COMPRESSED SENSING BLOCK MAP-LMS ADAPTIVE FILTER FOR SPARSE CHANNEL ESTIMATION AND A BAYESIAN CRAMER-RAO BOUND

H. Zayyani, M. Babaie-Zadeh*

C. Jutten

Sharif University of Technology, Department of
Electrical Engineering and Advanced Communication
Research Institute, Tehran, Iran

GIPSA-LAB, Grenoble and
Institut Universitaire de France

ABSTRACT

This paper suggests to use a Block MAP-LMS (BMAP-LMS) adaptive filter instead of an Adaptive Filter called MAP-LMS for estimating the sparse channels. Moreover to faster convergence than MAP-LMS, this block-based adaptive filter enables us to use a compressed sensing version of it which exploits the sparsity of the channel outputs to reduce the sampling rate of the received signal and to alleviate the complexity of the BMAP-LMS. Our simulations show that our proposed algorithm has faster convergence and less final MSE than MAP-LMS, while it is more complex than MAP-LMS. Moreover, some lower bounds for sparse channel estimation is discussed. Specially, a Cramer-Rao bound and a Bayesian Cramer-Rao bound is also calculated.

1. INTRODUCTION

Recently, sparse channels whose impulse responses consist of many zero taps, have gained interest in signal processing and communications [1, 2, 3]. Moreover, sparse channels have been intensively studied in geophysics and seismic, where each layer is associated to a reflection i.e. a delayed and attenuated Dirac in the time response. The mathematical model of the channel is:

$$y_l = \mathbf{x}_l^T \mathbf{w} + r_l \tag{1}$$

where y_l, $\mathbf{x}_l$, $\mathbf{w}$ and r_l are scalar channel output, $m \times 1$ input vector, $m \times 1$ sparse channel impulse response vector and the scalar additive Gaussian noise, respectively. In the training mode for channel estimation, the input vector to the channel is known and the problem is to estimate the sparse channel response $\mathbf{w}$ from the observations y_l.

Some approaches (refer to [1, 2, 3]) transmit a pulse shaped symbol through the channel and then using all the sampled received data try to estimate the channel response [2]. In [1], a Matching Pursuit (MP) algorithm is used to estimate the channel from all the received signal. [2] suggested to use some zero-tap detection schemes for channel estimation. In [3], an order-recursive Least Square MP is used for the same purpose.

One can exploit the sparsity response with adaptive filtering algorithms [4, 5]. The method in [4] is based on minimizing a regularized mean square error criterion with sparsity being promoted by the regularization term [4]. In [5], a Maximum A Posteriori (MAP) update of the adaptive filter taps is proposed which is called MAP-LMS.

Compressed Sensing (CS) is an emerging field between sampling and compression [6, 7, 8]. This field suggests to use a few random measurements of the sparse signal to reconstruct the original sparse signal.

This paper is organized as follows. In Section 2.1, we suggest to use a block-based MAP-LMS adaptive filter instead of MAP-LMS adaptive filter to increase the convergence rate. Then, in Section 2.2, we suggest to use a CS measurement matrix to reduce the sampling rate of the output channel to exploit the sparsity structure of the channel outputs. In Section 3.1, a Cramer-Rao Bound (CRB) for the problem will be computed and will be compared with the bounds derived in [9]. A Bayesian CRB is also calculated for our problem in Section 3.2. Finally, in Section 4, some simulation results are presented.

2. THE CS BLOCK MAP-LMS ADAPTIVE FILTER

First of all, we generalize the MAP-LMS adaptive filter proposed in [5] based on a block of data. Then, the update formula for the block MAP-LMS adaptive filter and some simpler updates are presented. Moreover, to estimate the channel more completely, a formula for computing the channel noise variance is suggested. This block-based adaptive filter enables us to use CS random matrices to reduce the number of samples obtained from the channel outputs. So, the CS version of the block adaptive filter is also suggested.

2.1. Block MAP-LMS Adaptive Filter (BMAP-LMS)

In our algorithm, we use a block of sparse channel outputs for updating the adaptive filter. We gather L successive output samples of channel output in (1) in a vector called output vector. So, this block of date can be represented in a vector form as:

$$\mathbf{y}_l = \mathbf{X}_l\mathbf{w} + \mathbf{r}_l \tag{2}$$

where $\mathbf{y}_l = [y_l, y_{l+1}, ..., y_{l+L-1}]^T$ is an $L \times 1$ vector obtained from channel outputs. $\mathbf{X}_l$ is an $L \times m$ matrix with the rows equal to $\mathbf{x}_l^T$, $\mathbf{x}_{l+1}^T$,... and $\mathbf{x}_{l+L-1}^T$ where $\mathbf{x}_l \triangleq [x_l, x_{l+1}, ..., x_{l+L-1}]^T$. The $m \times 1$ sparse vector $\mathbf{w}$ is the channel impulse response. Finally, $\mathbf{r}_l \triangleq [r_l, r_{l+1}, ..., r_{l+L-1}]^T$ is the Gaussian channel noise vector with the covariance matrix equal to $\sigma_n^2\mathbf{I}$.

Now, with the block-based notation of (1) in (2), we can derive our Block MAP-LMS adaptive filter (BMAP-LMS). Similar to [5], a MAP criterion can be used for estimating the vector $\mathbf{w}$ based on the block $\mathbf{y}_l$ at time index l. So, at each time index l, given the output block $\mathbf{y}_l$ and input training signals in $\mathbf{X}_l$, we want to update the vector $\mathbf{w}$ based on knowing previous vector $\mathbf{w}_{n-1}$. Hence, by MAP criterion, the posterior $p(\mathbf{w}|\mathbf{y}_l, \mathbf{w}_{n-1})$ should be maximized. Using Bayes rule, the posterior can be written as:

$$p(\mathbf{w}|\mathbf{y}_l, \mathbf{w}_{n-1}) \propto p(\mathbf{w}|\mathbf{w}_{n-1})p(\mathbf{y}_l|\mathbf{w}, \mathbf{w}_{n-1}) \tag{3}$$

where $p(\mathbf{w}|\mathbf{w}_{n-1})$ is the prior and $p(\mathbf{y}_l|\mathbf{w}, \mathbf{w}_{n-1}) = p(\mathbf{y}_l|\mathbf{w})$ is the likelihood. The log-likelihood can be easily written as:

$$\log p(\mathbf{y}_l|\mathbf{w}) = K + \frac{-1}{2\sigma_n^2}||\mathbf{y}_l - \mathbf{X}_l\mathbf{w}||_2^2 \tag{4}$$

where K is a constant not depending on $\mathbf{w}$. The log-prior can also be written as [5]:

$$\log p(\mathbf{w}|\mathbf{w}_{n-1}) \propto \frac{-1}{2}(\mathbf{w} - \mathbf{w}_{n-1})^T\mathbf{Q}^{-1}(\mathbf{w} - \mathbf{w}_{n-1}) \tag{5}$$

where a Gaussian distribution is assumed for $\mathbf{z} \triangleq \mathbf{w} - \mathbf{w}_{n-1}$ with covariance matrix $\mathbf{Q}$. So, the overall MAP update is equal to $\mathbf{w}_n = \text{argmax}_\mathbf{w} H(\mathbf{w})$ where function $H(\mathbf{w})$ is:

$$\frac{-1}{2}(\mathbf{w} - \mathbf{w}_{n-1})^T\mathbf{Q}^{-1}(\mathbf{w} - \mathbf{w}_{n-1}) + \frac{-1}{2\sigma_n^2}||\mathbf{y}_l - \mathbf{X}_l\mathbf{w}||_2^2 \tag{6}$$

To find the maximum, we write the above function in terms of $\mathbf{z}$ which is:

$$H(\mathbf{z}) = \frac{-1}{2\sigma_n^2}(\mathbf{e}_n - \mathbf{X}_l\mathbf{z})^T(\mathbf{e}_n - \mathbf{X}_l\mathbf{z}) - \frac{1}{2}\mathbf{z}^T\mathbf{Q}^{-1}\mathbf{z} \tag{7}$$

where $\mathbf{e}_n \triangleq \mathbf{y}_l - \mathbf{X}_l\mathbf{w}_{n-1}$. If $\mathbf{z}_n$ is the maximizer of H(z), then its gradient satisfies $\nabla H(\mathbf{z}_n) = 0$. Some manipulations show that the gradient is equal to:

$$\nabla H(\mathbf{z}) = \frac{1}{\sigma_n^2}\mathbf{X}_l^T\hat{\mathbf{e}} - \mathbf{Q}^{-1}\mathbf{z} \tag{8}$$

where $\hat{\mathbf{e}} \triangleq \mathbf{e}_n - \mathbf{X}_l\mathbf{z}$. It is easily obtained that $\hat{\mathbf{e}} = \mathbf{y}_l - \mathbf{X}_l\mathbf{w}$. So, $\hat{\mathbf{e}}_n = \mathbf{y}_l - \mathbf{X}_l\mathbf{w}_n = \mathbf{e}_n - \mathbf{X}_l\mathbf{z}_n$. At the maximum point we have $\frac{1}{\sigma_n^2}\mathbf{X}_l^T\hat{\mathbf{e}}_n - \mathbf{Q}^{-1}\mathbf{z}_n = 0$. Hence, $\mathbf{z}_n = \frac{1}{\sigma_n^2}\mathbf{Q}\mathbf{X}_l^T\hat{\mathbf{e}}_n$. Finally, replacing $\hat{\mathbf{e}}_n$ with the $(\mathbf{e}_n - \mathbf{X}_l\mathbf{z}_n)$ results in:

$$\mathbf{z}_n = (\mathbf{I}_m + \mathbf{C}\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{C}\mathbf{X}_l^T\mathbf{e}_n \tag{9}$$

where $\mathbf{C} \triangleq \frac{1}{\sigma_n^2}\mathbf{Q}$. Finally, the BMAP-LMS update will be:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + (\mathbf{I}_m + \mathbf{C}\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{C}\mathbf{X}_l^T\mathbf{e}_n \tag{10}$$

For the matrix $\mathbf{Q}$ which is the covariance matrix of $\mathbf{z}$, we assume that $\mathbf{Q} = \alpha\mathbf{I}_m$. Then, defining a parameter $\tau \triangleq \frac{\alpha}{\sigma_n^2}$ for BMAP-LMS, we can reach to a tradeoff in the performance of our adaptive filter. Moreover, using this parameter prevents that our adaptive filter design be dependent on the noise levels of the channel. This free parameter provides us a flexibility for choosing between the rate of convergence and the final level of Mean Square Error. It has the same role of parameter τ in [5]. Using the definition of τ, the matrix $\mathbf{C}$ will be $\mathbf{C} = \tau\mathbf{I}_m$ and the BMAP-LMS update will be:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \tau(\mathbf{I}_m + \tau\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{X}_l^T\mathbf{e}_n \tag{11}$$

where τ is the free parameter of the BMAP-LMS. Our simulations show that a small value of this parameter yields to better results. By this new formulation, we can reduce the complexity of our adaptive filter. Since the length of the sparse impulse response may be large in some applications, the value of m in some cases may be large and inversion of a matrix with a large size is very complex. So, we use a Matrix Inversion Lemma (MIL) to compute the inverse of the matrix in formula (11). This leads to the following formula:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \tau[\mathbf{I}_m - \mathbf{X}_l^T(\frac{1}{\tau}\mathbf{I}_L + \mathbf{X}_l\mathbf{X}_l^T)^{-1}\mathbf{X}_l]\mathbf{X}_l^T\mathbf{e}_n \tag{12}$$

where inversion is done with an $L \times L$ matrix instead of an $m \times m$ matrix in (10). The length of the block L can be less than m and hence using (12) is less complex than (11).

To further reduce the complexity, we can assume that matrix $\mathbf{B} \triangleq \tau\mathbf{X}_l^T\mathbf{X}_l$ will be a matrix with small elements. This can be satisfied by selecting the small input training sequences or selecting a small parameter τ for BMAP-LMS. Therefore, the the inverse matrix $(\mathbf{I}_m + \mathbf{B})^{-1}$ can be approximated by:

$$(\mathbf{I}_m + \mathbf{B})^{-1} = \mathbf{I}_m - \mathbf{B} + \mathbf{B}^2 - ... + (-1)^K\mathbf{B}^K \tag{13}$$

where only K term of the Tailor matrix series are used for the approximation. This approximation can avoid the matrix inversion in BMAP-LMS. In the simulation results, we show that approximating the inverse with (13) with a small value for K has a very low effect on decreasing the performance of BMAP-LMS.

In addition to the channel estimation, this method allows to estimate the noise level of the channel. The noise level of the channel is determined by the variance of the Gaussian noise which is σ_n. To estimate σ_n, we use the L samples of the block. If L is large, then the parameter σ_n can be approximated by $\hat{\sigma}_n \approx \frac{1}{L} \sum_{l=1}^{L} r_l^2$. Using (1) for r_l, it is obtained that $\sum_{l=1}^{L} y_l^2 \approx L\hat{\sigma}_n^2 + \sum_{l=1}^{L} \mathbf{x}_l^T \mathbf{w} \mathbf{w}^T \mathbf{x}_l$. Therefore, the following formula estimates the variance of the noise:

$$\hat{\sigma}_n^2 = \frac{||\mathbf{y}_l||_2^2 - \text{trace}(\mathbf{X}_l \mathbf{w} \mathbf{w}^T \mathbf{X}_l^T)}{L} \qquad (14)$$

2.2. Compressed Sensing BMAP-LMS (CS-BMAP-LMS)

Compressed Sensing or Compressive Sampling (CS) is an emerging field in signal processing [6, 7, 8]. The theory of CS suggests to use only a few random linear measurements of a sparse signal (in a basis) for reconstructing the original signal. The mathematical model of noise free CS is:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} \qquad (15)$$

where $\mathbf{x} = \mathbf{\Psi}\mathbf{w}$ is the original signal with length m and is sparse in the basis $\mathbf{\Psi}$ and $\mathbf{\Phi}$ is an $n \times m$ random measurement matrix where $n < m$. For near perfect recovery, in addition to the signal sparsity, the incoherence of the random measurement matrix $\mathbf{\Phi}$ with the basis $\mathbf{\Psi}$ is needed. The incoherence is satisfied with high probability for some types of random matrices such as i.i.d Gaussian elements or i.i.d Bernoulli ± 1 elements. Recent theoretical results show that under these two conditions (sparsity and incoherence), the original signal can be recovered from only a few linear measurements of the signal within a controllable error, even in the case of noisy measurements [6, 7, 8].

Since the channel response $\mathbf{w}$ is sparse, the channel output vector $\mathbf{y}_l$ is sparse in the domain $\mathbf{X}_l$ which is determined by the training sequences. So, we can use a $K \times L$ random measurement matrix $\mathbf{\Phi}$ which converts the $L \times 1$ block data $\mathbf{y}_l$ to a smaller block data $\tilde{\mathbf{y}} \triangleq \mathbf{\Phi}\mathbf{y}_l$ with K elements ($K < L$). Therefore, (2) can be written as:

$$\tilde{\mathbf{y}}_l = \tilde{\mathbf{X}}_l \mathbf{w} + \tilde{\mathbf{r}}_l \qquad (16)$$

where $\tilde{\mathbf{X}}_l \triangleq \mathbf{\Phi}\mathbf{X}_l$ and $\tilde{\mathbf{r}}_l = \mathbf{\Phi}\mathbf{r}_l$ can be viewed as the new training matrix and new noise vector. To ensure the Gaussianity of the new noise vector, we can normalize the rows of the matrix $\mathbf{\Phi}$ to have unit norm. The main advantage of the CS scheme for BMAP-LMS is the complexity reduction both in terms of hardware complexity and memory requirements. As we can see in the simulation results, another benefit of CS scheme is that the CS-BMAP-LMS has less final Mean Square Error (MSE) than BMAP-LMS and also MAP-LMS.

3. BOUNDS

3.1. Cramer-Rao bound

The CRB, which is the inverse of Fisher information matrix, bounds the performance of any unbiased parametric estimator in terms of the mean square error [10]. This lower bound gives a measure that if the performance of parameter estimation algorithms is near this lower bound or not and hence an effort is needed to solve the problem more efficiently or not. Channel response estimation can be regarded as a parametric estimation of $\mathbf{w}$ by observations $\mathbf{y}_l$ in (2). The CRB for estimating $\mathbf{w}$ in (2) is given by [10]:

$$\text{CRB} = \sigma_n^2 (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \qquad (17)$$

where this is the Cramer-Rao matrix when $\mathbf{X}_l$ is known and fixed. Similar to [11], when the matrix $\mathbf{X}_l$ is known but random which is in the case, we should add the matrix $\mathbf{X}_l$ as the additional observation[1]. So, the Fisher information matrix will be:

$$J_{ij} = E_{\mathbf{X}_l, \mathbf{y}_l} \left\{ \frac{\partial \log p(\mathbf{y}_l, \mathbf{X}_l | \mathbf{w})}{\partial w_i w_j} \right\} \qquad (18)$$

where by the Bayes rule we have $p(\mathbf{y}_l, \mathbf{X}_l | \mathbf{w}) = p(\mathbf{X}_l)p(\mathbf{y}_l | \mathbf{X}, \mathbf{w})$. So, the Fisher matrix will be:

$$\mathbf{J}_{\text{Rv,Known}} = E_{\mathbf{X}_l}\{\mathbf{J}\} \qquad (19)$$

where $\mathbf{J}_{\text{Rv,Known}}$ is the notation for Fisher information matrix when the training matrix $\mathbf{X}_l$ is known but random and $\mathbf{J} = \frac{1}{\sigma_n^2} \mathbf{X}_l \mathbf{X}_l^T$ is the Fisher matrix when the training matrix is known and fixed. Hence, $\mathbf{J}_{\text{Rv,Known}} = \frac{1}{\sigma_n^2} E\{\mathbf{X}_l \mathbf{X}_l^T\}$ where:

$$E\{\mathbf{X}_l \mathbf{X}_l^T\} = E\{\sum_{i=1}^{L} \mathbf{x}_i \mathbf{x}_i^T\} = \sum_{i=1}^{L} E\{\mathbf{x}_i \mathbf{x}_i^T\} \qquad (20)$$

where $E\{\mathbf{x}_i \mathbf{x}_i^T\}$ is the covariance matrix of $\mathbf{x}_i$ and is equal to $\sigma_r^2 \mathbf{I}$ since the training sequences are independent and zero mean random variables with variance σ_r^2. Therefore, the CRB will be:

$$\text{CRB}_{\text{Rv,Known}} = \frac{\sigma_n^2}{L\sigma_r^2} \mathbf{I}_m = \frac{1}{L.\text{SNR}} \mathbf{I}_m \qquad (21)$$

where $\text{SNR} \triangleq \frac{\sigma_r^2}{\sigma_n^2}$ is a measure of Signal to Noise Ratio (SNR) and $L \gg 1$ can be regarded as the number of training sequences. Then, if we add all inequalities $E\{(w_i - \hat{w}_i)^2\}) \geq \text{CRB}_{ii}$, we have the following CRB for the ℓ^2-norm of the error $(\mathbf{w} - \hat{\mathbf{w}})$:

$$E[||\hat{\mathbf{w}} - \mathbf{w}||_2^2] \geq \frac{m\sigma_n^2}{L\sigma_r^2} \qquad (22)$$

---

[1] Because knowledge of $\mathbf{X}_l$ effects the estimation of $\mathbf{w}$.

In addition to the CRB, in [9], a deterministic lower MSE for an oracle channel estimator is computed as:

$$E[||\mathbf{w}^* - \mathbf{w}||_2^2] \geq \frac{S\sigma_n^2}{||\mathbf{x}_l||_2^2} \quad (23)$$

where S is the number of nonzero taps in sparse vector $\mathbf{w}$ and $\mathbf{w}^*$ is a special oracle estimator which is defined in [9]. We prove that this bound is compatible with our CRB and we also generalize it to all oracle estimators which know the nonzero locations of the channel response. If we know the location of active taps, we can write $E[||\mathbf{w}^* - \mathbf{w}||_2^2] = \sum_{i=1}^{S} E[(w_i - \hat{w}_i)^2]$. Each error variance satisfying (21), the CRB oracle estimator is:

$$E[||\mathbf{w}^* - \mathbf{w}||_2^2] \geq \frac{S\sigma_n^2}{L\sigma_r^2} \quad (24)$$

where $\mathbf{w}^*$ is a general oracle estimator (not a deterministic one as in [9]). The oracle CRB bound (24) is approximately the same as the deterministic oracle bound (23) because we can write $||\mathbf{x}_l||_2^2 \approx L\sigma_r^2$ for large values of L.

3.2. Bayesian Cramer-Rao bound

The Posterior Cramer-Rao Bound (PCRB) or Bayesian Cramer-Rao Bound (BCRB) of a vector of parameters $\boldsymbol{\theta}$ estimated from data vector $\mathbf{y}$ is the inverse of the Fisher information matrix, and bounds the estimation error in the following form [12]:

$$E\left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\right] \geq \mathbf{J}^{-1} \quad (25)$$

where $\hat{\boldsymbol{\theta}}$ is the estimate of $\boldsymbol{\theta}$ and $\mathbf{J}$ is the Fisher information matrix with the elements [12]:

$$J_{ij} = E_{\mathbf{y},\boldsymbol{\theta}}\left[-\frac{\partial^2 \log p(\mathbf{y},\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] \quad (26)$$

where $p(\mathbf{y},\boldsymbol{\theta})$ is the joint Probability Density Function (PDF) between the observations and the parameters. Unlike CRB, the BCRB (25) is satisfied for any estimator (even for biased estimators) under some mild conditions [13], [12] which we assume that are fulfilled in our problem. Using Bayes rule, the Fisher information matrix can be decomposed into two matrices [12]:

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P \quad (27)$$

where $\mathbf{J}_D$ represents data information matrix and $\mathbf{J}_P$ represents prior information matrix which their elements are [12]:

$$J_{D_{ij}} \triangleq E_{\mathbf{y},\boldsymbol{\theta}}\left[-\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = E_{\boldsymbol{\theta}}(J_{s_{ij}}) \quad (28)$$

$$J_{P_{ij}} \triangleq E_{\boldsymbol{\theta}}\left[-\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] \quad (29)$$

where $\mathbf{J}_s$ is the standard Fisher information matrix [10] and $p(\boldsymbol{\theta})$ is the prior distribution of the parameter vector.

Now, we want to find the BCRB for our problem which is the estimation of $\mathbf{w}$ by the observations $\mathbf{y}_l$. In this case, following the previous section, the standard Fisher information matrix is $\mathbf{J}_s = L.\text{SNR}.\mathbf{I}_m$. It is independent of $\mathbf{w}$ which is the parameter vector. Hence, the data information matrix is equal to $\mathbf{J}_D = \mathbf{J}_s = L.\text{SNR}.\mathbf{I}_m$.

To compute the prior information matrix $\mathbf{J}_P$ from (29), we should assume a sparse prior distribution for our parameter vector elements w_i. We assume w_i's are independent and have a Gaussian distribution similar to [14]:

$$p(w_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp(-\frac{w_i^2}{2\sigma_i^2}) \quad (30)$$

where the variance σ_i^2 determines the prior information about the corresponding coefficient. It can be easily seen that in this case, the prior information matrix is:

$$\mathbf{J}_P = \text{diag}(\frac{1}{\sigma_i^2}) \quad (31)$$

Finally, the BCRB results in:

$$E\left[(w_i - \hat{w}_i)^2\right] \geq \left(L.\text{SNR} + \frac{1}{\sigma_i^2}\right)^{-1} \quad (32)$$

and finally the bound derived from CRB for the ℓ^2-norm of the error $(\mathbf{w} - \hat{\mathbf{w}})$ is:

$$E[||\hat{\mathbf{w}} - \mathbf{w}||_2^2] \geq \sum_{i=1}^{m} \left(L.\text{SNR} + \frac{1}{\sigma_i^2}\right)^{-1} \quad (33)$$

If we have only S nonzero elements of the channel response with the variance σ_w^2 and the other elements are zero (variances are equal to zero), then the above BCRB is reduced to:

$$E[||\hat{\mathbf{w}} - \mathbf{w}||_2^2] \geq S\left(L.\text{SNR} + \frac{1}{\sigma_w^2}\right)^{-1} \quad (34)$$

If we assume $L.\text{SNR} \gg \frac{1}{\sigma_w^2}$, then we can neglect $\frac{1}{\sigma_w^2}$ in comparison to $L.\text{SNR}$. So, the final approximated BCRB is the same as the oracle bound (24).

4. SIMULATION RESULTS

In this section, we investigate our proposed adaptive filter for sparse channel estimation. In our experiment, we used a sparse channel impulse response with length $m = 100$ and with only 10% nonzero coefficients (i.e., $S = 10$). The training sequence x_l is selected randomly from a zero-mean Gaussian random variable with variance $\sigma_r^2 = 1$. We used $T = 2000$ samples of the training input sequence and so we had $T - m = 1900$ samples of the channel output. The
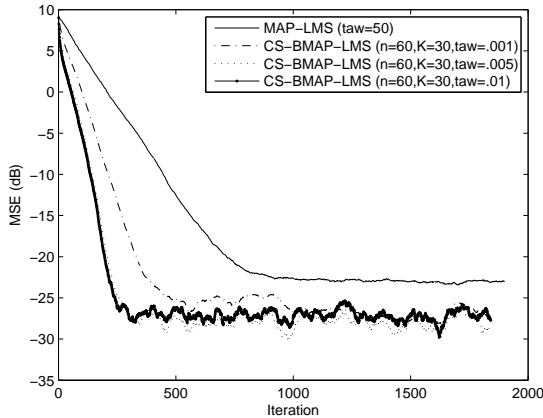
Fig. 1. The performance of CS-BMAP-LMS for various values of τ in comparison to MAP-LMS with the fastest convergence ($\tau = 50$).
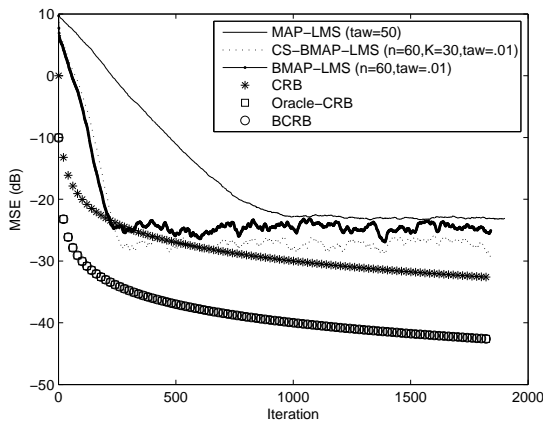
Fig. 2. Performance of CS-BMAP-LMS in comparison with BMAP-LMS, MAP-LMS and also with CRB's.

standard deviation of the Gaussian noise is selected as $\sigma_n = 0.1$. To compare the performances, we compute the MSE between the true sparse vector $\mathbf{w}$ and the estimated vector $\hat{\mathbf{w}}$ which is defined as $\mathrm{MSE(dB)} = 10\log_{10}||\mathbf{w} - \hat{\mathbf{w}}||_2^2$.

In the first experiment, we investigated the effect of parameter τ on the performance of the CS-BMAP-LMS. The effect of τ on the performance of MAP-LMS is investigated in [5]. Here, we just report the performance of MAP-LMS for $\tau = 50$ because it results in the fastest rate of convergence among all values of τ. Similar to [5], the parameter γ was selected as $\gamma = 0.98$ (refer to [5] for the details). In CS-BMAP-LMS, we used the block length as $n = 60$ and then a random measurement matrix $\boldsymbol{\Phi}$ with $K = 30$ rows and $L = 60$ columns is used to reduce the sampling rate by a factor of 0.5. Figure 1 shows the simulation results of

CS-BMAP-LMS for various values of τ in comparison to MAP-LMS with the fastest convergence ($\tau = 50$). As we can see the best value for parameter τ with respect to the convergence rate is $\tau = 0.01$. We use this value for the next experiment.

In the second experiment, we compared the CS-BMAP-LMS with BMAP-LMS and MAP-LMS. For MAP-LMS and CS-BMAP-LMS, the parameters are selected as the first experiment. The CS-BMAP-LMS was compared with BMAP-LMS with the same block length $n = 60$ and with the same parameter $\tau = .01$. This parameter is selected for fastest rate of convergence as explained in the first experiment. Figure 2 shows the simulation results. It can be seen that the CS-BMAP-LMS and BMAP-LMS are faster than MAP-LMS (the convergence is approximately three to four times faster than MAP-LMS convergence). This fast rate of convergence is very useful in the cases where we have rapid time-varying channels [3]. Another benefit of CS-BMAP-LMS and BMAP-LMS is that they have lower final MSE than MAP-LMS. Moreover, CS-BMAP-LMS has slightly lower final MSE than BMAP-LMS.

To compare the complexity of our proposed algorithms and MAP-LMS, we use the average CPU time of the simulations. Our simulations were performed in MATLAB7.0 environment using an AMD Athlon Dual core 4600 with 896 MB of RAM and under Windows Xp operating system. The average simulation times for CS-BMAP-LMS, BMAP-LMS and MAP-LMS are 12.2, 13.2 and 3.6 seconds for processing 1900 samples of the channel output. So, our suggested methods are approximately four times more complex in terms of simulation time.

Finally, we compare the performance of BMAP-LMS and CS-BMAP-LMS with CRB (33) and Oracle bound (24) and BCRB (34). As it was stated, since $L.\mathrm{SNR} \gg \frac{1}{\sigma_w^2}$ is satisfied in our simulations, the BCRB is coincided with the oracle bound. The final note is that there is a gap between the performance of the adaptive filters and the BCRB. So, in spite of devising the fast adaptive filters with high convergence rate (what we did in this paper), enhancing the performance (reducing the final MSE) of the adaptive filters is remained as future works.

5. CONCLUSIONS

In this paper, we introduced a block-based adaptive filter for estimating sparse channels. We also suggested to use a CS scheme of the previous adaptive filter to exploit the sparsity of the channel outputs. We also calculated a CRB and a BCRB where training data are zero-mean random variables. Our simulation results show that the proposed algorithm has faster convergence rate than MAP-LMS algorithm and less final MSE, while it is more complex than MAP-LMS. We also showed that with some conditions, the BCRB is ap-

proximately equivalent to the Oracle bound (24).

6. REFERENCES

[1] S.F. Cotter and B.D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Comm.*, vol. 50, pp. 374–377, March 2002.

[2] S. Vedantam C. Carbonelli and U. Mitra, "Sparse channel estimation with zero tap detection," *IEEE Trans. Wireless Comm.*, vol. 6, pp. 1743–1763, May 2007.

[3] Li. Weichang and J.C. Preisig, "Estimation of rapidly time-varying sparse channels," *IEEE Journal of Oceanic Engineering*, vol. 32, pp. 927–939, October 2007.

[4] B.D. Rao and S. Bongyong, "Adaptive filtering algorithms for promoting sparsity," *ICASSP'03*, pp. 361–364, 2003.

[5] G. Deng, "Partial update and sparse adaptive filters," *IET Signal Proc.*, vol. 1, pp. 9–17, March 2007.

[6] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, April 2006.

[7] E. Candes, "Near-optimal signal recovery from random projections: universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, December 2006.

[8] R. Baraniuk, "Compressive sensing," *IEEE Signal. Process. Magazine*, vol. 24, pp. 118–121, July 2007.

[9] G. Raz W.U. Bajwa, J. Haupt and R. Nowak, "Compressed channel sensing," *CISS 2008*, pp. 5–10, 2008.

[10] S.M. Kay, "Fundamentals of statistical signal processing: Estimation theory," *Prentice-Hall*, 1993.

[11] Y. C. Eldar A. Wiesel and A. Yeredor, "Linear regression with gaussian model uncertainty: Algorithms and bounds," *IEEE Trans. Signal. Processing*, vol. 56, pp. 2194–2205, June 2008.

[12] C. H. Muravchik P. Tichavsky and A. Nehorai, "Posterior cramer-rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal. Processing*, vol. 46, pp. 1386–1395, May 1998.

[13] H.L. Van Trees, "Detection, estimation and modulation theory," *Wiley*, 1968.

[14] D. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. Signal. Processing*, vol. 52, pp. 2153–2164, August 2004.