathbf{x}^{(2)} = \begin{bmatrix} -10 \\ -10 \end{bmatrix}, \quad \mathbf{C}_\mathbf{xx}^{(1)} = \mathbf{C}_\mathbf{xx}^{(2)} = \mathbf{I}$$

- $\mathbf{n}$ is also GM distributed ($|\mathcal{L}| = 2$), with parameters

$$q_1 = q_2 = \frac{1}{2}, \quad \mathbf{u}_\mathbf{n}^{(1)} = \mathbf{u}_\mathbf{n}^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{C}_\mathbf{nn}^{(1)} = \frac{a}{2}\mathbf{I}, \quad \mathbf{C}_\mathbf{nn}^{(2)} = a\mathbf{I}$$

- The scalar a is adjusted to produce different SNR values.

Figure 1 shows the genie lower bound, the estimated MMSE, and the upper bound, all in dB, versus increasing SNR. The MMSE is estimated from 10^5 independent $\mathbf{y}$'s for each SNR value. Figure 1 is in line our previous findings: At very low and very high SNR, the MMSE coincides with the MSE of the LMMSE estimator. Also, the bounds approach each other with increasing SNR. Interestingly, for the current setup, the MMSE estimator performs as if helped by a genie for SNR>10 dB. Thus, in practice, the MMSE estimator pinpoints the correct Gaussian component (k,l), and places nearly all weight in $\alpha^{(k,l)}(\mathbf{y})$, already from modest SNR.

From Figure 1 it is apparent that the MMSE estimator out-performs the LMMSE estimator, especially at intermediate SNR levels. For designers, it could therefore be of interest to know

the approximate computational cost of this MSE gain. It can be determined as follows. Note from (14) that $\mathbf{u}_{\mathbf{x}|\mathbf{y}}^{(k,l)}$ depends on the (say $N \times M$) matrix $\mathbf{C}_{\mathbf{xx}}^{(k)}\mathbf{H}^T\mathbf{C}_{\mathbf{yy}}^{-(k,l)}$. When all parameters are known, this matrix can be computed offline. Defining $|\mathcal{K}| \cdot |\mathcal{L}| = S$, it can be verified that the *online* cost for computing (15) is in the order of $\mathcal{O}(SNM) + \mathcal{O}(SM^2)$ multiplications and additions. The corresponding cost for the LMMSE estimate in (30) is only $\mathcal{O}(NM)$. Roughly, the LMMSE estimator's complexity depends on the dimensions of the system. In contrast, the MMSE estimator's complexity depends on the dimensions multiplied by the number of mixture components, S.

For the interested reader, and in order to facilitate reproducible research, the MATLAB code which produced Figure 1 can be downloaded from: http://sites.google.com/site/johntorjusflaam.

VI. Conclusion

Much motivation for this paper derives from the applicability and generality of GMs. Assuming GM inputs to a Bayesian linear model, we show that the MMSE does not come in analytical form. Existing Bayesian lower bounds are, however, not attractive in this setup. Instead the MMSE can be bounded analytically - from above and below. The LMMSE estimator yields an upper bound, and a genie aided MMSE estimator offers a lower one. The genie-aided estimator consists of a perfect decision device, followed by a decision dependent Gaussian signal and Gaussian noise MMSE estimator. We have shown that the upper and lower bounds approach each other with increasing SNR. We have also studied the MMSE in the extreme cases of zero and infinite SNR. In both of these cases, the MMSE becomes analytical and corresponds to the MSE of the LMMSE estimator. A numerical example displays the behavior of these bounds as a function of SNR. It indicates that, in practice, the MMSE estimator has performance comparable to the genie-aided estimator already from quite modest SNR.

VII. Acknowledgments

References

[1] Jonathan Q. Li and Andrew R. Barron, "Mixture Density Estimation," in *In Advances in Neural Information Processing Systems 12*. 1999, pp. 279–285, MIT Press.

[2] H.W. Sorenson and D.L. Alspach, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, vol. 7, no. 4, pp. 465 – 479, 1971.

[3] S. Dasgupta, "Learning Mixtures of Gaussians," in *Foundations of Computer Science, 1999. 40th Annual Symposium on*, 1999, pp. 634 –644.

[4] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[5] Zoubin Ghahramani, "Solving Inverse Problems using an EM approach to Density Estimation," in *Proceedings of the 1993 Connectionist Models Summer School*, 1993, pp. 316–323.

[6] David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[7] Michael P. Perrone and Leon N. Cooper, "When Networks Disagree: Ensemble Methods for Hybrid Neural Networks," in *Artificial Neural Networks for Speech and Vision*. 1993, pp. 126–142, Chapman and Hall.

[8] D. Ormoneit and V. Tresp, "Averaging, Maximum Penalized Likelihood and Bayesian Estimation for Improving Gaussian Mixture Probability Density Estimates," *Neural Networks, IEEE Transactions on*, vol. 9, no. 4, pp. 639 –650, jul 1998.

[9] G. Yu, G. Sapiro, and S. Mallat, "Solving Inverse Problems With Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity," *Image Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2011.

[10] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *Automatic Control, IEEE Transactions on*, vol. 17, no. 4, pp. 439 – 448, aug 1972.

[11] Igal Bilik and Joseph Tabrikian, "MMSE-Based Filtering in Presence of Non-Gaussian System and Measurement Noise," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 3, pp. 1153 – 1170, july 2010.

[12] I. Bilik and J. Tabrikian, "Maneuvering Target Tracking in the Presence of Glint using the Nonlinear Gaussian Mixture Kalman Filter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 46, no. 1, pp. 246 –262, jan. 2010.

[13] J.H. Kotecha and P.M. Djuric, "Gaussian Sum Particle Filtering," *Signal Processing, IEEE Transactions on*, vol. 51, no. 10, pp. 2602 – 2612, oct. 2003.

[14] G. Ackerson and K. Fu, "On State Estimation in Switching Environments," *Automatic Control, IEEE Transactions on*, vol. 15, no. 1, pp. 10 – 17, feb 1970.

[15] J. Tugnait and A. Haddad, "Adaptive Estimation in Linear Systems with Unknown Markovian Noise Statistics," *Information Theory, IEEE Transactions on*, vol. 26, no. 1, pp. 66 – 78, jan 1980.

[16] M. Afify, X. Cui, and Y. Gao, "Stereo-Based Stochastic Mapping for Robust Speech Recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1325 –1334, sept. 2009.

[17] Xiaodong Cui, M. Afify, and Yuqing Gao, "MMSE-based Stereo Feature Stochastic Mapping for Noise Robust Speech Recognition," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 4077 –4080.

[18] L. Buera, E. Lleida, A. Miguel, A. Ortega, and O. Saz, "Cepstral Vector Normalization Based on Stereo Data for Robust Speech Recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 3, pp. 1098 –1113, march 2007.

[19] J. Samuelsson and P. Hedelin, "Recursive Coding of Spectrum Parameters," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 5, pp. 492 –503, jul 2001.

[20] D. Persson and T. Eriksson, "Mixture Model- and Least Squares-Based Packet Video Error Concealment," *Image Processing, IEEE Transactions on*, vol. 18, no. 5, pp. 1048 –1054, may 2009.

[21] A. Kundu, S. Chatterjee, and T.V. Sreenivas, "Subspace Based Speech Enhancement Using Gaussian Mixture Model," in *Interspeech 2008*, Brisbane, Australia, september 2008, pp. 395 –398.

[22] A. Kundu, S. Chatterjee, and T.V. Sreenivas, "Speech Enhancement using Intra-Frame Dependency in DCT Domain," in *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 25-29 2008.

[23] A.D. Subramaniam, W.R. Gardner, and B.D. Rao, "Low-Complexity Source Coding Using Gaussian Mixture Models, Lattice Vector Quantization, and Recursive Coding with Application to Speech Spectrum Quantization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 2, pp. 524 – 532, march 2006.

[24] Li Deng, J. Droppo, and A. Acero, "Estimating Cepstrum of Speech Under the Presence of Noise Using a Joint Prior of Static and Dynamic Features," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 3, pp. 218 – 233, may 2004.

[25] Veronique Stouten, Hugo Van hamme, and Patrick Wambacq, "Model-based Feature Enhancement with Uncertainty Decoding for Noise Robust ASR," *Speech Communication*, vol. 48, no. 11, pp. 1502 – 1514, 2006.

[26] Wooil Kim and John H.L. Hansen, "Feature compensation in the cepstral domain employing model combination," *Speech Communication*, vol. 51, no. 2, pp. 83 – 96, 2009.

[27] J.A. Gonzandlez, A.M. Peinado, A.M. Gandmez, and J.L. Carmona, "Efficient MMSE Estimation and Uncertainty Processing for Multienvironment Robust Speech Recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1206 –1220, july 2011.

[28] Harry L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 1*, John Wiley and Sons, 1968.

[29] B. Bobrovsky and M. Zakai, "A Lower Bound on the Estimation Error for Certain Diffusion Processes," *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 45 – 52, jan 1976.

[30] A. Weiss and E. Weinstein, "A Lower Bound on the Mean-Square Error in Random Parameter Estimation (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 31, no. 5, pp. 680 – 682, sep 1985.

[31] I. Reuven and H. Messer, "A Barankin-Type Lower Bound on the Estimation Error of a Hybrid Parameter Vector," *Information Theory, IEEE Transactions on*, vol. 43, no. 3, pp. 1084 –1093, may 1997.

[32] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation; Part II: Bayesian Bounds," *Information Theory, IEEE Transactions on*, vol. 56, no. 10, pp. 5064 –5082, oct. 2010.

[33] A. Kundu, S. Chatterjee, A. Sreenivasa Murthy, and T.V. Sreenivas, "GMM based Bayesian approach to Speech Enhancement in Signal / Transform domain," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 4893 –4896.

[34] John T. Flåm, Joakim Jaldén, and Saikat Chatterjee, "Gaussian Mixture Modeling for Source Localization," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 2604 –2607.