PERFORMANCE ANALYSIS OF THE BAYESIAN BEAMFORMER

Chunwei Jethro Lam, Andrew C. Singer

University of Illinois at Urbana Champaign
Coordinated Science Laboratory
1308 West Main St., Urbana, IL 61801
217-244-9263
clam@uiuc.edu, acsinger@uiuc.edu

ABSTRACT

We present an analysis of the performance of Bayesian beamformers that are able to estimate signals from unknown source directions by balancing multiple optimal estimates according to the a posteriori probability mass function (PMF). We show that the conditional mean square error (MSE) of the Bayesian beamformer asymptotically achieves the conditional MSE of an estimator that has prior knowledge of the true direction of arrival. The convergence rate depends on both the signal-to-noise ratio (SNR) and the Kullback Leibler distance between certain probability distributions on which the Bayesian model is defined.

1. INTRODUCTION

We consider a set of D narrowband signals arriving at an array of N sensors. At time k, the observed data vector is given by

$$\mathbf{x}_k = \sum_{m=0}^{D-1} \mathbf{a}(u_m) s_{m,k} + \mathbf{w}_k \tag{1}$$

where $s_{m,k}, m = 0, ..., D-1$ are the D equivalent discrete-time baseband zero-mean source signals and $\mathbf{w_k}$ is a vector of noise samples and where $s_{m,k}$ and $\mathbf{w}_k$ are modeled as independent identically distributed (i.i.d.) Gaussian random processes that are mutually independent; $\mathbf{a}(u_m)$ is the Nx1 array manifold vector defined on the normalized spatial frequency (or direction) u_m.

We assume that one of the signals is the desired signal and denote it by s and its direction by u_0. The observed data can be expressed as

$$\mathbf{x}_k = \mathbf{a}(u_0) s_k + \mathbf{n}_k. \tag{2}$$

This work was supported in part by the National Science Foundation under grant number CCR-0092598 (career) and by NASA under grant number NAG2-2840.

where $\mathbf{n}_k$ is the interference-plus-noise component with covariance R_n. The data covariance matrix has the form

$$R_x = E[\mathbf{x}_k \mathbf{x}_k^H] \tag{3}$$
$$= \sigma_s^2 \mathbf{a}(u_0)\mathbf{a}(u_0)^H + R_n \tag{4}$$

where σ_s^2 is the power of the desired signal.

We consider the estimation of the desired signal from all the available data $\mathbf{X}_k = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ using a linear array processor. Given the direction u_0, the minimum mean square error (MMSE) optimal estimate is [1]

$$\hat{s}_{k,\text{MMSE}} = E[s_k | \mathbf{X}_k, u_0] \tag{5}$$
$$= \sigma_s^2 \mathbf{a}(u_0)^H R_x^{-1} \mathbf{x}_k. \tag{6}$$

Next, we introduce a new estimate $\hat{s}_k(u)$ that has the form

$$\hat{s}_k(u) = \sigma_s^2 \mathbf{a}(u)^H R_x^{-1} \mathbf{x}_k. \tag{7}$$

where the variable u is referred as the look direction. When $u = u_0$, the estimate is the MMSE estimate. We consider the conditional MSE of this estimator in general:

$$\zeta(u) = E[|\hat{s}_k(u) - s_k|^2 | u_0] \tag{8}$$
$$= \sigma_s^2 (1 - \sigma_s^2 \mathbf{a}_0^H R_x^{-1} \mathbf{a}_0) + \sigma_s^4 \|\mathbf{a}(u) - \mathbf{a}_0\|_x^2 \tag{9}$$

where,

$$\mathbf{a}_0 \triangleq \mathbf{a}(u_0) \tag{10}$$
$$\|\mathbf{a}\|_x^2 \triangleq \mathbf{a}^H R_x^{-1} \mathbf{a}. \tag{11}$$

When the look direction matches with the true direction, the MSE achieves its minimum, that is

$$\zeta_{\text{opt}} = \zeta(u_0) = \sigma_s^2 (1 - \sigma_s^2 \mathbf{a}_0^H R_x^{-1} \mathbf{a}_0). \tag{12}$$

Rewriting (9), the MSE of the estimator with look direction u is a sum of the optimal term and a redundancy term:

$$\zeta(u) = \zeta_{\text{opt}} + \sigma_s^4 \|\mathbf{a}(u) - \mathbf{a}_0\|_x^2. \tag{13}$$

2. THE BAYESIAN APPROACH

Using a Bayesian approach [2], the unknown source direction is assumed to be a discrete random variable u that takes on values in the set $U = \{u_1, \ldots, u_L\}$. The data samples $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are conditionally independent given u. However, without prior knowledge of the value of u, the data samples are statistically dependent. Each sample $\mathbf{x}_k$ can be described by the conditional probability density function (PDF) $p(\mathbf{x}|u)$ and the a priori PMF $p(u)$, and $p(\mathbf{x}_k) = \sum_v p(v)p(\mathbf{x}_k|v)$.

Using the Bayesian model, the MMSE estimate of s_k can be written

$$\begin{aligned}
\hat{s}_{B,k} &= E[s_k|\mathbf{X}_k] && (14)\\
&= E_u[E[s_k|\mathbf{X}_k, u]] && (15)\\
&= \sum_{u\in U} p(u|\mathbf{X}_k)E[s_k|\mathbf{X}_k, u] && (16)\\
&= \sum_{u\in U} p(u|\mathbf{X}_k)\hat{s}_k(u) && (17)
\end{aligned}$$

where

$$p(u|\mathbf{X}_k) = \frac{p(u)p(\mathbf{X}_k|u)}{\sum_v p(v)p(\mathbf{X}_k|v)}. \qquad (18)$$

The estimate can be viewed as a weighted average over multiple optimal estimates, each with different look directions. The weighting coefficients are governed by the a posteriori PMF of u given the available data. When k becomes large, the a posteriori PMF adapts to the environment and the Bayesian estimator converges to a single MMSE estimator with the optimal look direction in a mean square sense.

To implement this estimator, we need both $p(\mathbf{x}_k|u)$ and $p(u)$. The former can be derived from prior knowledge of the statistics of the noise and interference. The a priori PMF $p(u)$ may be derived from prior statistical knowledge of the true direction for the problem at hand.

3. PERFORMANCE ANALYSIS

We evaluate the performance of the Bayesian estimate $\hat{s}_{B,k}$ by calculating the corresponding conditional MSE at time k:

$$\zeta_{B,k} = E[|\hat{s}_{B,k} - s_k|^2|u_0]. \qquad (19)$$

Note that we are interested in the conditional MSE, as we would like to extend these results to cases when our Bayesian model does not hold, i.e., when $u_0 \notin U$. When the Bayesian model does hold, i.e, when $u_0 \in U$, the unconditional MSE is related to the conditional MSE as

$$E[|\hat{s}_{B,k} - s_k|^2] = \sum_u p(u)E[|\hat{s}_{B,k} - s_k|^2|u]. \qquad (20)$$

For cases when u_0 is not included in the parameter space U, we will show that the algorithm converges to a particular direction u^*, where u^* is the nearest point to u_0 with respect to certain distance metric that will be described later.

We begin by introducing an arbitrary estimate, $\tilde{s}_{B,k}$, which is defined to be

$$\tilde{s}_{B,k} = \sum_{u\in U} p(u|\mathbf{X}_{k-1})\hat{s}_k(u). \qquad (21)$$

This estimate is different from the Bayesian estimate in (17) by the composition of the weighting coefficients. The former uses all of the k available data samples to form the weighting coefficients, while the latter uses only the first $k-1$ data samples.

The MSE of $\tilde{s}_{B,k}$ is upper bounded as

$$\begin{aligned}
\tilde{\zeta}_{B,k} &= E[|\tilde{s}_{B,k} - s_k|^2|u_0] && (22)\\
&= E[|\sum_u p(u|\mathbf{X}_{k-1})\hat{s}_k(u) - s_k|^2|u_0] && (23)\\
&\leq \sum_u E[p(u|\mathbf{X}_{k-1})|\hat{s}_k(u) - s_k|^2|u_0]. && (24)
\end{aligned}$$

Since the expectation is taken over $\mathbf{X}_k$ conditioned on u_0 and the data samples $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are conditionally independent, the expected value of the weighting coefficients and the mean square error term can be separated into a product of two expectations, that is

$$\begin{aligned}
&E[p(u|\mathbf{X}_{k-1})|\hat{s}_k(u) - s_k|^2|u_0] && (25)\\
&= E[p(u|\mathbf{X}_{k-1})|u_0] \cdot E[|\hat{s}_k(u) - s_k|^2|u_0]. && (26)
\end{aligned}$$

The first term involves $\mathbf{X}_{k-1}$, and the second expectation involves only $\mathbf{x}_k$.

To aid our discussion about the closeness of one direction to another, we consider the Kullback Leibler distance between $p(\mathbf{x}_k|u_0)$ and $p(\mathbf{x}_k|u)$, that is

$$\begin{aligned}
D(p(\mathbf{x}_k|u_0)||p(\mathbf{x}_k|u)) &= E\left[\ln\frac{p(\mathbf{x}_k|u_0)}{p(\mathbf{x}_k|u)}\right] && (27)\\
&= \int p(\mathbf{x}_k|u_0)\ln\frac{p(\mathbf{x}_k|u_0)}{p(\mathbf{x}_k|u)}d\mathbf{x}_k && (28)\\
&\triangleq D(u_0||u). && (29)
\end{aligned}$$

The value is 0 if and only if $u = u_0$. For cases when $u_0 \notin U$, we consider $u^* \in U$, which minimizes $D(u_0||u)$ over all $u \in U$, that is,

$$u^* = \arg\min_{u\in U} D(u_0||u). \qquad (30)$$

Using a technique similar to that in [3], the following theorem is established to describe the mean convergence behavior of the weighting coefficients $E[p(u|\mathbf{X}_{k-1})|u_0]$ for every $u \in U$.

Theorem 1: *If $D(u_0||u)$ exists $\forall u \in U$ and there is an interval $\pi_T = (0, T] \subset (0, 1]$ such that $E[(p(\mathbf{x}|u)/p(\mathbf{x}|u^*))^t]$ exists $\forall t \in \pi_T, u \in U$, then the conditional expectation of the weighting coefficients are bounded as*

$$\begin{cases} 0 \leq E[p(u|\mathbf{X}_k)|u_0] < c(u)\ \eta^k(u) & u \neq u^* \\ (1 + \sum_{v \neq u^*} c(v)\eta^k(v))^{-1} < E[p(u|\mathbf{X}_k)|u_0] \leq 1 & u = u^* \end{cases} \tag{31}$$

where $c(u) > 0$ is a constant and $0 \leq \eta(u) < 1$ for all $u \neq u^*$. It follows that

$$\lim_{k \to \infty} E[p(u|\mathbf{X}_k)|u_0] = \begin{cases} 0 & u \neq u^* \\ 1 & u = u^*. \end{cases} \tag{32}$$

The proof is given in the next section.

The second expectation term in equation (26) is the same as the general conditional MSE $\zeta(u)$ in equation (9). Thus, the inequality in (22) becomes

$$\tilde{\zeta}_{B,k} \leq \sum_u E[p(u|\mathbf{X}_{k-1})|u_o]\zeta(u) \tag{33}$$

$$\leq \zeta_{\text{opt}} + \sigma_s^4 \sum_u E[p(u|\mathbf{X}_{k-1})|u_0] \cdot \|\mathbf{a}(u) - \mathbf{a}(u_0)\|_x^2. \tag{34}$$

From Theorem 1, the coefficient $E[p(u|\mathbf{X}_{k-1})|u_o]$ is upper bounded by 1 or $c(u)\ \eta^{k-1}(u)$ depending on different values of u. Using $\zeta(u^*) = \zeta_{\text{opt}} + \sigma_s^4\|\mathbf{a}(u^*) - \mathbf{a}_0\|_x^2$ from (9), we have

$$\tilde{\zeta}_{B,k} \leq \zeta(u^*) + \sigma_s^4 \sum_{u \neq u^*} c(u)\|\mathbf{a}(u) - \mathbf{a}_0\|_x^2\ \eta^{k-1}(u). \tag{35}$$

The analysis is completed by noting that $\zeta_{B,k} \leq \tilde{\zeta}_{B,k}$ by the definition of the MMSE estimate of s_k given $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$.

Overall, the conditional MSE $\zeta_{B,k}$ of the Bayesian estimate $\hat{s}_{B,k}$ is bounded as

$$\zeta_{B,k} \leq \zeta(u^*) + \sigma_s^4 \sum_{u \neq u^*} c(u)\|\mathbf{a}_0 - \mathbf{a}(u)\|_x^2\ \eta^{k-1}(u). \tag{36}$$

When $k \to \infty$, the performance of the Bayesian estimator achieves $\zeta(u^*)$. When u_0 is included in the set U, then $u^* = u_0$ and $\zeta(u^*) = \zeta_{\text{opt}}$.

4. PROOF OF THEOREM 1

Because $p(u|\mathbf{X}_k) \leq 1\ \ \forall u \in U$,

$$p(u|\mathbf{X}_k) \leq p(u|\mathbf{X}_k)^t, \quad t \in \pi_T \tag{37}$$

Case 1: $(u \neq u^*)$ From equation (18),

$$p(u|\mathbf{X}_k) \leq p(u|\mathbf{X}_k)^t = \left(\frac{p(u)p(\mathbf{X}_k|u)}{\sum_v p(v)p(\mathbf{X}_k|v)}\right)^t \tag{38}$$

$$\leq \left(\frac{p(u)p(\mathbf{X}_k|u)}{p(u^*)p(\mathbf{X}_k|u^*)}\right)^t \tag{39}$$

where the last step is obtained by removing all terms in the denominator except the term $p(u^*)p(\mathbf{X}_k|u^*)$. Taking the conditional expectation, we have

$$E[p(u|\mathbf{X}_k)|u_0] \leq E\left[\left(\frac{p(u)p(\mathbf{X}_k|u)}{p(u^*)p(\mathbf{X}_k|u^*)}\right)^t |u_0\right] \tag{40}$$

$$= E\left[\left(\frac{p(u)\prod_{i=1}^k p(\mathbf{x}_i|u)}{p(u^*)\prod_{i=1}^k p(\mathbf{x}_i|u^*)}\right)^t |u_0\right] \tag{41}$$

$$= \left(\frac{p(u)}{p(u^*)}\right)^t E\left[\left(\frac{p(\mathbf{x}_k|u)}{p(\mathbf{x}_k|u^*)}\right)^t |u_0\right]^k \tag{42}$$

We need to show that $E[(p(\mathbf{x}_k|u)/p(\mathbf{x}_k|u^*))^t|u_0]$ is strictly less than 1 [3]. We consider the difference between two Kullback Leibler distances $D(u_0||u^*)$ and $D(u_0||u)$:

$$D(u_0||u^*) - D(u_0||u)$$

$$= E\{\ln[p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]|u_0\} \tag{43}$$

$$= E\left\{\frac{d}{dt}[p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]^t|_{t=0}|u_0\right\} \tag{44}$$

$$= E\{\lim_{t \to 0} \frac{([p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]^t - 1)}{t}|u_0\}. \tag{45}$$

According to the Lebesgue dominated-convergence theorem [4], the expectation and limit can be interchanged such that

$$D(u_0||u^*) - D(u_0||u) = \lim_{t \to 0} \frac{E[p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]^t\} - 1}{t}. \tag{46}$$

For any $\delta > 0$, there exists $t = t(\delta) \in \pi_T$ such that

$$\frac{E\{[p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]^t\} - 1}{t} \leq [D(u_0||u^*) - D(u_0||u)](1 + \delta) \tag{47}$$

or

$$E\{[p(\mathbf{x}|u)/p(\mathbf{x}|u^*)]^t\} \leq 1 - t(1 + \delta)[D(u_0||u) - D(u_0||u^*)] \tag{48}$$

By the definition of u^*, $D(u_0||u) > D(u_0||u^*)$, and thus the right hand side is upper bounded by 1. This is true for all $u \neq u^*$. The left hand side is always non-negative. By defining the right hand side of (48) to be $\eta(u)$ and $(p(u)/p(u^*))^t$ to be $c(u)$, we have for $u \neq u^*$

$$E[p(u|\mathbf{X}_k)|u_0] \leq c(u)\ \eta^k(u) \tag{49}$$

where

$$0 \leq \eta(u) < 1. \tag{50}$$

Case 2: $(u = u^*)$ From equation(18),

$$p(u^*|\mathbf{X}_k) = \frac{p(u^*)p(\mathbf{X}_k|u^*)}{\sum_v p(v)p(\mathbf{X}_k|v)} \tag{51}$$

$$= \left(1 + \sum_{v \neq u^*} \frac{p(v)p(\mathbf{X}_k|v)}{p(u^*)p(\mathbf{X}_k|u^*)}\right)^{-1} \tag{52}$$

Taking expectation of both sides and by the Jensen's inequality, we have

$$E[p(u^*|\mathbf{X}_k)|u_0] \geq \left(1 + \sum_{v \neq u^*} E\left[\frac{p(v)p(\mathbf{X}_k|v)}{p(u^*)p(\mathbf{X}_k|u^*)}|u_0\right]\right)^{-1} \tag{53}$$

$$> \left(1 + \sum_{u \neq u^*} c(u)\eta^k(u)\right)^{-1}. \tag{54}$$

The last step is obtained from the result of the previous case.

Lastly, since $p(u|\mathbf{X}_k)$ is a PMF, it must lie between 0 and 1 for all u, and this completes the proof of Theorem 1.

5. DISCUSSION

The convergence rate of the conditional MSE is governed by the convergence rate of the weighting coefficient $p(u|\mathbf{X}_k)$. Each coefficients has its corresponding convergence rate governed by $\eta(u)$, which is related to the difference between two Kullback Leibler distances $D(u_0||u)$ and $D(u_0||u^*)$ according to (48). This quantity can be viewed as the theoretical measurement of the distance between u and u^* with respect to u_0, Thus, one can claim that the weighting coefficients of all $u \in U$ converge at different rates proportional to their relative distances to u^*, whereas a large distance induces a fast convergence rate and vice versa.

The signal-to-noise ratio also plays a role in determining the convergence rate. Low SNR reduces the variation of the a posteriori PMF along u. When this happens, the a posteriori PMF's of two directions may become more alike. Both Kullback Leibler distances $D(u_0||u)$ and $D(u_0||u^*)$ become closer to 0 and thus closer to each other, inducing a slower convergence rate.

6. SIMULATION

We simulate a Bayesian beamformer in the following settings: Uniform linear array with half-wavelength spacing is used. $N = 8$. $U = \{-0.5, -0.3, -0.1, 0.1, 0.3, 0.5\}$. $u_0 = 0.08$. SNR = 0dB. The a priori PMF $p(u)$ is uniform over U and thus $c(u) = 1$. The simulation results are averaged over 1000 trials. Figure 1 shows the conditional MSE of the Bayesian beamformer and the upper bound (36) in log scale.

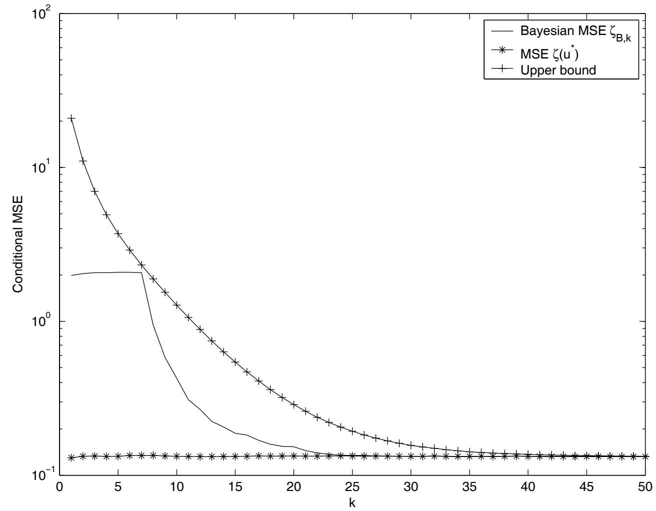

Fig. 1. Conditional MSE of the Bayesian beamformer

7. REFERENCES

[1] H. L. V. Trees, *Optimum Array Processing*. New York: Wiley Interscience, 2002.

[2] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Processing*, vol. 48, pp. 386–398, Feburary 2000.

[3] L. Liporace, "Variance of bayes estimates," *IEEE Trans. Information Theory*, vol. 17, pp. 665–669, Novenmber 1971.

[4] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1964.