Lower and Upper Bounds on the Minimum Mean-Square Error in Composite Source Signal Estimation

Yariv Ephraim, *Senior Member, IEEE,* and Neri Merhav, *Member, IEEE*

***Abstract*—Performance analysis of a minimum mean-square error (mmse) estimator for the output signal from a composite source model (CSM), which has been degraded by statistically independent additive noise, is performed for a wide class of discrete as well as continuous time models. The noise in the discrete time case is assumed to be generated by another CSM. For the continuous time case only Gaussian white noise, or a single state CSM noise, is considered. In both cases, the mmse is decomposed into the mmse of the estimator which is informed of the exact states of the signal and noise, and an additional error term. This term is tightly upper and lower bounded. The bounds for the discrete time signals are developed using distribution tilting and Shannon's lower bound on the probability of a random variable to exceed a given threshold. The analysis for the continuous time signals is performed using Duncan's theorem. The bounds in this case are developed by applying the data processing theorem to sampled versions of the state process and its estimate, and by using Fano's inequality. The bounds in both cases are explicitly calculated for CSM's with Gaussian subsources. For causal estimation, these bounds approach zero harmonically as the duration of the observed signals approaches infinity.**

***Index Terms*—Composite sources, minimum mean-square error estimation, distribution tilting, Duncan's theorem.**

I. INTRODUCTION

MINIMUM mean-square error (mmse) estimation performed using discrete time composite source models (CSM's) [1] for the signal and for an additive statistically independent noise is of primary interest in speech enhancement applications [7] for the following reasons.

1) CSM's have proven useful for speech signals [2] and for frequently encountered noise sources [7]. Furthermore, the mmse estimator is optimal for a large class of difference distortion measures, not only the mean-squared error (mse) measure, provided that the posterior probability density function (pdf) of the clean signal given the noisy signal is symmetric about its mean [3, pp. 60–63]. This class includes all convex U difference distortion measures. Hence, CSM based mmse estimators are potentially good estimators for speech signals, since the pdf of these signals, and often also the pdf of the noise process, are not available, and the most perceptually meaningful distortion measure is unknown.
2) The mmse estimator of the signal is the optimal preprocessor in mmse waveform vector quantization (VQ) [4]–[5]. Furthermore, the mmse estimator of the sample spectrum of the signal is the optimal preprocessor in autoregressive (AR) model VQ in the Itakura–Saito sense [5].
3) The *causal* mmse estimator of the signal is the optimal preprocessor in minimum probability of error classification of *any* finite energy *continuous* time signal contaminated by white Gaussian noise [6].

A discrete time CSM is a finite set of statistically independent subsources that are controlled by a switch [1]. The position of the switch at each time instant is randomly selected according to some probability law. Throughout this paper, each subsource is assumed a statistically independent identically distributed (i.i.d.) vector source, and the switch is assumed to be governed by a first-order Markov chain. The model obtained in this way is referred to as a hidden Markov model (HMM) in the speech literature [2]. Each position of the switch defines a state of the source. A pair of states of the signal and noise defines a composite state of the noisy signal.

The CSM based mmse estimator comprises a weighted sum of conditional mean estimators for the composite states of the noisy signal [7]. For causal mmse estimation of a vector of the clean signal, the weights are the posterior probabilities of the composite states given all past and present vectors of the noisy signal. The causality of the estimator in this case is with respect to vectors of the signals rather than the samples within each vector. These samples, except for the last one, are not estimated in a causal manner. The mmse estimator was originally developed by Magill [8], and subsequently in [9]–[10], for a particular case CSM and white Gaussian noise. This model assumes that the switch remains fixed at a randomly

Manuscript received May 21, 1991; revised February 11, 1992. This work was presented at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 24–28, 1991.

Y. Ephraim is with the Speech Research Department, AT & T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974. He is also with the C[3]I Center, George Mason University, Fairfax, VA 22030.

N. Merhav was with the Speech Research Department, AT & T Bell Laboratories, Murray Hill, NJ 07974. He is now with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9201543.

selected initial position. Hence, in essence, the model used in [8]–[10] is a mixture model [11]. The CSM used here is more general since it allows state transitions each time a new output vector is generated.

The purpose of this paper is to theoretically analyze the performance of the CSM based mmse signal estimator which has proven useful in speech enhancement applications [7]. A second-order analysis is performed. Since the estimator is unbiased in the sense that the expected value of the error signal is zero, only the mmse is studied. The analysis is performed for a wide class of CSM's whose initial state probabilities and state transition probabilities are strictly positive. The subsources are assumed to satisfy only mild technical regularity conditions. It is shown that the mmse can be decomposed into two error components. The first is the mmse of the estimator that is informed of the exact composite state of the noisy signal at each time instant. The second error component represents the sum of cross error terms corresponding to pairs of composite states. This component is evaluated using the "sandwich" approach. Specifically, tight upper and lower bounds are developed for each cross error term. The bounds are first shown to be dependent on the probability of classification error in a two class hypothesis testing problem. Then, the probability of misclassification is upper and lower bounded using distribution tilting [3], [29], and Shannon's lower bound on the probability of a random variable to exceed a given threshold [18]. These bounds resemble the Chernoff bound [29]. The bounds are explicitly evaluated for the most commonly used CSM's, i.e., those whose subsources are asymptotically weakly stationary (AWS) [14], [15] Gaussian processes. Examples of such sources are Gaussian AR processes. For this case, the bounds are shown to converge exponentially to zero as the vector dimension of the output signal approaches infinity. Hence, the asymptotic mmse is the mmse of the informed estimator.

An intuitive suboptimal detection-estimation scheme is also analyzed. In this scheme, the composite state at each given time instant is first estimated from all past and present vectors of the noisy signal. Then, the conditional mean estimator associated with the estimated state is applied to the noisy signal. It is shown that the mse associated with this scheme can be decomposed similarly to the mmse, and that the cross error terms can be upper and lower bounded by bounds similar to those developed for the mmse estimator. Hence, for CSM's with AWS Gaussian subsources, the detection-estimation scheme is asymptotically optimal in the mmse sense.

Next, the mmse in causal estimation of the output signal from a continuous time CSM, which has been degraded by statistically independent additive Gaussian white noise, is analyzed. The continuous time CSM is defined analogously to the discrete time CSM. A Markov chain whose state transition may occur every T seconds is assumed. During each T second interval, a random output process whose statistics depend on the state is generated. The mmse analysis for the continuous time CSM's is performed using Duncan's theorem [13]. This theorem relates the mmse in strictly causal estimation of the signal to the average mutual information between the clean and the noisy signals assuming Gaussian white noise. Similarly to the discrete-time case, the mmse can be decomposed into the mmse of the informed estimator, and an additional error term for which upper and lower bounds are developed. The error term in this case equals the average mutual information between the state process and the noisy signal. For CSM's with AWS Gaussian subsources, these upper and lower bounds are shown to converge harmonically to zero as the signal duration approaches infinity. The difference in convergence rate for discrete and continuous time mmse signal estimation, is attributed to the fact that in the continuous time case strictly causal estimation is performed while in the discrete time case noncausal estimation is essentially performed.

This paper is organized as follows. In Section II, we develop the upper and lower bounds on the mmse for discrete time CSM's. In Section III, we provide explicit expressions for those bounds for the case of CSM's with AWS Gaussian subsources. In Section IV, we develop similar bounds for the detection-estimation scheme. In Section V, we focus on the bounds for the continuous time CSM's. In Section VI, we demonstrate that the bonding technique used here is useful for mmse parameter estimation. Comments are given in Section VII.

II. MMSE Analysis for Discrete Time CSM's

A. Preliminaries

Let $y_t \in R^K$ be a K-dimensional vector of the clean signal. Similarly, let $\nu_t \in R^K$ be a K-dimensional vector of the noise process. Assume that the noise is additive and statistically independent of the signal. Let $z_t = y_t + \nu_t$ be a K-dimensional vector of the noisy process. Let $y_0^t \triangleq \{y_\tau, \tau = 0,\cdots,t\}$, $\nu_0^t \triangleq \{\nu_\tau, \tau = 0,\cdots,t\}$, and $z_0^t \triangleq \{z_\tau, \tau = 0,\cdots,t\}$.

Let $p(y_0^t)$ be the pdf of an M-state discrete time CSM for the clean signal. Let $x_0^t \triangleq \{x_\tau, \tau = 0,\cdots,t\}$ denote a sequence of signal states corresponding to y_0^t. For each τ, $x_\tau \in \{1,\cdots,M\}$. For CSM's with i.i.d. vector subsources and a first-order Markov switch, the pdf $p(y_0^t)$ can be written as

$$p(y_0^t) = \sum_{x_0^t} p(x_0^t) p(y_0^t|x_0^t) = \sum_{x_0^t} \prod_{\tau=0}^{t} a_{x_{\tau-1}x_\tau} b(y_\tau|x_\tau), \quad (1)$$

where $a_{x_{\tau-1}x_\tau}$ denotes the transition probability from state $x_{\tau-1}$ at time $\tau-1$ to state x_τ at time τ, $a_{x_{-1}x_0} \triangleq \pi_{x_0}$ denotes the probability of the initial state x_0, and $b(y_\tau|x_\tau)$ denotes the pdf of the output vector y_τ from the subsource x_τ. Such a model will be referred to as a first-order M-state discrete time CSM. For simplicity of notation and terminology, we assume that $b(y_\tau|x_\tau)$ is the pdf of an absolutely continuous probability distribution (pd). The analysis performed here, however, will be applicable to mixtures of discrete and continuous pd's that satisfy some regularity conditions that will be specified shortly.

Similarly, let $p(\nu_0^t)$ be the pdf of a first-order $\tilde{M}$-state discrete time CSM for the noise process. This pdf is given by

$$p(\nu_0^t) = \sum_{\tilde{x}_0^t} \prod_{\tau=0}^{t} a_{\tilde{x}_{\tau-1}\tilde{x}_\tau} b(\nu_\tau|\tilde{x}_\tau), \tag{2}$$

where $\tilde{x}_0^t \triangleq \{\tilde{x}_\tau, \tau = 0,\cdots,t\}$ denotes a sequence of noise states, and $b(\nu_\tau|\tilde{x}_\tau)$ is the pdf of the output vector ν_τ from the noise subsource $\tilde{x}_\tau$.

It is easy to show that $p(z_0^t)$, the pdf of the model for the noisy signal, is a first-order discrete time CSM with $\overline{M} \triangleq M \times \tilde{M}$ composite states. A composite state of the noisy signal at time t is defined as $\bar{x}_t \triangleq (x_t, \tilde{x}_t)$. This pdf is given by

$$p(z_0^t) = \sum_{\bar{x}_0^t} \prod_{\tau=0}^{t} a_{\bar{x}_{\tau-1}\bar{x}_\tau} b(z_\tau|\bar{x}_\tau), \tag{3}$$

where

$$a_{\bar{x}_{\tau-1}\bar{x}_\tau} \triangleq a_{x_{\tau-1}x_\tau} a_{\tilde{x}_{\tau-1}\tilde{x}_\tau},$$

$$b(z_\tau|\bar{x}_\tau) = \int b(z_\tau - y_\tau|\tilde{x}_\tau) b(y_\tau|x_\tau)\, dy_\tau. \tag{4}$$

Note that we use generic notation for the state transition probabilities, and the state dependent pdf's, for the CSM's for the signal, the noise, and the noisy process. The distinction between the models is made here through the different notation used for the state sequences from these models. Thus, $a_{x_{t-1}x_t}$, $a_{\tilde{x}_{t-1}\tilde{x}_t}$, and $a_{\bar{x}_{t-1}\bar{x}_t}$ denote, respectively, transition probabilities between states of the model for the clean signal, the noise process, and the noisy process. Similarly, $b(y_t|x_t)$, $b(\nu_t|\tilde{x}_t)$, and $b(z_t|\bar{x}_t)$ denote, respectively, the pdf's of the output vectors at time t from the subsources of the models for the clean signal, the noise process, and the noisy signal.

Similarly to (3)–(4), it can be shown that the pdf of y_t given z_0^T, $T \geq t$, is given by

$$p(y_t|z_0^T) = \sum_{\bar{x}_t} p(\bar{x}_t|z_0^T) b(y_t|z_t, \bar{x}_t), \tag{5}$$

where $p(\bar{x}_t|z_0^T)$ denotes the posterior probability of the composite state of the noisy signal at time t given the observed signal z_0^T, and $b(y_t|z_t, \bar{x}_t)$ is the conditional pdf of y_t given z_t and $\bar{x}_t$. The conditional probability $p(\bar{x}_t|z_0^T)$ in (5), and the pdf $p(z_0^t)$ in (3), can be efficiently calculated using the "forward–backward" formulas for HMM's (see, e.g., [16, (25)–(27)]). Since we focus here on causal estimation only, we provide the forward recursion for calculating $p(\bar{x}_t|z_0^t)$ and $p(z_0^t)$. The extension of the discussion to noncausal estimation can be found in [7]. We have that

$$F(\bar{x}_0, z_0) = p(\bar{x}_0, z_0) = \pi_{\bar{x}_0} b(z_0|\bar{x}_0), \tag{6}$$

$$F(\bar{x}_t, z_0^t) = p(\bar{x}_t, z_0^t) = \sum_{\bar{x}_{t-1}} F(\bar{x}_{t-1}, z_0^{t-1}) a_{\bar{x}_{t-1}\bar{x}_t} b(z_t|\bar{x}_t), \quad 0 < t \leq T. \tag{7}$$

Hence,

$$p(\bar{x}_t|z_0^t) = \frac{F(\bar{x}_t, z_0^t)}{\sum_{\bar{x}_t} F(\bar{x}_t, z_0^t)} \tag{8}$$

and

$$p(z_0^t) = \sum_{\bar{x}_t} F(\bar{x}_t, z_0^t). \tag{9}$$

B. MMSE Estimation

The causal (in the vector sense) mmse estimator of y_t given z_0^t can be obtained from (5). This estimator is given by

$$\hat{y}_t = E\{y_t|z_0^t\} = \sum_{\bar{x}_t} p(\bar{x}_t|z_0^t) E\{y_t|\bar{x}_t, z_t\} \triangleq \sum_{\bar{x}_t} p(\bar{x}_t|z_0^t) \hat{y}_{t|\bar{x}_t}. \tag{10}$$

A block diagram of this estimator is shown in Fig. 1. This estimator is unbiased in the sense that

$$E\{\hat{y}_t\} = E\{y_t\}, \tag{11}$$

as can be shown by using the rule of iterated expectation [17, p. 161].

The mse associated with $\hat{y}_t$ can be calculated using the orthogonality principle [17, p. 164]

$$E\{(y_t - \hat{y}_t)\hat{y}_t^{\#}\} = 0. \tag{12}$$

Using (12), the rule of iterated expectation, (5), and (10) in that order, we can write the mmse as

$$\overline{\epsilon_t^2} \triangleq \frac{1}{K} \operatorname{tr} E\{(y_t - \hat{y}_t)(y_t - \hat{y}_t)^{\#}\} = \frac{1}{K} \operatorname{tr} E\{(y_t - \hat{y}_t) y_t^{\#}\} = \frac{1}{K} \operatorname{tr} E\{E\{y_t y_t^{\#}|z_0^t\} - \hat{y}_t \hat{y}_t^{\#}\}$$

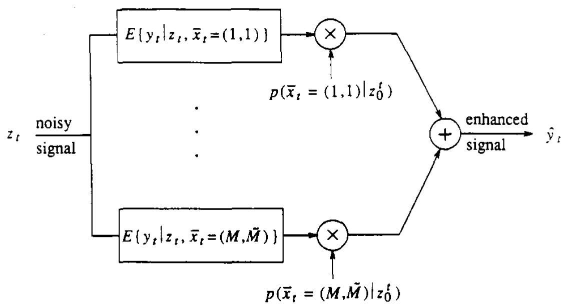

Fig. 1. Causal mmse estimator.

$$\begin{aligned}
&= \frac{1}{K} \operatorname{tr} E\left\{\sum_{\bar{x}_t} p(\bar{x}_t|z_0^t) E\{y_t y_t^{\#}|\bar{x}_t, z_t\} - \hat{y}_t \hat{y}_t^{\#}\right\} \\
&= \frac{1}{K} \operatorname{tr} E\left\{\sum_{\bar{x}_t} p(\bar{x}_t|z_0^t) \operatorname{cov}(y_t|\bar{x}_t, z_t) \right. \\
&\quad \left. + \sum_{\bar{x}_t, \bar{s}_t} p(\bar{x}_t|z_0^t) p(\bar{s}_t|z_0^t) \hat{y}_{t|\bar{x}_t}\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t}\right)^{\#}\right\} \\
&= \frac{1}{K} \operatorname{tr} E\{\operatorname{cov}(y_t|\bar{x}_t, z_t)\} \\
&\quad + \frac{1}{2}\frac{1}{K} \operatorname{tr} \sum_{\bar{x}_t, \bar{s}_t} E\left\{p(\bar{x}_t|z_0^t) p(\bar{s}_t|z_0^t) \right. \\
&\quad \left. \cdot\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t}\right)\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t}\right)^{\#}\right\},
\end{aligned} \tag{13}$$

where # denotes vector transpose and $\bar{s}_t$ is defined similarly to $\bar{x}_t$. Define

$$\overline{\xi_t^2} \triangleq \frac{1}{K} \operatorname{tr} E\{\operatorname{cov}(y_t|\bar{x}_t, z_t)\}, \tag{14}$$

$$g(\bar{x}_t, \bar{s}_t, z_t) \triangleq \frac{1}{K} \operatorname{tr}\left\{\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t}\right)\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t}\right)^{\#}\right\}, \tag{15}$$

$$\begin{aligned}
\overline{\eta_t^2} &\triangleq \sum_{\bar{x}_t, \bar{s}_t} E\{p(\bar{x}_t|z_0^t) p(\bar{s}_t|z_0^t) g(\bar{x}_t, \bar{s}_t, z_t)\} \\
&\triangleq \sum_{\bar{x}_t, \bar{s}_t} \delta_{\bar{x}_t \bar{s}_t}(K).
\end{aligned} \tag{16}$$

Hence,

$$\overline{\epsilon_t^2} = \overline{\xi_t^2} + \frac{1}{2}\overline{\eta_t^2}. \tag{17}$$

Equation (17) shows that the mmse $\overline{\epsilon_t^2}$ can be decomposed into two terms, $\overline{\xi_t^2}$ and $\overline{\eta_t^2}$. The first is the mmse of the estimator $\hat{y}_{t|\bar{x}_t}$ which is informed of the exact composite state of z_t. Since $\hat{y}_{t|\bar{x}_t}$ is a "completely informed" mmse estimator, $\overline{\xi_t^2}$ represents the minimum achievable mse among all estimators in general and informed estimators in particular. The second term $\overline{\eta_t^2}$ represents the sum of cross error terms which depend on pairs of composite states of the noisy signal. Since this term is difficult to evaluate even for CSM's with Gaussian subsources (see Section III), it will be bounded from above and below, and the asymptotic bounds obtained when $K \to \infty$ will be studied. Thus, upper and lower bounds on the mmse $\overline{\epsilon_t^2}$ can be obtained by adding the upper and lower bounds on $\overline{\eta_t^2}$, respectively, to the mmse of the completely informed estimator $\overline{\xi_t^2}$.

In developing the bounds on $\overline{\eta_t^2}$ we shall make the following assumptions:

1) $E\{y_t^{\#} y_t\} < \infty$,
2) $b(y_t|x_t)$ and $b(y_t|\bar{x}_t)$ are such that $\hat{y}_{t|\bar{x}_t}$ is square integrable with respect to $b(z_t|\bar{x}_t)$,
3) $a_{x_{t-1}x_t}$ and $a_{\bar{x}_{t-1}\bar{x}_t}$ are time invariant, $a_{x_{t-1}x_t} \geq a_{\min}^{1/2} > 0$ for all x_{t-1}, x_t, and t, and $a_{\bar{x}_{t-1}\bar{x}_t} \geq a_{\min}^{1/2} > 0$ for all $\bar{x}_{t-1}$, $\bar{x}_t$, and t.

Assumption 1) implies that $\overline{\epsilon_t^2} < \infty$, since under this condition an estimator that results in finite mmse can always be found, e.g., $\hat{y}_t = E\{y_t\}$. Assumption 2) implies that $g(\bar{x}_t, \bar{s}_t, z_t)$ defined in (15) is integrable with respect to $b(z_t|\bar{s}_t)$. Finally, Assumption 3), together with (4), imply that $a_{\bar{x}_{t-1}\bar{x}_t} \geq a_{\min}^{1/2} > 0$ for all $\bar{x}_{t-1}$, $\bar{x}_t$, and t. Hence, from (6)–(8) we have that

$$p(\bar{x}_t|z_0^t) \leq a_{\min}^{-1} \frac{b(z_t|\bar{x}_t)}{\sum_{\bar{x}_t} b(z_t|\bar{x}_t)} \leq a_{\min}^{-1} \frac{b(z_t|\bar{x}_t)}{\max_{\bar{x}_t} b(z_t|\bar{x}_t)}, \tag{18}$$

$$p(\bar{x}_t|z_0^t) \geq a_{\min} \frac{b(z_t|\bar{x}_t)}{\sum_{\bar{x}_t} b(z_t|\bar{x}_t)} \geq \frac{a_{\min}}{\bar{M}} \frac{b(z_t|\bar{x}_t)}{\max_{\bar{x}_t} b(z_t|\bar{x}_t)}. \tag{19}$$

In deriving the bounds on $\overline{\eta_t^2}$ we shall use the following notation:

$$l(z_t) \triangleq \ln \frac{b(z_t|\bar{x}_t)}{b(z_t|\bar{s}_t)}, \tag{20}$$

$$\begin{aligned}
\Omega_{\bar{x}_t} &\triangleq \{z_t : l(z_t) > 0\}, \\
\Omega_{\bar{s}_t} &\triangleq R^K \setminus \Omega_{\bar{x}_t} = R^K \cap \Omega_{\bar{x}_t}^c,
\end{aligned} \tag{21}$$

$$I_{\bar{x}_t}(\bar{s}_t) \triangleq \int_{\Omega_{\bar{x}_t}} b(z_t|\bar{s}_t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t. \tag{22}$$

We first show that both the upper and lower bounds on $\overline{\eta_t^2}$ depend only on $I_{\bar{x}_t}(\bar{s}_t)$ and $I_{\bar{s}_t}(\bar{x}_t)$, and then we develop upper and lower bounds on those integrals. Note that if $g(\bar{x}_t, \bar{s}_t, z_t)$ in (22) is replaced by a unity, then $I_{\bar{x}_t}(\bar{s}_t)$ is the probability of misclassification of the state $\bar{s}_t$ as the state $\bar{x}_t$. Hence, the problem is essentially that of developing bounds for the error probability in classification systems.

C. Upper and Lower bounds on $\overline{\eta_t^2}$

The upper bound on $\overline{\eta_t^2}$ is obtained from an upper bound on $\delta_{\bar{x}_t \bar{s}_t}(K)$. The latter is obtained from (16), (18) and (21) as follows:

$$\begin{aligned}\delta_{\bar{x}_t\bar{s}_t}(K) &= E\{p(\bar{x}_t|z_0^t)p(\bar{s}_t|z_0^t)g(\bar{x}_t,\bar{s}_t,z_t)\}\\ &\le a_{\min}^{-2} E\left\{\frac{b(z_t|\bar{x}_t)b(z_t|\bar{s}_t)}{\left[\max_{\bar{x}_t} b(z_t|\bar{x}_t)\right]^2} g(\bar{x}_t,\bar{s}_t,z_t)\right\}\\ &= a_{\min}^{-2}\sum_{\bar{u}_t} p(\bar{u}_t)\int_{R^K} b(z_t|\bar{u}_t)\frac{b(z_t|\bar{x}_t)b(z_t|\bar{s}_t)}{\left[\max_{\bar{x}_t} b(z_t|\bar{x}_t)\right]^2}\\ &\quad\cdot g(\bar{x}_t,\bar{s}_t,z_t)\,dz_t\\ &\le a_{\min}^{-2}\int_{R^K}\frac{b(z_t|\bar{x}_t)b(z_t|\bar{s}_t)}{\max_{\bar{x}_t} b(z_t|\bar{x}_t)} g(\bar{x}_t,\bar{s}_t,z_t)\,dz_t\\ &\le a_{\min}^{-2}\int_{\Omega_{\bar{x}_t}} b(z_t|\bar{s}_t)g(\bar{x}_t,\bar{s}_t,z_t)\,dz_t\\ &\quad + a_{\min}^{-2}\int_{\Omega_{\bar{s}_t}} b(z_t|\bar{x}_t)g(\bar{x}_t,\bar{s}_t,z_t)\,dz_t\\ &= a_{\min}^{-2}\left[I_{\bar{x}_t}(\bar{s}_t) + I_{\bar{s}_t}(\bar{x}_t)\right].\end{aligned} \tag{23}$$

The lower bound on $\overline{\eta_t^2}$ cannot be straightforwardly obtained from a lower bound on $\delta_{\bar{x}_t\bar{s}_t}(K)$, since the latter is difficult to derive when the number of composite states is greater than two. To derive the desired bound, we study the performance of a partially informed mmse estimator of y_t that outperforms the completely uninformed mmse estimator $\hat{y}_t$. The partially informed estimator chosen here is provided with the information that the composite state $\bar{x}_t$ can take one of two possible values, say $\overline{\alpha}$ and $\overline{\beta}$. The pair $(\overline{\alpha}, \overline{\beta})$ is randomly chosen according to some probability measure, defined on the space of all possible $\overline{M} \times (\overline{M} - 1)$ different pairs of composite states, which agrees with the marginal probability measures $p(\bar{x}_t = \overline{\alpha})$ and $p(\bar{x}_t = \overline{\beta})$. The mmse of the partially informed estimator is obviously obtained from the expected value of the squared error over all realizations of clean and noisy signal vectors as well as all possible pairs of states $(\overline{\alpha}, \overline{\beta})$. We show that

$$\overline{\epsilon_t^2} \ge \overline{\xi_t^2} + \frac{1}{2}\overline{\zeta_t^2}, \tag{24}$$

where $\overline{\xi_t^2}$ is the mmse of the completely informed estimator (14), and $\overline{\zeta_t^2}$ is the expected value of the sum of cross error terms $\delta_{\bar{x}_t\bar{s}_t}(K)$ obtained under the assumption that $\bar{x}_t, \bar{s}_t \in \{\overline{\alpha}, \overline{\beta}\}$. The expected value is taken with respect to the probability measure of the pairs of composite states. Comparing (24) with (17) shows that $\overline{\eta_t^2} \ge \overline{\zeta_t^2}$. Hence, a lower bound on $\overline{\eta_t^2}$ can be obtained from a lower bound on $\delta_{\bar{x}_t\bar{s}_t}(K)$ assuming only two composite states for z_t.

Let $\hat{y}_{t|\overline{\alpha}\overline{\beta}}$ be the mmse estimator of y_t given z_0^t and the pair $\{\overline{\alpha}, \overline{\beta}\}$ of possible composite states for z_t:

$$\begin{aligned}\hat{y}_{t|\overline{\alpha}\overline{\beta}} &= E\{y_t|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t\}\\ &= \sum_{\bar{x}_t\in\{\overline{\alpha},\overline{\beta}\}} p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t)\hat{y}_{t|\bar{x}_t},\end{aligned} \tag{25}$$

where $p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha}, \overline{\beta}\}, z_0^t)$ is the restriction of $p(\bar{x}_t|z_0^t)$ to $\bar{x}_t \in \{\overline{\alpha}, \overline{\beta}\}$ defined by

$$p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t) = \frac{p(\bar{x}_t|z_0^t)}{\sum_{\bar{x}_t\in\{\overline{\alpha},\overline{\beta}\}} p(\bar{x}_t|z_0^t)}, \tag{26}$$

if $\bar{x}_t \in \{\overline{\alpha}, \overline{\beta}\}$, and $p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha}, \overline{\beta}\}, z_0^t) = 0$, otherwise. For this estimator, we have the following orthogonality principle:

$$E\{(y_t - \hat{y}_{t|\overline{\alpha}\overline{\beta}})\hat{y}^{\#}_{t|\overline{\alpha}\overline{\beta}}\} = 0, \tag{27}$$

which results from $E\{(y_t - \hat{y}_{t|\overline{\alpha}\overline{\beta}})\hat{y}^{\#}_{t|\overline{\alpha}\overline{\beta}}|\bar{x}_t \in \{\overline{\alpha}, \overline{\beta}\}, z_0^t\} = 0$. Following a procedure similar to (13), using (27), (26), and (25) in that order, we obtain the following lower bound on $\overline{\epsilon_t^2}$:

$$\begin{aligned}\overline{\epsilon_t^2} &\ge \frac{1}{K}\operatorname{tr} E\{(y_t - \hat{y}_{t|\overline{\alpha}\overline{\beta}})(y_t - \hat{y}_{t|\overline{\alpha}\overline{\beta}})^{\#}\}\\ &= \overline{\xi_t^2} + \frac{1}{K}\operatorname{tr} E\{\hat{y}_{t|\bar{x}_t}\hat{y}^{\#}_{t|\bar{x}_t} - \hat{y}_{t|\overline{\alpha}\overline{\beta}}\hat{y}^{\#}_{t|\overline{\alpha}\overline{\beta}}\}\\ &= \overline{\xi_t^2} + \frac{1}{K}\operatorname{tr} E\{E\{\hat{y}_{t|\bar{x}_t}\hat{y}^{\#}_{t|\bar{x}_t}|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_t\}\\ &\quad - \hat{y}_{t|\overline{\alpha}\overline{\beta}}\hat{y}^{\#}_{t|\overline{\alpha}\overline{\beta}}\}\\ &= \overline{\xi_t^2} + \frac{1}{K}\operatorname{tr} E\left\{\sum_{\bar{x}_t,\bar{s}_t\in\{\overline{\alpha},\overline{\beta}\}} p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t)\right.\\ &\quad\left.\cdot p(\bar{s}_t|\bar{s}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t)\hat{y}_{t|\bar{x}_t}(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{s}_t})^{\#}\right\}\\ &= \overline{\xi_t^2} + \frac{1}{2}E_{\overline{\alpha}\overline{\beta}}\left\{\sum_{\bar{x}_t,\bar{s}_t\in\{\overline{\alpha},\overline{\beta}\}} E\{p(\bar{x}_t|\bar{x}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t)\right.\\ &\quad\left.\cdot p(\bar{s}_t|\bar{s}_t \in \{\overline{\alpha},\overline{\beta}\}, z_0^t)g(\bar{x}_t,\bar{s}_t,z_t)|\overline{\alpha},\overline{\beta}\}\right\}\\ &= \overline{\xi_t^2} + \frac{1}{2}E_{\overline{\alpha}\overline{\beta}}\left\{\sum_{\bar{x}_t,\bar{s}_t\in\{\overline{\alpha},\overline{\beta}\}} \delta_{\bar{x}_t\bar{s}_t}(K)\right\}\\ &\triangleq \overline{\xi_t^2} + \frac{1}{2}\overline{\zeta_t^2},\end{aligned} \tag{28}$$

where $E_{\overline{\alpha},\overline{\beta}}$ is the expected value with respect to the probability measure defined over pairs of different composite states.

The lower bound on $\delta_{\bar{x}_t\bar{s}_t}(K)$, $\bar{x}_t, \bar{s}_t \in \{\overline{\alpha}, \overline{\beta}\}$, is obtained as follows. We assume, without loss of generality, that $\bar{x}_t \ne \bar{s}_t$ since $g(\bar{x}_t, \bar{s}_t, z_t) = 0$, and hence $\delta_{\bar{x}_t\bar{s}_t}(K) = 0$, whenever $\bar{x}_t = \bar{s}_t$. Furthermore, since the lower bound in (28) depends only on $\delta_{\overline{\alpha}\overline{\beta}}(K)$ and $\delta_{\overline{\beta}\overline{\alpha}}(K)$, and $\delta_{\overline{\alpha}\overline{\beta}}(K) = \delta_{\overline{\beta}\overline{\alpha}}(K)$, we can assume, without loss of generality, that

$\bar{x}_t = \bar{\alpha}$ and $\bar{s}_t = \bar{\beta}$. Using $\bar{M} = 2$, we have from (19) for $\bar{x}_t$ (and similarly for $\bar{s}_t$) that

$$p(\bar{x}_t | \bar{x}_t \in \{\bar{\alpha}, \bar{\beta}\}, z_0^t) \geq \frac{a_{\min}}{2} \frac{b(z_t|\bar{x}_t)}{\max\{b(z_t|\bar{\alpha}), b(z_t|\bar{\beta})\}}. \tag{29}$$

Hence, for $\bar{x}_t = \bar{\alpha}$ and $\bar{s}_t = \bar{\beta}$ we have that

$$\begin{aligned}
\delta_{\bar{x}_t \bar{s}_t}(K) &= E\{p(\bar{x}_t|\bar{x}_t \in \{\bar{\alpha}, \bar{\beta}\}, z_0^t) p(\bar{s}_t|\bar{s}_t \in \{\bar{\alpha}, \bar{\beta}\}, z_0^t) \\
&\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t) | \bar{\alpha}, \bar{\beta}\} \\
&\geq \frac{a_{\min}^2}{4} E\left\{ \frac{b(z_t|\bar{x}_t) b(z_t|\bar{s}_t)}{[\max\{b(z_t|\bar{\alpha}), b(z_t|\bar{\beta})\}]^2} \right. \\
&\quad \left. \cdot g(\bar{x}_t, \bar{s}_t, z_t) | \bar{\alpha}, \bar{\beta} \right\} \\
&= \frac{a_{\min}^2}{4} \sum_{\bar{u}_t \in \{\bar{\alpha}, \bar{\beta}\}} p(\bar{u}_t) \int_{R^K} \\
&\quad \cdot b(z_t|\bar{u}_t) \frac{b(z_t|\bar{x}_t) b(z_t|\bar{s}_t)}{[\max\{b(z_t|\bar{\alpha}), b(z_t|\bar{\beta})\}]^2} \\
&\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t) \, dz_t |_{\{\bar{x}_t = \bar{\alpha}, \bar{s}_t = \bar{\beta}\}} \\
&= \frac{a_{\min}^2}{4} p(\bar{\alpha}) \int_{R^K} \frac{b(z_t|\bar{\alpha}) b(z_t|\bar{x}_t) b(z_t|\bar{s}_t)}{[\max\{b(z_t|\bar{\alpha}), b(z_t|\bar{\beta})\}]^2} \\
&\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t) \, dz_t |_{\{\bar{x}_t = \bar{\alpha}, \bar{s}_t = \bar{\beta}\}} \\
&\quad + \frac{a_{\min}^2}{4} p(\bar{\beta}) \int_{R^K} \frac{b(z_t|\bar{\beta}) b(z_t|\bar{x}_t) b(z_t|\bar{s}_t)}{[\max\{b(z_t|\bar{\alpha}), b(z_t|\bar{\beta})\}]^2} \\
&\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t) \, dz_t |_{\{\bar{x}_t = \bar{\alpha}, \bar{s}_t = \bar{\beta}\}} \\
&\geq \frac{a_{\min}^2}{4} p(\bar{\alpha}) \int_{\Omega_{\bar{\alpha}}} b(z_t|\bar{\beta}) g(\bar{\alpha}, \bar{\beta}, z_t) \, dz_t \\
&\quad + \frac{a_{\min}^2}{4} p(\bar{\beta}) \int_{\Omega_{\bar{\beta}}} b(z_t|\bar{\alpha}) g(\bar{\alpha}, \bar{\beta}, z_t) \, dz_t \\
&= \frac{a_{\min}^2}{4} [p(\bar{\alpha}) I_{\bar{\alpha}}(\bar{\beta}) + p(\bar{\beta}) I_{\bar{\beta}}(\bar{\alpha})].
\end{aligned} \tag{30}$$

D. Upper and Lower Bounds on $I_{\bar{x}_t}(\bar{s}_t)$

We now turn to develop upper and lower bounds for $I_{\bar{x}_t}(\bar{s}_t)$ (or $I_{\bar{s}_t}(\bar{x}_t)$), which appear in the bounds (23) and (30). Define

$$q(z_t|\bar{x}_t, \bar{s}_t) \triangleq \Phi^{-1}(\bar{x}_t, \bar{s}_t) b(z_t|\bar{s}_t) g(\bar{x}_t, \bar{s}_t, z_t), \tag{31}$$

where

$$\Phi(\bar{x}_t, \bar{s}_t) \triangleq \int_{R^K} b(z_t|\bar{s}_t) g(\bar{x}_t, \bar{s}_t, z_t) \, dz_t. \tag{32}$$

By Assumption 2), $\Phi(\bar{x}_t, \bar{s}_t) < \infty$. Hence, $q(z_t|\bar{x}_t, \bar{s}_t)$ is a pdf on R^K since it is a nonnegative function which integrates to one. Expressing $I_{\bar{x}_t}(\bar{s}_t)$ in terms of $q(z_t|\bar{x}_t, \bar{s}_t)$ gives

$$I_{\bar{x}_t}(\bar{s}_t) = \Phi(\bar{x}_t, \bar{s}_t) \int_{\Omega_{\bar{x}_t}} q(z_t|\bar{x}_t, \bar{s}_t) \, dz_t, \tag{33}$$

and the problem becomes that of bounding from above and below the probability of the set $\Omega_{\bar{x}_t}$ with respect to $q(z_t|\bar{x}_t, \bar{s}_t)$, i.e.,

$$J_{\bar{x}_t}(\bar{s}_t) \triangleq \int_{\Omega_{\bar{x}_t}} q(z_t|\bar{x}_t, \bar{s}_t) \, dz_t. \tag{34}$$

This is done by using distribution tilting (see, e.g., [3], [29]), and Shannon's lower bound (see, e.g., [18, Lemma 5]) on the probability of a random variable to exceed a given threshold.

Let $q(l(z_t)|\bar{x}_t, \bar{s}_t)$ be the pdf of $l(z_t)$ as can be obtained from (20) and (31). Define the tilted pdf of $l(z_t)$ as

$$q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t) = e^{-\mu(\lambda)} e^{\lambda l(z_t)} q(l(z_t)|\bar{x}_t, \bar{s}_t), \quad \lambda > 0, \tag{35}$$

where

$$\mu(\lambda) \triangleq \ln \int_{R^K} e^{\lambda l(z_t)} q(l(z_t)|\bar{x}_t, \bar{s}_t) \, dl(z_t) \tag{36}$$

is the logarithm of the moment generating function of $l(z_t)$ with respect to q or the semi-invariant moment generating function of $l(z_t)$ [29, p. 188]. Since $\mu(\lambda)$ is the logarithm of the expected value of some function of z_t, it can be evaluated by

$$\mu(\lambda) = \ln \int_{R^K} e^{\lambda l(z_t)} q(z_t|\bar{x}_t, \bar{s}_t) \, dz_t. \tag{37}$$

By substituting (20) and (31)–(32) into (37), it can be shown (see Appendix) that $\mu(\lambda) < \infty$ for $0 \leq \lambda \leq 1$. For $\lambda \notin [0, 1]$, $\mu(\lambda)$ may still be finite, depending on the specific pdf's $b(z_t|\bar{x}_t)$ and $b(z_t|\bar{s}_t)$ of the CSM's. This is demonstrated in the next section, where we discuss CSM's with Gaussian subsources. Nevertheless, the case where $\lambda \in [0, 1]$ will be of particular interest, since convergence of the bounds can be proved for λ within this interval. We also have the following useful relations:

$$\begin{aligned}
\dot{\mu}(\lambda) &\triangleq \frac{\partial}{\partial \lambda} \mu(\lambda) = E_{q_\lambda}\{l(z_t)\}, \\
\ddot{\mu}(\lambda) &\triangleq \frac{\partial^2}{\partial \lambda^2} \mu(\lambda) = \mathrm{var}_{q_\lambda}\{l(z_t)\},
\end{aligned} \tag{38}$$

where $E_{q_\lambda}\{\cdot\}$ and $\mathrm{var}_{q_\lambda}\{\cdot\}$ are the expected value and variance with respect to q_λ, respectively.

The upper bound on $J_{\bar{x}_t}(\bar{s}_t)$ is obtained from (21), (34), and (35) as follows:

$$\begin{aligned} J_{\bar{x}_t}(\bar{s}_t) &= \int_{\Omega_{\bar{x}_t}} q(z_t|\bar{x}_t, \bar{s}_t)\, dz_t \\ &= \int_0^\infty q(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &= e^{\mu(\lambda)} \int_0^\infty e^{-\lambda l(z_t)} q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &\le e^{\mu(\lambda)} \int_0^\infty q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &\le e^{\mu(\lambda)} \int_{-\infty}^\infty q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &= e^{\mu(\lambda)}. \end{aligned} \tag{39}$$

Hence,

$$I_{\bar{x}_t}(\bar{s}_t) \le \Phi(\bar{x}_t, \bar{s}_t) e^{\mu(\lambda)}, \qquad \lambda > 0. \tag{40}$$

To obtain the lower bound, assume that there exists λ such that $\dot{\mu}(\lambda) > 0$. Let γ and ρ be two numbers that satisfy

$$\gamma > \dot{\mu}(\lambda) > \rho \ge 0. \tag{41}$$

Define

$$\Omega_{\bar{x}_t}(\rho) = \{z_t : l(z_t) > \rho\}. \tag{42}$$

Since $\Omega_{\bar{x}_t}(\rho) \subseteq \Omega_{\bar{x}_t}$, as defined in (21) and (35), we have following [18, Lemma 5] that

$$\begin{aligned} J_{\bar{x}_t}(\bar{s}_t) &\ge \int_{\Omega_{\bar{x}_t}(\rho)} q(z_t|\bar{x}_t, \bar{s}_t)\, dz_t \\ &= \int_\rho^\infty q(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &= e^{\mu(\lambda)} \int_\rho^\infty e^{-\lambda l(z_t)} q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &\ge e^{\mu(\lambda)} \int_\rho^\gamma e^{-\lambda l(z_t)} q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &\ge e^{\mu(\lambda) - \lambda\gamma} \int_\rho^\gamma q_\lambda(l(z_t)|\bar{x}_t, \bar{s}_t)\, dl(z_t) \\ &= e^{\mu(\lambda) - \lambda\gamma} \Pr\{\rho \le l(z_t) \le \gamma\}. \end{aligned} \tag{43}$$

By way of choosing ρ and γ we can assume that

$$\begin{aligned} \gamma &= \dot{\mu}(\lambda) + \delta \\ \rho &= \dot{\mu}(\lambda) - \delta, \qquad \delta > 0. \end{aligned} \tag{44}$$

Hence, by applying the Chebycheff inequality to (43) and using (33) and (38), we obtain

$$I_{\bar{x}_t}(\bar{s}_t) \ge \Phi(\bar{x}_t, \bar{s}_t) e^{\mu(\lambda) - \lambda(\dot{\mu}(\lambda) + \delta)} \left(1 - \frac{\ddot{\mu}(\lambda)}{\delta^2}\right). \tag{45}$$

The bound is useful if $\ddot{\mu}(\lambda) < \delta^2$. Combining this condition with (41) for $\rho = \dot{\mu}(\lambda) - \delta$, we obtain that (45) is useful for any λ that satisfies

$$\dot{\mu}(\lambda) > \delta > (\ddot{\mu}(\lambda))^{1/2}. \tag{46}$$

From (40) and (45), we see that the upper and lower bounds on $I_{\bar{x}_t}(\bar{s}_t)$, and hence, on $\overline{\eta_t^2}$, depend on $\Phi(\bar{x}_t, \bar{s}_t)$, and on the semi-invariant moment generating function $\mu(\lambda)$ and its first two derivatives. In the next section, we explicitly calculate those functions for CSM's with AWS Gaussian subsources that have mostly been used in practice [2]. For this important case, we show that there exists $0 < \lambda < 1$ which satisfies (46), and that the upper and lower bounds on $I_{\bar{x}_t}(\bar{s}_t)$ converge exponentially to zero at the same rate as the frame length K approaches infinity. This means that the bounds (40) and (45) are tight, and that the mmse $\overline{\epsilon_t^2}$ exponentially approaches the mmse $\overline{\xi_t^2}$ of the completely informed estimator.

III. CSM's with AWS Gaussian Subsources

Consider an M-state discrete time CSM with zero-mean AWS Gaussian subsources for the signal and an $\tilde{M}$-state CSM with zero-mean AWS Gaussian subsources for the noise. In this case, $b(y_t|x_t)$ and $b(y_t|\bar{x}_t)$ are zero-mean Gaussian pdf's with asymptotically Toeplitz covariance matrices S_{x_t} and $S_{\bar{x}_t}$, respectively. A covariance matrix, say S_{x_t}, is asymptotically Toeplitz if there exists a sequence of nested Toeplitz covariance matrices $T_K(S_{x_t}(\theta))$, where $S_{x_t}(\theta) \le U < \infty$ denotes the asymptotic power spectral density of the subsource x_t, and θ is the angular frequency, such that S_{x_t} and $T_K(S_{x_t}(\theta))$ are uniformly bounded in the strong norm, and $S_{x_t} \to T_K(S_{x_t}(\theta))$ as $K \to \infty$ in the weak or Hilbert–Schmidt norm [14], [15]. The fact that S_{x_t} and $S_{\bar{x}_t}$ are asymptotically Toeplitz is denoted by

$$\begin{aligned} S_{x_t} &\sim T_K(S_{x_t}(\theta)), \\ S_{\bar{x}_t} &\sim T_K(S_{\bar{x}_t}(\theta)). \end{aligned}$$

Assume that $S_{x_t}(\theta) \ge m > 0$ and $S_{\bar{x}_t}(\theta) \ge m > 0$; so that inverses of S_{x_t} and $S_{\bar{x}_t}$ are also asymptotically Toeplitz [14], [15].

Under these assumptions, $b(z_t|\bar{x}_t)$ is zero-mean Gaussian with covariance $Q_{\bar{x}_t} \triangleq S_{x_t} + S_{\bar{x}_t}$, and $Q_{\bar{x}_t}$ is asymptotically Toeplitz with power spectral density $Q_{\bar{x}_t}(\theta) = S_{x_t}(\theta) + S_{\bar{x}_t}(\theta)$. Furthermore,

$$\begin{aligned} \hat{y}_{t|\bar{x}_t} &= E\{y_t|\bar{x}_t, z_t\} \\ &= S_{x_t}(S_{x_t} + S_{\bar{x}_t})^{-1} z_t \\ &\triangleq H_{\bar{x}_t} z_t, \end{aligned} \tag{47}$$

where $H_{\bar{x}_t}$ is the Wiener filter for the output processes from the signal state x_t and the noise state $\bar{x}_t$. $H_{\bar{x}_t}$ is asymptotically Toeplitz with power spectral density given by

$$H_{\bar{x}_t}(\theta) = \frac{S_{x_t}(\theta)}{S_{x_t}(\theta) + S_{\bar{x}_t}(\theta)}. \tag{48}$$

$H_{\bar{x}_t}(\theta)$ is often referred to as the frequency response of $H_{\bar{x}_t}$. The conditional covariance of y_t given $\bar{x}_t$ and z_t is given by

$$\operatorname{cov}(y_t|\bar{x}_t, z_t) = H_{\bar{x}_t} S_{\bar{x}_t}, \tag{49}$$

and it is independent of z_t. This covariance is also asymptotically Toeplitz with power spectral density $H_{\bar{x}_t}(\theta)S_{\bar{x}_t}(\theta)$. Note that $\operatorname{cov}(y_t|\bar{x}_t, z_t)$ is the mmse in estimating y_t from z_t given $\bar{x}_t$.

The mmse of the completely informed estimator is obtained from (14) and (49),

$$\overline{\xi_t^2} = \frac{1}{K} \operatorname{tr} \sum_{\bar{x}_t} p(\bar{x}_t) H_{\bar{x}_t} S_{\bar{x}_t}. \tag{50}$$

This constitutes the statistical average of the mmse obtained under explicit knowledge of the composite state of z_t. Applying the Toeplitz distribution theorem [14]–[15] to (49)–(50) results in the asymptotic mmse of the completely informed estimator given by

$$\lim_{K\to\infty} \overline{\xi_t^2} = \sum_{\bar{x}_t} p(\bar{x}_t) \int_0^{2\pi} H_{\bar{x}_t}(\theta) S_{\bar{x}_t}(\theta) \frac{d\theta}{2\pi}. \tag{51}$$

The upper bound on $\overline{\eta_t^2}$ can be obtained from (16), (23), and (40), and the lower bound from (28), (30), and (45). In both cases, we have to calculate the upper and lower bounds on $I_{\bar{x}_t}(\bar{s}_t)$ given in (40) and (45), respectively. These bounds depend on $\Phi(\bar{x}_t, \bar{s}_t)$, which is given in (32), on $\mu(\lambda)$ given in (37), and on $\dot{\mu}(\lambda)$ and $\ddot{\mu}(\lambda)$. These functions are now evaluated for the CSM's with AWS Gaussian subsources considered here.

From (15) and (47), we have that

$$\begin{aligned} g(\bar{x}_t, \bar{s}_t, z_t) &= \frac{1}{K} z_t^{\#} \left(H_{\bar{x}_t} - H_{\bar{s}_t}\right)^{\#} \left(H_{\bar{x}_t} - H_{\bar{s}_t}\right) z_t \\ &\triangleq \frac{1}{K} z_t^{\#} H_{\bar{x}_t \bar{s}_t}^2 z_t, \end{aligned} \tag{52}$$

where $H_{\bar{x}_t\bar{s}_t}^2$ is asymptotically Toeplitz with power spectral density

$$H_{\bar{x}_t\bar{s}_t}^2(\theta) = |H_{\bar{x}_t}(\theta) - H_{\bar{s}_t}(\theta)|^2. \tag{53}$$

Hence, from (32) and the fact that $b(z_t|\bar{s}_t)$ is zero-mean Gaussian with covariance $Q_{\bar{s}_t}$ we obtain

$$\Phi(\bar{x}_t, \bar{s}_t) = \frac{1}{K} \operatorname{tr} \left\{ H_{\bar{x}_t\bar{s}_t}^2 Q_{\bar{s}_t} \right\}. \tag{54}$$

Using the Toeplitz distribution theorem [14]–[15], we have

$$\begin{aligned} \Phi_\infty(\bar{x}_t, \bar{s}_t) &\triangleq \lim_{K\to\infty} \Phi(\bar{x}_t, \bar{s}_t) \\ &= \int_0^{2\pi} H_{\bar{x}_t\bar{s}_t}^2(\theta) Q_{\bar{s}_t}(\theta) \frac{d\theta}{2\pi}. \end{aligned} \tag{55}$$

By substituting (20), (31), and (52) in (37), we obtain

$$\begin{aligned} \mu(\lambda) = \ln \int_{R^K} & b^\lambda(z_t|\bar{x}_t) b^{1-\lambda}(z_t|\bar{s}_t) z_t^{\#} H_{\bar{x}_t\bar{s}_t}^2 z_t \, dz_t \\ & - \ln\{K\Phi(\bar{x}_t, \bar{s}_t)\}. \end{aligned} \tag{56}$$

Note that

$$b^\lambda(z_t|\bar{x}_t) b^{1-\lambda}(z_t|\bar{s}_t) = \frac{|R_\lambda(\bar{x}_t, \bar{s}_t)|^{1/2}}{|Q_{\bar{x}_t}|^{\frac{\lambda}{2}} |Q_{\bar{s}_t}|^{\frac{1-\lambda}{2}}} \cdot \frac{\exp\left\{-\frac{1}{2} z_t^{\#} R_\lambda^{-1}(\bar{x}_t, \bar{s}_t) z_t\right\}}{(2\pi)^{K/2} |R_\lambda(\bar{x}_t, \bar{s}_t)|^{1/2}}, \tag{57}$$

where

$$R_\lambda^{-1}(\bar{x}_t, \bar{s}_t) \triangleq \lambda Q_{\bar{x}_t}^{-1} + (1-\lambda) Q_{\bar{s}_t}^{-1}, \tag{58}$$

provided that $|R_\lambda(\bar{x}_t, \bar{s}_t)| > 0$. Hence, if $R_\lambda(\bar{x}_t, \bar{s}_t)$ is positive definite, then $b^\lambda(z_t|\bar{x}_t) b^{1-\lambda}(z_t|\bar{s}_t)$ is proportional to a zero-mean Gaussian pdf with covariance $R_\lambda(\bar{x}_t, \bar{s}_t)$. The values of λ that satisfy this condition are obtained as follows. Let $\sigma(Q_{\bar{x}_t})$ be an eigenvalue of $Q_{\bar{x}_t}$ and assume that $\sigma_{\min}(Q_{\bar{x}_t}) \le \sigma(Q_{\bar{x}_t}) \le \sigma_{\max}(Q_{\bar{x}_t})$. From [19, p. 285], $\sigma_{\min}(Q_{\bar{x}_t})$ and $\sigma_{\max}(Q_{\bar{x}_t})$ are, respectively, the minimum and maximum values of the Rayleigh quotient of the symmetric matrix $Q_{\bar{x}_t}$. Define

$$A(\bar{x}_t, \bar{s}_t) \triangleq \frac{\sigma_{\max}^{-1}(Q_{\bar{s}_t})}{\sigma_{\max}^{-1}(Q_{\bar{x}_t}) - \sigma_{\min}^{-1}(Q_{\bar{s}_t})}.$$

It is easy to show that $R_\lambda(\bar{x}_t, \bar{s}_t)$ is positive definite if $\lambda > -A(\bar{x}_t, \bar{s}_t)$ provided that $\sigma_{\max}^{-1}(Q_{\bar{x}_t}) > \sigma_{\min}^{-1}(Q_{\bar{s}_t})$ or $\lambda < -A(\bar{x}_t, \bar{s}_t)$ assuming that $\sigma_{\max}^{-1}(Q_{\bar{x}_t}) < \sigma_{\min}^{-1}(Q_{\bar{s}_t})$. If $\sigma_{\max}^{-1}(Q_{\bar{x}_t}) = \sigma_{\min}^{-1}(Q_{\bar{s}_t})$, then $R_\lambda(\bar{x}_t, \bar{s}_t)$ is positive definite for all $\lambda > 0$. The matrix $R_\lambda(\bar{x}_t, \bar{s}_t)$ is asymptotically Toeplitz with power spectral density given by

$$R_\lambda^{-1}(\theta) \triangleq \lambda Q_{\bar{x}_t}^{-1}(\theta) + (1-\lambda) Q_{\bar{s}_t}^{-1}(\theta). \tag{59}$$

On substituting (54) and (57) into (56), we obtain

$$\mu(\lambda) = \ln \frac{|R_\lambda(\bar{x}_t, \bar{s}_t)|^{1/2}}{|Q_{\bar{x}_t}|^{\frac{\lambda}{2}} |Q_{\bar{s}_t}|^{\frac{1-\lambda}{2}}} + \ln \frac{\operatorname{tr}\left\{H_{\bar{x}_t\bar{s}_t}^2 R_\lambda(\bar{x}_t, \bar{s}_t)\right\}}{\operatorname{tr}\left\{H_{\bar{x}_t\bar{s}_t}^2 Q_{\bar{s}_t}\right\}}. \tag{60}$$

Applying the Toeplitz distribution theorem to (60) and using Jensen's inequality yield

$$\begin{aligned} \mu_\infty(\lambda) &\triangleq \lim_{K\to\infty} \frac{1}{K} \mu(\lambda) \\ &= -\frac{1}{2} \int_0^{2\pi} \ln\left[\frac{\lambda}{Q_{\bar{x}_t}(\theta)} + \frac{1-\lambda}{Q_{\bar{s}_t}(\theta)}\right] \frac{d\theta}{2\pi} \\ &\quad + \frac{1}{2} \int_0^{2\pi} \left[\lambda \ln \frac{1}{Q_{\bar{x}_t}(\theta)} + (1-\lambda) \ln \frac{1}{Q_{\bar{s}_t}(\theta)}\right] \frac{d\theta}{2\pi} \\ &< 0, \qquad \text{for all} \quad 0 < \lambda < 1. \end{aligned} \tag{61}$$

In evaluating $\dot{\mu}(\lambda)$ it will be convenient to define the "distance"

$$d\left(R_\lambda(\bar{x}_t,\bar{s}_t),Q_{\bar{x}_t}\right) \triangleq \frac{1}{K}\left[\operatorname{tr}\left\{R_\lambda(\bar{x}_t,\bar{s}_t)Q_{\bar{x}_t}^{-1}\right\} - \ln|R_\lambda(\bar{x}_t,\bar{s}_t)Q_{\bar{x}_t}^{-1}| - 1\right]. \quad (62)$$

The asymptotic value of this distance is obtained from applying the Toeplitz distribution theorem to (62). This results in the so called Itakura–Saito distortion measure [20]–[21], [28, pp. 196–199]:

$$d_{\mathrm{IS}}\left(R_\lambda,Q_{\bar{x}_t}\right) \triangleq \lim_{K\to\infty} d\left(R_\lambda(\bar{x}_t,\bar{s}_t),Q_{\bar{x}_t}\right) = \int_0^{2\pi}\left(\frac{R_\lambda(\theta)}{Q_{\bar{x}_t}(\theta)} - \ln\frac{R_\lambda(\theta)}{Q_{\bar{x}_t}(\theta)} - 1\right)\frac{d\theta}{2\pi}. \quad (63)$$

Furthermore, we shall use the identity

$$E\{z_t^{\#}Az_t z_t^{\#}Bz_t\} = \operatorname{tr}\{AR\}\operatorname{tr}\{BR\} + \operatorname{tr}\{(AR)(RB)^{\#}\} + \operatorname{tr}\{(AR)(BR)\}, \quad (64)$$

where A and B are two $K\times K$ matrices, and the expected value is taken with respect to a zero-mean Gaussian pdf with covariance R. Taking the first derivative of (37) we obtain (see Appendix)

$$\dot{\mu}(\lambda) = -\frac{K}{2}\left[d\left(R_\lambda(\bar{x}_t,\bar{s}_t),Q_{\bar{x}_t}\right) - d\left(R_\lambda(\bar{x}_t,\bar{s}_t),Q_{\bar{s}_t}\right)\right] - \frac{\operatorname{tr}\left\{H^2_{\bar{x}_t\bar{s}_t}R_\lambda(\bar{x}_t,\bar{s}_t)\left(Q_{\bar{x}_t}^{-1}-Q_{\bar{s}_t}^{-1}\right)R_\lambda(\bar{x}_t,\bar{s}_t)\right\}}{\operatorname{tr}\left\{H^2_{\bar{x}_t\bar{s}_t}R_\lambda(\bar{x}_t,\bar{s}_t)\right\}}. \quad (65)$$

The asymptotic value of $\dot{\mu}(\lambda)/K$ is given by

$$\dot{\mu}_\infty(\lambda) \triangleq \lim_{K\to\infty}\frac{1}{K}\dot{\mu}(\lambda) = -\frac{1}{2}\left[d_{\mathrm{IS}}\left(R_\lambda,Q_{\bar{x}_t}\right) - d_{\mathrm{IS}}\left(R_\lambda,Q_{\bar{s}_t}\right)\right]. \quad (66)$$

$\ddot{\mu}(\lambda)$ can be evaluated using the following identity, which can be derived using [22, p. 97]:

$$E\left\{(z_t^{\#}Az_t)(z_t^{\#}Bz_t)^2\right\} = \operatorname{tr}\{AR\}[\operatorname{tr}\{BR\}]^2 + 2\operatorname{tr}\{AR\}\operatorname{tr}\left\{(BR)^2\right\} + 8\operatorname{tr}\left\{AR(BR)^2\right\} + 4\operatorname{tr}\{BR\}\operatorname{tr}\{AR\,BR\}, \quad (67)$$

where A and B are $K\times K$ symmetric matrices, and the expected value is taken with respect to a zero-mean Gaussian pdf with covariance matrix R. Taking the second derivative of (37) using (67), and applying the Toeplitz distribution theorem, we obtain (see Appendix)

$$\ddot{\mu}_\infty(\lambda) \triangleq \lim_{K\to\infty}\frac{1}{K^2}\ddot{\mu}(\lambda) = 0. \quad (68)$$

We now show that there exists λ such that (46) is satisfied provided that K is sufficiently large. From (66) and (59), it is easy to show that $\dot{\mu}_\infty(\lambda)$ is continuous on $[0,1]$ and that

$$\dot{\mu}_\infty(\lambda)_{|_{\lambda=0}} = -\frac{1}{2}d_{\mathrm{IS}}\left(Q_{\bar{s}_t},Q_{\bar{x}_t}\right) < 0,$$
$$\dot{\mu}_\infty(\lambda)_{|_{\lambda=1}} = \frac{1}{2}d_{\mathrm{IS}}\left(Q_{\bar{x}_t},Q_{\bar{s}_t}\right) > 0. \quad (69)$$

Hence, there must exist $\lambda^* \in (0,1)$ and $\Delta\lambda > 0$ such that $\dot{\mu}_\infty(\lambda^*) = 0$ and $\dot{\mu}_\infty(\lambda^*+\Delta\lambda) \ge \delta' > 0$, where δ' is independent of K. Note from (66) that R_{λ^*} yields equal Itakura–Saito distances with respect to $Q_{\bar{x}_t}$ and $Q_{\bar{s}_t}$. Combining this result with (68) we obtain, for $\lambda = \lambda^* + \Delta\lambda$,

$$\dot{\mu}_\infty(\lambda) \ge \delta' > (\ddot{\mu}_\infty(\lambda))^{1/2}. \quad (70)$$

Hence, for sufficiently large K there exists λ such that (46) is satisfied with $\delta = K\delta'$. For these K, λ, and δ, the bounds (40) and (45) can be approximated by

$$I_{\bar{x}_t}(\bar{s}_t) \le \Phi_\infty(\bar{x}_t,\bar{s}_t)e^{K\mu_\infty(\lambda)},$$
$$I_{\bar{x}_t}(\bar{s}_t) \ge \Phi_\infty(\bar{x}_t,\bar{s}_t)e^{K\mu_\infty(\lambda)}e^{-\lambda K(\dot{\mu}_\infty(\lambda)+\delta')}. \quad (71)$$

If λ is chosen such that $\dot{\mu}_\infty(\lambda) + \delta' \cong 0$, then the upper and lower bounds are essentially the same. Furthermore, since $\mu_\infty(\lambda) < 0$ for $\lambda \in (0,1)$, both bounds approach zero exponentially. This means that the lower and upper bounds developed here are tight and that the asymptotic mmse converges exponentially with rate $-\mu_\infty(\lambda)$ to the asymptotic mmse of the completely informed estimator given in (51).

IV. MSE in Detection–Estimation

In this section, we study the performance of a suboptimal intuitive estimator which first detects the composite state of the noisy signal and then applies the conditional mean estimator associated with this state to the given noisy signal. Using the notation of Section II, this estimator is given by

$$\hat{y}_t = E\{y_t|\bar{x}_t^*, z_t\} = \hat{y}_{t|\bar{x}_t^*}, \quad (72)$$

where

$$\bar{x}_t^* = \arg\max_{\bar{x}_t} p(\bar{x}_t|z_0^t). \quad (73)$$

A block diagram of this estimator is shown in Fig. 2. We show that the mse associated with this estimator comprises the sum of the mmse $\overline{\xi_t^2}$ of the completely informed estimator, and the expected value of cross error terms that can be bounded similarly to $I_{\bar{x}_t}(\bar{s}_t)$ in Section II. For CSM with AWS Gaussian subsources, this means that the mse of the detection–estimation scheme exponentially converges to the mmse of the completely informed estimator as $K\to\infty$. Hence, for these sources, the

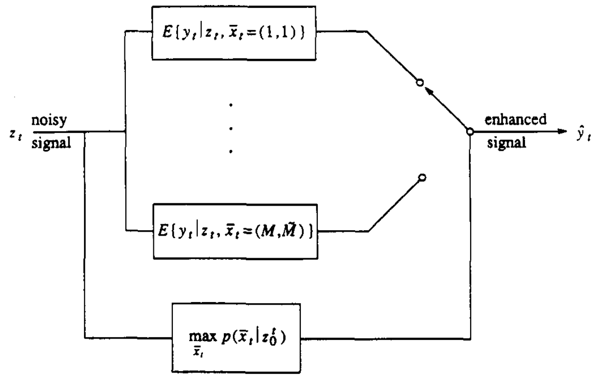

Fig. 2. Detection-Estimation scheme.

detection–estimation scheme is asymptotically optimal in the mmse sense. Convergence of the mse of the detection-estimation scheme to the mmse of the completely informed estimator is not surprising, since if $p(\bar{x}_t|z_0^t)$ approaches one for some $\bar{x}_t$ and zero for all other states as $K \to \infty$, then the estimator (72)–(73) and the mmse estimator are essentially the same. This indeed was shown in [8] to be the case for CSM's with Gaussian ergodic subsources. It is less obvious, however, that the exponential rate of convergence of the mse's associated with the detection-estimation scheme and the mmse estimator should be the same.

The mse associated with the estimator (72)–(73) is calculated using the orthogonality principle

$$E\left\{\left(y_t - \hat{y}_{t|\bar{x}_t}\right)\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{x}_t^*}\right)^{\#} | \bar{x}_t, z_0^t\right\} = 0. \quad (74)$$

Hence, by adding and subtracting $\hat{y}_{t|\bar{x}_t}$ to $y_t - \hat{y}_{t|\bar{x}_t^*}$, we obtain, using (74),

$$\begin{aligned}\overline{\epsilon_t^2} &= \frac{1}{K} \operatorname{tr} E\left\{\left(y_t - \hat{y}_{t|\bar{x}_t^*}\right)\left(y_t - \hat{y}_{t|\bar{x}_t^*}\right)^{\#}\right\} \\ &= \overline{\xi_t^2} + \overline{\zeta_t^2}, \end{aligned} \quad (75)$$

where $\overline{\xi_t^2}$ is the mmse of the completely informed estimator given in (14), and $\overline{\zeta_t^2}$ is defined by

$$\overline{\zeta_t^2} \triangleq E\{g(\bar{x}_t, \bar{x}_t^*, z_0^t)\}, \quad (76)$$

with

$$g(\bar{x}_t, \bar{x}_t^*, z_0^t) \triangleq \frac{1}{K} \operatorname{tr}\left\{\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{x}_t^*}\right)\left(\hat{y}_{t|\bar{x}_t} - \hat{y}_{t|\bar{x}_t^*}\right)^{\#}\right\}. \quad (77)$$

We now develop upper and lower bounds on $\overline{\zeta_t^2}$. Let

$$L(z_0^t) \triangleq \ln \frac{p(\bar{x}_t|z_0^t)}{p(\bar{s}_t|z_0^t)}. \quad (78)$$

Define

$$\begin{aligned}\Theta_{\bar{x}_t}(\rho) &\triangleq \{z_0^t : l(z_t) > \rho, \text{ for all } \quad \bar{s}_t \neq \bar{x}_t\}, \\ \Psi_{\bar{x}_t}(\rho) &\triangleq \{z_0^t : L(z_0^t) > \rho, \text{ for all } \quad \bar{s}_t \neq \bar{x}_t\}. \end{aligned} \quad (79)$$

From (18) and (19), we have that

$$l(z_t) - \rho \le L(z_0^t) \le l(z_t) + \rho, \quad (80)$$

for

$$\rho = -\ln \frac{a_{\min}^2}{M\bar{M}} > 0. \quad (81)$$

Hence,

$$\Theta_{\bar{x}_t}(\rho) \subseteq \Psi_{\bar{x}_t}(0) \subseteq \Theta_{\bar{x}_t}(-\rho). \quad (82)$$

Let $\{\Psi_{\bar{s}_t}(0)\}$ be a partition of the space of noisy signals $\{z_0^t\}$. By definition, $z_0^t \in \Psi_{\bar{s}_t}(0)$ if and only if $p(\bar{s}_t|z_0^t) \ge p(\bar{u}_t|z_0^t)$ for all $\bar{u}_t \neq \bar{s}_t$. Hence, from (76)–(77), we obtain

$$\begin{aligned}\overline{\zeta_t^2} &= \sum_{\bar{x}_t} \int p(\bar{x}_t, z_0^t) g(\bar{x}_t, \bar{x}_t^*, z_0^t)\, dz_0^t \\ &= \sum_{\bar{x}_t, \bar{s}_t} \int_{\Psi_{\bar{s}_t}(0)} p(\bar{x}_t, z_0^t) g(\bar{x}_t, \bar{x}_t^*, z_0^t)\, dz_0^t \\ &= \sum_{\bar{x}_t, \bar{s}_t} \int_{\Psi_{\bar{s}_t}(0)} p(\bar{x}_t, z_0^t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_0^t. \end{aligned} \quad (83)$$

The upper and lower bounds on $\overline{\zeta_t^2}$ are obtained by applying (82) to (83) as follows:

$$\begin{aligned}\overline{\zeta_t^2} &\le \sum_{\bar{x}_t, \bar{s}_t} \int_{\Theta_{\bar{s}_t}(-\rho)} p(\bar{x}_t, z_0^t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_0^t \\ &= \sum_{\bar{x}_t, \bar{s}_t} p(\bar{x}_t) \int_{\Omega_{\bar{s}_t}(-\rho)} b(z_t|\bar{x}_t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t. \end{aligned} \quad (84)$$

Similarly,

$$\begin{aligned}\overline{\zeta_t^2} &\ge \sum_{\bar{x}_t, \bar{s}_t} \int_{\Theta_{\bar{s}_t}(\rho)} p(\bar{x}_t, z_0^t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_0^t \\ &= \sum_{\bar{x}_t, \bar{s}_t} p(\bar{x}_t) \int_{\Omega_{\bar{s}_t}(\rho)} b(z_t|\bar{x}_t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t. \end{aligned} \quad (85)$$

The integrals in (84) and (85) are analogous to the integral $I_{\bar{s}_t}(\bar{x}_t)$ defined in (22), where the latter is taken over $\Omega_{\bar{s}_t}(-\rho)$ and $\Omega_{\bar{s}_t}(\rho)$, respectively. Hence, upper and lower bounds similar to those developed in Section II-B can be applied to (84) and (85), respectively. In this case, the lower bound on $I_{\bar{s}_t}(\bar{x}_t)$ is identical to that given in (43), and the upper bound is given by the product of (39) and $\exp(\lambda\rho)$.

In summary, if $\overline{\epsilon_t^2}$(mmse) denotes the mmse of the estimator (10), and $\overline{\epsilon_t^2}$(des) denotes the mse of the detection-estimation scheme, we have shown that

$$\begin{aligned}&\lim_{K\to\infty}\left[-\frac{1}{K}\ln\left(\overline{\epsilon_t^2}(\text{mmse}) - \overline{\xi_t^2}\right)\right] \\ &\qquad = \lim_{K\to\infty}\left[-\frac{1}{K}\ln\left(\overline{\epsilon_t^2}(\text{des}) - \overline{\xi_t^2}\right)\right] \end{aligned} \quad (86)$$

for CSM's with AWS Gaussian subsources.

V. MMSE Analysis for Continuous Time CSM's

In this section, we analyze the mmse in causal estimation of the output signal from a continuous time CSM which has been contaminated by statistically independent additive Gaussian white noise. A continuous time CSM is defined analogously to the discrete time CSM. It is a random process whose statistics at each time instant depend on the state of the driving Markov chain. The analysis is performed using Duncan's theorem [13] that relates the mmse in causal estimation to the average mutual information between the clean and noisy signals assuming additive Gaussian white noise. Similarly to the discrete case, we show that the mmse can be decomposed into the mmse of the completely informed estimator, and an additional error term for which upper and lower bounds are developed.

The continuous time CSM is defined as follows. Let $\{u_i, i = 0, 1, \cdots\}$ be a first-order M-state time-homogeneous Markov chain with initial state probabilities $\{\pi_\beta\}$ and state transition probabilities $\{a_{\alpha\beta}\}$, where $\alpha, \beta = 1, \cdots, M$. Assume that state transitions may occur every T seconds. In this case, $a_{\alpha\beta}$ is interpreted as a time function that denotes the transition probability from state α at some time instant, say τ, to state β at $\tau + T$. The Markov process associated with $a_{\alpha\beta}$ [26, p. 236], denoted here by $x_0^t \triangleq \{x_\tau, 0 \le \tau \le t\}$, is defined by

$$\begin{aligned}\Pr\{x_{\tau_1} = \nu_1, \cdots, x_{\tau_N} = \nu_N\} &= \Pr\{u_{\lfloor \tau_1/T \rfloor} = \nu_1, \cdots, u_{\lfloor \tau_N/T \rfloor} = \nu_N\} \\ &= \pi_{\nu_1} a_{\nu_1\nu_2} \cdots a_{\nu_{N-1}\nu_N} \end{aligned} \tag{87}$$

for every finite set $0 \le \tau_1 < \cdots < \tau_N \le t$, where $\nu_i \in \{1, \cdots, M\}$ for $i = 1, \cdots, N$, and $\lfloor \gamma \rfloor$ denotes the largest integer which does not exceed γ. Since the first state transition may occur only at time $\tau = T$, $a_{\alpha\beta}$ is a continuous function at $\tau = 0$, and the process x_τ is continuous in probability [26, p. 239]. Now, during each T second interval, a random process whose statistics depend on the state is generated. Let the output process be denoted by $y_0^t \triangleq \{y_\tau, 0 \le \tau \le t\}$ where now y_τ is a real scalar ($y_\tau \in R^1$). As with the discrete case, we assume that the T second output signals generated from a given sequence of states are statistically independent, and that $a_{\alpha\beta} \ge a_{\min} > 0$. Furthermore, we assume that the process y_0^t is continuous in probability, and y_0^t has finite energy, i.e.,

$$\int_0^t E\{y_\tau^2\}\, d\tau < \infty. \tag{88}$$

The noisy signal $z_0^t \triangleq \{z_\tau, 0 \le \tau \le t\}$ is obtained from

$$dz_\tau = y_\tau\, d\tau + dw_\tau, \tag{89}$$

where w_τ is a standard Brownian motion. We assume that $w_0^t \triangleq \{w_\tau, 0 \le \tau \le t\}$ is statistically independent of $y_0^t \triangleq \{y_\tau, 0 \le \tau \le t\}$ and of $x_0^t \triangleq \{x_\tau, 0 \le \tau \le t\}$. Since these processes are continuous in probability, there exists a version of each process defined on the same sample space, which is separable relative to the closed Borel sets, measurable, and which equals the original process with probability one [26, Theorem 2.6, p. 61]. Hence, in the subsequent discussion, where mutual information and conditional mean are used, the original processes can be substituted by their separable measurable versions. To simplify the notation, however, we shall not make explicit distinction between the processes and their measurable separable versions.

Let $x \triangleq x_0^t$, $y \triangleq y_0^t$, and $z \triangleq z_0^t$. Let $I(y; z)$ be the average mutual information between the two processes y and z. Let $I(y; z|x)$ be the conditional average mutual information between y and z given x. Let $I((x, y); z)$ be the average mutual information between (x, y) and z. Let P_{xyz} be the distribution of (x, y, z), and let $P_z \times P_{xy}$ be a product measure of the marginal distributions. Assume that $P_{xyz} \ll P_z \times P_{xy}$, that is, P_{xyz} is absolutely continuous with respect to $P_z \times P_{xy}$. From [12, corollary 5.5.3] this condition guarantees the existence of $I((x, y); z)$, and hence of $I(x; z)$, $I(y; z)$, $I(y; z|x)$, and $I(x; z|y)$, for random variables x, y, z with standard alphabets. The average mutual information $I(y; z)$ is defined by [12, (5.5.4)], [13]:

$$I(y; z) = \int \ln \frac{dP_{yz}}{d(P_y \times P_z)}\, dP_{yz}. \tag{90}$$

The conditional average mutual information $I(y; z|x)$ is defined by [12, (5.5.5)]:

$$I(y; z|x) = \int \ln \frac{dP_{xyz}}{dP_{y\times z|x}}\, dP_{xyz}, \tag{91}$$

where $P_{y\times z|x}$ is the distribution on x, y, z which agrees with P_{xyz} on the conditional distributions of y given x and z given x and with the marginal distribution of x, but which is such that $y \to x \to z$ forms a Markov chain [17, p. 171]. From Kolmogorov's formula [12, corollary 5.5.3], [30] we have that

$$\begin{aligned} I((x, y); z) &= I(y; z|x) + I(x; z) \\ &= I(x; z|y) + I(y; z). \end{aligned} \tag{92}$$

Furthermore, since $x \to y \to z$ forms a Markov chain under P_{xyz}, we have from [12, Lemma 5.5.2] that $I(x; z|y) = 0$. Hence, $I((x, y); z) = I(y; z)$, and

$$I(y_0^t; z_0^t) = I(y_0^t; z_0^t|x_0^t) + I(x_0^t; z_0^t). \tag{93}$$

From Duncan's theorem, we have

$$I(y_0^t; z_0^t) = \frac{1}{2}\int_0^t E\{(y_\tau - \hat{y}_\tau)^2\}\, d\tau, \tag{94}$$

where

$$\hat{y}_\tau \triangleq E\{y_\tau | z_0^\tau\} \tag{95}$$

is the causal mmse estimator of y_τ given z_0^τ. Hence,

$$\overline{\epsilon_t^2} \triangleq \frac{1}{t}\int_0^t E\{(y_\tau - \hat{y}_\tau)^2\}\, d\tau = \frac{2}{t} I(y_0^t; z_0^t) \tag{96}$$

is the mmse obtained in estimating y_t by $\hat{y}_t$. Similarly, since the conditions of Duncan's theorem are satisfied when the state sequence x_0^τ is given, we have that

$$I(y_0^t; z_0^t | x_0^t) = \frac{1}{2}\int_0^t E\left\{\left(y_\tau - \hat{y}_{\tau|x_0^\tau}\right)^2\right\} d\tau, \quad (97)$$

where

$$\hat{y}_{\tau|x_0^\tau} \triangleq E\{y_\tau | z_0^\tau, x_0^\tau\} \quad (98)$$

is the causal mmse informed estimator of y_τ given x_0^τ. Hence,

$$\overline{\xi_t^2} \triangleq \frac{1}{t}\int_0^t E\left\{\left(y_\tau - \hat{y}_{\tau|x_0^\tau}\right)^2\right\} d\tau = \frac{2}{t} I(y_0^t; z_0^t | x_0^t) \quad (99)$$

is the mmse obtained in estimating y_t using the informed estimator (98). Substituting (96) and (99) into (93) gives

$$\overline{\epsilon_t^2} = \overline{\xi_t^2} + \frac{2}{t} I(x_0^t; z_0^t). \quad (100)$$

This equation shows that similarly to the discrete case, the mmse equals the mmse of the informed estimator and an additional error term. The error term for the continuous time signals is given by the average mutual information between the state process and the noisy signal. Note that this result is not specific to our continuous time CSM's, and it can be applied to any signals x, y, z continuous in probability, which form a Markov chain $x \to y \to z$ and satisfy (88)–(89). In our model, however, (100) has an interesting interpretation since x_0^t is a state process. Note that for the trivial case of a deterministic switch $I(x_0^t; z_0^t) = 0$. Hence, $\overline{\epsilon_t^2} = \overline{\xi_t^2}$ as expected.

The relationship in (100) can be specialized for the particular continuous time CSM's considered in this section as follows. Let $\tau = mT + \tau'$ for some integer m. Using the assumption that signals generated from a given sequence of states are statistically independent and the assumption that the signal is degraded by white Gaussian noise, we have that

$$\begin{aligned} \hat{y}_{\tau|x_0^\tau} &= E\{y_\tau | z_0^\tau, x_0^\tau\} \\ &= E\{y_\tau | z_{mT}^\tau, x_\tau\} \\ &\triangleq \hat{y}_{\tau|x_\tau}. \end{aligned} \quad (101)$$

Furthermore, applying the rule of iterated expectation [17, p. 161] to (95) results in the desired estimator given by

$$\hat{y}_\tau = \sum_{x_\tau} \Pr\{x_\tau | z_0^\tau\} \hat{y}_{\tau|x_\tau}. \quad (102)$$

Following a derivation similar to (13) it can be shown that

$$\frac{1}{t}\int_0^t E\{(y_\tau - \hat{y}_\tau)^2\} d\tau = \frac{1}{t}\int_0^t E\left\{\left(y_\tau - \hat{y}_{\tau|x_\tau}\right)^2\right\} d\tau + \overline{\eta_t^2}, \quad (103)$$

where $\overline{\eta_t^2}$ is defined similarly to (16),

$$\overline{\eta_t^2} \triangleq \frac{1}{2t}\int_0^t \sum_{x_\tau, s_\tau} E\left\{\Pr(x_\tau | z_0^\tau)\Pr(s_\tau | z_0^\tau)\left(\hat{y}_{\tau|x_\tau} - \hat{y}_{\tau|s_\tau}\right)^2\right\} d\tau. \quad (104)$$

Hence, from (96), (99)–(100), and (103) we obtain

$$\overline{\eta_t^2} = \frac{2}{t} I(x_0^t; z_0^t). \quad (105)$$

The error term in (100), or equivalently $\overline{\eta_t^2}$ in (104), will be evaluated by developing upper and lower bounds on $I(x_0^t; z_0^t)/t$. We assume, without loss of generality, that $t = nT$ for some integer n, and study the asymptotic behavior of the bounds as $T \to \infty$. Since only causal estimation is considered, the significance of letting T go to infinity is that asymptotic estimation of y_t is performed from $z_{-\infty}^t$. Note that the situation here is analog to estimating the last sample in the K-dimensional vector y_t from z_0^t in the discrete case. In that case, however, the entire vector y_t was simultaneously estimated, and hence the first $K-1$ samples of each vector were estimated in a noncausal manner. The estimation problems for the discrete and the continuous time models were formulated differently, since normally vector estimation is performed in practice using discrete time models (see, e.g., [7]), and the analysis for the continuous time models uses Duncan's theorem which can only be applied to causal estimation.

The lower bound on $I(x_0^t; z_0^t)/t$ is developed by analyzing the system whose block diagram is shown in Fig. 3. In this system, u_n is a discrete process obtained from sampling x_t at T second intervals starting from $\tau = 0$. Similarly, z_n is a discrete time vector process obtained from sampling z_t at $\Delta \triangleq T/K$ second intervals, where K is a given integer. Hence, z_n is a K-dimensional vector ($z_n \in R^K$). Finally, $\hat{u}_n$ in Fig. 3 denotes an estimate of the state u_n as obtained from the sampled noisy signal. From the data processing theorem [12, p. 129] we have that

$$I(x_0^t; z_0^t) \geq I(u_0^n; z_0^n), \quad (106)$$

where $u_0^n \triangleq \{u_0, \cdots, u_{n-1}\}$, and $z_0^n \triangleq \{z_0, \cdots, z_{n-1}\}$. Hence, a lower bound on $I(x_0^t; z_0^t)$ can be obtained from a lower bound on $I(u_0^n; z_0^n)$. Since [12, lemma 5.5.6]

$$I(u_0^n; z_0^n) = H(u_0^n) - H(u_0^n | z_0^n), \quad (107)$$

and $H(u_0^n)$ is finite, a lower bound on $I(u_0^n; z_0^n)$ can be obtained from an upper bound on $H(u_0^n | z_0^n)$. This bound is provided by Fano's inequality [12, corollary 4.2.1]:

$$H(u_0^n | z_0^n)/n \leq P_e^{(n)}(T)\log(M-1) + h_2(P_e^{(n)}(T)) \triangleq \epsilon_n(T), \quad (108)$$

where

$$P_e^{(n)}(T) \triangleq \frac{1}{n}\sum_{i=0}^{n-1} \Pr\{\hat{u}_i \neq u_i\},$$

and $h_2(\cdot)$ denotes the binary entropy function. Combining (106)–(108) we obtain the desired lower bound for

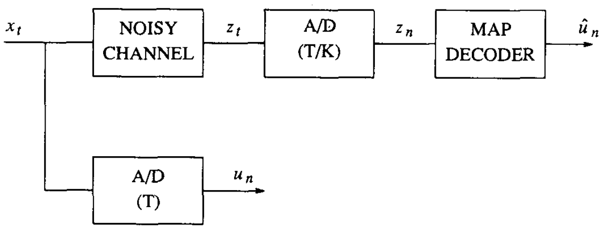

Fig. 3. Lower bound estimation scheme.

$I(x_0^t; z_0^t)$:

$$I(x_0^t; z_0^t) \geq H(u_0^n) - n\epsilon_n(T). \tag{109}$$

The upper bound on $I(x_0^t; z_0^t)$ results from

$$I(x_0^t; z_0^t) = H(x_0^t) - H(x_0^t|z_0^t) \leq H(x_0^t) = H(u_0^n), \tag{110}$$

since $H(x_0^t|z_0^t) \geq 0$ and x_t can be obtained from u_n.

For the stationary first-order Markov chain considered here, we have, from the chain rule for entropy [12, corollary 2.5.1] and from [12, lemma 2.5.2], that

$$H(u_0^n) = H(u_0) + (n-1)H(u_1|u_0) \leq H(u_0) + (n-1)H(u_1) = nH(u_1). \tag{111a}$$

Similarly,

$$H(u_0^n) = H(u_1) + (n-1)H(u_1|u_0) \geq H(u_1|u_0) + (n-1)H(u_1|u_0) = nH(u_1|u_0). \tag{111b}$$

Hence, using $t = nT$ we obtain, from (109)–(111),

$$H(u_1|u_0)/T - \epsilon_n(T)/T \leq I(x_0^t; z_0^t)/t \leq H(u_1)/T. \tag{112}$$

These bounds approach zero harmonically as $T \to \infty$ and n is fixed, provided that $\epsilon_n(T) \to 0$ as $T \to \infty$. This is now shown for CSM's with Gaussian subsources. Specifically, we develop an upper bound on $\epsilon_n(T)$ and show that it converges to zero exponentially.

The upper bound on $\epsilon_n(T)$ can be obtained from an upper bound on $P_e^{(n)}(T)$. Consider the single letter probability of misclassification error $\Pr\{\hat{u}_k \neq u_k\}$, $k = 1, \cdots, n$, when u_k is estimated from z_k using the maximum *a posteriori* (MAP) decoder. From [27] we have that

$$\Pr\{\hat{u}_k \neq u_k\} \leq \frac{1}{2} \sum_{i<j} e^{\mu_{ij}(0.5)}, \tag{113}$$

where

$$\mu_{ij}(\lambda) \triangleq \ln \int_{R^K} b^{\lambda}(z_k|i) b^{1-\lambda}(z_k|j)\, dz_k,$$

for $i, j = 1, \cdots, M$. Hence, for CSM's with AWS Gaussian subsources, we have from Section III that this bound approaches zero exponentially as $K \to \infty$ for a fixed Δ, or, the bound approaches zero exponentially as $T \to \infty$. From the definitions of $P_e^{(n)}(T)$ and $h_2(P_e^{(n)}(T))$ it follows that the upper bound on $\epsilon_n(T)$ approaches zero exponentially as $T \to \infty$.

We have seen that for CSM's with AWS Gaussian subsources, the bounds on the error term $\overline{\eta_t^2} = 2I(x_0^t; z_0^t)/t$ in (100), converge to zero harmonically as $T \to \infty$. Hence, similarly to discrete time case, the asymptotic mmse in the continuous time case is the mmse of the completely informed estimator. The convergence rate of the bounds for the discrete and continuous time models, however, appear different. In Section III exponential convergence of $\overline{\eta_t^2}$ was proven for the discrete case, while here the convergence rate was shown to be harmonic. This difference can be explained as follows.

Consider first causal estimation of the discrete time signal under conditions similar to those used for estimation of the continuous time signal. Specifically, consider the mmse estimation of the last sample of the vector y_t, $y_{t,K}$, from the vectors of noisy signal z_0^t. This estimator is obtained from the minimization of

$$\overline{\epsilon_{t,K}^2} = E\{(y_{t,K} - \tilde{y}_{t,K})^2\} \tag{114}$$

over $\tilde{y}_{t,K}$, and is given by $\hat{y}_{t,K} = E\{y_{t,K}|z_0^t\}$. In this case, it is easy to show, using an analysis similar to that given in Sections II–III, that the mmse (114) can be decomposed into the mmse of the informed estimator and an additional error term; the error term is given by

$$\overline{\eta_{t,K}^2} = \frac{1}{2} \sum_{\bar{x}_t, \bar{s}_t} p(\bar{x}_t|z_0^t) p(\bar{s}_t|z_0^t) E\left\{\left(\hat{y}_{t,K|\bar{x}_t} - \hat{y}_{t,K|\bar{s}_t}\right)^2\right\}; \tag{115}$$

the bounds on this term depend only on z_t but not on z_0^{t-1}; and these bounds approach zero exponentially. Assume that the exponential bounds are proportional to $\exp(-BK)$ (see (71)). Note that since causal estimation of the last sample of y_t is not different from causal estimation of any other sample of y_t, then estimation of say the lth sample of y_t, results in exponential bounds which are proportional to $\exp(-Bl)$. If the lth sample of the vector y_t is estimated in a noncausal manner from z_0^t, however, then it can be shown that the bounds on the error term in this case are proportional to $\exp(-BK)$.

Consider now the time average mmse (13), or equivalently,

$$\frac{1}{K} \sum_{l=1}^{K} \overline{\epsilon_{t,l}^2}. \tag{116}$$

This time average mmse for discrete signals is analogous to the time average mmse (96) used for the continuous time signals. For causal estimation of $y_{t,l}$, the time average bounds are proportional to

$$\frac{1}{K} \sum_{l=1}^{K} e^{-Bl} = \frac{1}{K} e^{-B} \frac{e^{-BK} - 1}{e^{-B} - 1}, \tag{117}$$

which has a harmonic convergence rate. In the case of noncausal estimation of $y_{t,l}$, we similarly obtain that the time average bounds are proportional to

$$\frac{1}{K}\sum_{l=1}^{K} e^{-BK} = e^{-BK}, \tag{118}$$

which has an exponential convergence rate.

The foregoing discussion shows that the bounds on the time average error terms of the mmse in *causal* estimation of discrete as well as continuous time signals, (117) and (105), respectively, have similar harmonic convergence rate. The convergence rate of the bounds on the time average error term of the mmse in noncausal estimation of discrete time signals was shown to be exponential. For the discrete case, we were also able to calculate the bounds on the individual error terms obtained in causal as well as noncausal mmse estimation of each sample of the vector at time t, and we showed that in both cases the convergence rate of these bounds is exponential. For the continuous case, we do not have parallel results on the convergence of the individual error terms due to nature of the analysis performed here.

VI. An Example: MMSE Parameter Estimation

The bounds developed in Section II can be useful in mmse parameter estimation problems as is demonstrated in this section. Let θ be a random vector of N parameters of some random process. Let $p(y|\theta)$ be the pdf of a $K-$ dimensional vector y of that process given θ. Let $p(\theta)$ be the *a priori* pdf of θ. Let $\{\omega_j, j=1,\cdots,M\}$ be a partition of the parameter space of θ, and let $\{\theta_j, j=1,\cdots,M\}$ be a grid in that parameter space such that $\theta_j \in \omega_j$, $j=1,\cdots,M$. The mmse estimator of θ from y is given by

$$\hat{\theta} = \int \theta p(\theta|y)\, d\theta = \sum_{j=1}^{M} p(j|y) E\{\theta|j,y\}, \tag{119}$$

where $p(j|y)$ denotes the posterior probability of $\theta \in \omega_j$ given y, and $E\{\theta|j,y\}$ is the conditional mean of θ given that $\theta \in \omega_j$ and y. The mmse associated with this estimator can be evaluated using a similar analysis to that presented in Section II. The mmse will be composed of the mmse of the informed estimator, and a cross error term which can be bounded from above and below.

An interesting particular case, considered in [23]–[25], results when θ can only take a finite number of values, i.e.,

$$p(\theta) = \sum_{j=1}^{M} p(\theta_j)\delta(\theta-\theta_j), \tag{120}$$

where $\delta(\cdot)$ denotes the Kronecker delta function. In this case,

$$E\{\theta|j,y\} = \theta_j, \tag{121}$$

and the problem becomes a detection rather than an estimation problem. Hence, we expect the mmse to approach zero as $K \to \infty$. The mmse estimator of θ is given by

$$\hat{\theta} = \sum_{j=1}^{M} \theta_j p(\theta_j|y), \tag{122}$$

where $p(\theta_j|y)$ is the *a posteriori* probability of $\theta = \theta_j$ given y. The mmse associated with this estimator can be calculated similarly to (13). We have that

$$\overline{\epsilon^2} \triangleq \frac{1}{N}E\{\|\theta-\hat{\theta}\|^2\} = \frac{1}{2}\sum_{i,j=1}^{M}\frac{1}{N}(\theta_i-\theta_j)^{\#}(\theta_i-\theta_j) \cdot E\{p(\theta_i|y)p(\theta_j|y)\}. \tag{123}$$

Hence, $\overline{\epsilon^2}$ can be upper and lower bounded by applying the bounds of Section II to $E\{p(\theta_i|y)p(\theta_j|y)\}$ using $a_{\min} = \min_j\{p(\theta_j)\} > 0$, $\bar{x}_t = i$, $\bar{s}_t = j$, $z_0^t = y$, $b(y_t|\bar{x}_t) = p(y|i)$, and $g(\bar{x}_t, \bar{s}_t, z_t) \equiv 1$. If, for example, $p(y|j)$ is Gaussian, then from the results of Section III we know that the upper and lower bounds on $\overline{\epsilon^2}$ approach zero exponentially.

The major difference between our approach and the approach used in [23]–[25] is that here $E\{p(\theta_i|y)p(\theta_j|y)\}$ is bounded while in [23]–[25] only $E\{p(\theta_i|y)\}$ was bounded using the fact that $p(\theta_j|y) \le 1$. Furthermore, the problem of finding a lower bound on $\overline{\epsilon^2}$ was not considered in [23]–[25].

VII. Comments

We studied the performance of the mmse estimator of the output signal from a CSM given a noisy version of that signal. The analysis was performed for discrete as well as continuous time CSM's. In both cases the noise was assumed additive and statistically independent of the signal. In the discrete case, the noise was assumed to be another CSM, while in the continuous case only Gaussian white noise was considered.

In the discrete case, estimation of vectors of the clean signal from past and present vectors of the noisy signal was studied. This problem was motivated by the way CSM based mmse estimation is used in practice. In this case, vectors of the signal were estimated in a causal manners, but the samples within each vector (except for the last one) were estimated in a noncausal manner. The criterion used for this vector estimation problem was naturally chosen to be the time average mmse over all samples of the vector. Causal and noncausal mmse estimation of the individual samples of the clean signal was also considered and compared with the vector estimation. In the continuous case, the analysis was more restricted, as only causal estimation using the time average mmse over the time duration of the signal was considered. The restriction on the noise statistics and the analysis conditions in the continuous case resulted from using Duncan's theorem which can only be applied under these conditions.

For both discrete and continuous time CSM's, it was shown that the mmse is composed of the mmse of the completely informed estimator, and an additional error component for which upper and lower bounds were developed. The convergence rate of these bounds depends on the causality of the estimators as well as on whether the mmse or the time average mmse is considered. For discrete time CSM's with AWS Gaussian subsources, it was shown that the bounds corresponding to the mmse of each sample converge exponentially to zero in causal as well as noncausal estimation. The bounds that correspond to the time average mmse converge to zero harmonically in causal estimation, and exponentially in noncausal estimation. For the continuous time case, it was shown that the bounds which correspond to the time average mmse in causal estimation converges to zero harmonically.

ACKNOWLEDGMENT

The authors are grateful to Prof. S. Shamai (Shitz), Prof. A. Dembo, and Prof. R. M. Gray, for helpful discussions during this work. They also acknowledge the useful comments made by the anonymous referees that improved the presentation of this paper.

APPENDIX

Lemma: The semi-invariant moment generating function $\mu(\lambda)$ defined in (36) is finite for $0 \le \lambda \le 1$ provided that Assumption 2) holds.

Proof: By substituting (20) and (31)–(32) into (37), we obtain

$$\begin{aligned}\mu(\lambda) &= -\ln \Phi(\bar{x}_t, \bar{s}_t) + \ln \int_{R^K} b^{\lambda}(z_t|\bar{x}_t) b^{1-\lambda}(z_t|\bar{s}_t) \\ &\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t \\ &\le -\ln \Phi(\bar{x}_t, \bar{s}_t) + \ln \int_{R^K} \max\{b(z_t|\bar{x}_t), b(z_t|\bar{s}_t)\} \\ &\quad \cdot g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t \\ &\le -\ln \Phi(\bar{x}_t, \bar{s}_t) + \ln \left\{ \int_{R^K} b(z_t|\bar{x}_t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t \right. \\ &\quad \left. + \int_{R^K} b(z_t|\bar{s}_t) g(\bar{x}_t, \bar{s}_t, z_t)\, dz_t \right\} \\ &= -\ln \Phi(\bar{x}_t, \bar{s}_t) + \ln (\Phi(\bar{x}_t, \bar{s}_t) + \Phi(\bar{s}_t, \bar{x}_t)) \\ &< \infty, \end{aligned} \tag{A.1}$$

since $\Phi(\bar{x}_t, \bar{s}_t) < \infty$ for all $\{\bar{x}_t, \bar{s}_t\}$ as follows from Assumption 2).

Derivation of $\dot{\mu}(\lambda)$ and $\ddot{\mu}(\lambda)$: Let

$$\begin{aligned} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) &= b^{\lambda}(z_t|\bar{x}_t) b^{1-\lambda}(z_t|\bar{s}_t) \\ &= C_{\lambda}(\bar{x}_t, \bar{s}_t) N(0, R_{\lambda}(\bar{x}_t, \bar{s}_t)), \end{aligned} \tag{A.2}$$

where λ is such that $R_{\lambda}(\bar{x}_t, \bar{s}_t)$ is positive definite, $N(0, R_{\lambda}(\bar{x}_t, \bar{s}_t))$ denotes a zero-mean Gaussian pdf with covariance $R_{\lambda}(\bar{x}_t, \bar{s}_t)$ given in (58), and $C_{\lambda}(\bar{x}_t, \bar{s}_t)$ is independent of z_t and can be obtained from (57). From (56), we have that

$$\begin{aligned}\mu(\lambda) &= \ln \int_{R^K} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) z_t^{\#} H^2_{\bar{x}_t \bar{s}_t} z_t\, dz_t \\ &\quad - \ln \{K \Phi(\bar{x}_t, \bar{s}_t)\}. \end{aligned} \tag{A.3}$$

Using

$$\frac{\partial}{\partial \lambda} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) = f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) \ln \frac{b(z_t|\bar{x}_t)}{b(z_t|\bar{s}_t)}, \tag{A.4}$$

we obtain from (A.3) the following expression for the first derivative of $\mu(\lambda)$ with respect to λ:

$$\dot{\mu}(\lambda) = \frac{\int_{R^K} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) z_t^{\#} H^2_{\bar{x}_t \bar{s}_t} z_t \ln \frac{b(z_t|\bar{x}_t)}{b(z_t|\bar{s}_t)}\, dz_t}{\int_{R^K} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) z_t^{\#} H^2_{\bar{x}_t \bar{s}_t} z_t\, dz_t}. \tag{A.5}$$

The second derivative of $\mu(\lambda)$ with respect to λ is obtained from (A.5) by using (A.4), the normality of $b(z_t|\bar{x}_t)$ and of $b(z_t|\bar{s}_t)$, the fact that $f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t)$ is proportional to a Gaussian pdf, and (64). This results in

$$\begin{aligned}\ddot{\mu}(\lambda) &= \left[\frac{1}{2} \ln |Q_{\bar{x}_t} Q^{-1}_{\bar{s}_t}|\right]^2 \\ &\quad + \frac{1}{2} \ln |Q_{\bar{x}_t} Q^{-1}_{\bar{s}_t}| \operatorname{tr} \left\{ \left(Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}\right) R_{\lambda}(\bar{x}_t, \bar{s}_t) \right\} \\ &\quad + \ln |Q_{\bar{x}_t} Q^{-1}_{\bar{s}_t}| \frac{\operatorname{tr}\left\{ H^2_{\bar{x}_t \bar{s}_t} R_{\lambda}(\bar{x}_t, \bar{s}_t) \left(Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}\right) R_{\lambda}(\bar{x}_t, \bar{s}_t) \right\}}{\operatorname{tr}\left\{ H^2_{\bar{x}_t \bar{s}_t} R_{\lambda}(\bar{x}_t, \bar{s}_t) \right\}} \\ &\quad + \frac{1}{4} \frac{\int_{R^K} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) \left(z_t^{\#} H^2_{\bar{x}_t \bar{s}_t} z_t\right) \left(z_t^{\#} \left(Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}\right) z_t\right)^2 dz_t}{\int_{R^K} f_{\lambda}(z_t|\bar{x}_t, \bar{s}_t) z_t^{\#} H^2_{\bar{x}_t \bar{s}_t} z_t\, dz_t} \\ &\quad - (\dot{\mu}(\lambda))^2. \end{aligned} \tag{A.6}$$

Now, the integral in (A.6) can be calculated using (67) with $A = H^2_{\bar{x}_t \bar{s}_t}$, $B = Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}$, and $R = R_{\lambda}(\bar{x}_t, \bar{s}_t)$. This results in

$$\begin{aligned}\ddot{\mu}(\lambda) &= \left[\frac{1}{2} \ln |Q_{\bar{x}_t} Q^{-1}_{\bar{s}_t}|\right]^2 \\ &\quad + \frac{1}{2} \ln |Q_{\bar{x}_t} Q^{-1}_{\bar{s}_t}| \operatorname{tr} \left\{ \left(Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}\right) R_{\lambda}(\bar{x}_t, \bar{s}_t) \right\} \\ &\quad + \left[\frac{1}{2} \operatorname{tr} \left\{ \left(Q^{-1}_{\bar{x}_t} - Q^{-1}_{\bar{s}_t}\right) R_{\lambda}(\bar{x}_t, \bar{s}_t) \right\} \right]^2 \\ &\quad - (\dot{\mu}(\lambda))^2 + \epsilon(K), \end{aligned} \tag{A.7}$$

where $\epsilon(K)/K^2 \to 0$ as $K \to \infty$. Applying the Toeplitz distribution theorem to $\ddot{\mu}(\lambda)/K^2$ and using (66) we arrive at (68).

REFERENCES

[1] T. Berger, *Rate Distortion Theory.* Engelwood Cliffs, NJ: Prentice-Hall Inc., 1971.
[2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE,* vol. 79, pp. 257–286, Feb. 1989.
[3] H. L. van-Trees, *Detection, Estimation and Modulation Theory, Part I.* New York: Wiley, 1968.
[4] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory,* vol. IT-16, pp. 406–411, July 1970.

[5] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory,* vol. 34, pp. 826–834, July 1988.

[6] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory,* vol. IT-15, pp. 350–361, May 1969.

[7] Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," *IEEE Trans. Signal Processing,* vol. 40, pp. 725–735, Apr. 1992.

[8] D. T. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. Automat. Contr.,* vol. AC-10, pp. 434–439, Oct. 1965. cf. Author's reply, *IEEE Trans. Automat. Contr.,* vol. AC-14, pp. 216–218, Apr. 1969.

[9] F. L. Sims and D. G. Lainiotis, "Recursive algorithm for calculation of the adaptive Kalman filter weighing coefficients," *IEEE Trans. Automat. Contr.,* vol. AC-14, pp. 215–217, Apr. 1969.

[10] D. G. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaptation," *IEEE Trans. Automat. Contr.,* vol. AC-16, pp. 160–170, Apr. 1971.

[11] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.,* vol. 26, no. 2, pp. 195–239, Apr. 1984.

[12] R. M. Gray, *Entropy and Information Theory.* New York: Springer-Verlag, 1990.

[13] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.,* vol. 19, no. 1, pp. 215–220, July 1970.

[14] R. M. Gray, "Toeplitz and circulant matrices: II," Stanford Electron. Lab., Tech. Rep. 6504-1, Apr. 1977.

[15] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications.* New York: Chelsea, 1984.

[16] Y. Ephraim, D. Malah, and B.-H. Juang, "On the application of hidden Markov models for enhancing noisy speech," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 1846–1856, Dec. 1989.

[17] R. M. Gray, *Probability, Random Processes, and Ergodic Properties.* New York: Springer-Verlag, 1988.

[18] A. D. Wyner, "On the asymptotic distribution of a certain functional of the Wiener process," *Ann. Math. Statist.,* vol. 40, no. 4, pp. 1409–1418, 1969.

[19] P. Lancaster and M. Tismenetsky, *The Theory of Matrices,* 2nd ed. New York: Academic Press, 1985.

[20] R. M. Gray, A. Buzo, A. H. Gray, Jr., and Y. Matsuyama, "distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-28, pp. 367–376, Aug. 1980.

[21] R. M. Gray, A. H. Gray, Jr., G. Rebolledo, and J. E. Shore, "Rate-distortion speech coding with a minimum discrimination information distortion measure," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 708–721, Nov. 1981.

[22] A. D. Whalen, *Detection of Signals in Noise.* New York: Academic Press, 1971.

[23] L. A. Liporace, "Variance of Bayes estimators," *IEEE Trans. Inform. Theory,* vol. IT-17, pp. 665–669, Nov. 1971.

[24] D. Kazakos, "New convergence bounds for Bayes estimators," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 97–104, Jan. 1981.

[25] L. Merakos and D. Kazakos, "Comments and corrections to New convergence bounds for Bayes estimators," *IEEE Trans. Inform. Theory,* vol. IT-29, pp. 318–320, Mar. 1983.

[26] J. L. Doob, *Stochastic Processes,* Wiley Classics Library Edition. New York: Wiley, 1990.

[27] D. G. Lainiotis, "A class of upper bounds on probability of error for multihypotheses pattern recognition," *IEEE Trans. Inform. Theory,* vol. IT-15, pp. 730–731, Nov. 1969.

[28] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco, CA: Holden-Day, 1964.

[29] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[30] T. T. Kadota, M. Zakai, and J. Ziv, "Capacity of a continuous memoryless channel with feedback," *IEEE Trans. Inform. Theory,* vol. IT-17, pp. 372–378, July 1971.