CRAMER RAO BOUNDS FOR CHANNEL ESTIMATION WITH SYMBOL A PRIORI INFORMATION

Xavier Wautelet, Cédric Herzet, Luc Vandendorpe

Communications and Remote Sensing Laboratory, Université catholique de Louvain
place du Levant 2, B-1348 Louvain-la-Neuve, Belgium
{wautelet,herzet,vandendorpe}@tele.ucl.ac.be

ABSTRACT

In this paper we derive the Cramer-Rao bound (CRB) for the estimation of a multiple-input multiple-output (MIMO) channel when a given a priori information about the transmitted sequence is available. This approach also enables to compute the non-data-aided (NDA) bound, i.e. the CRB when the receiver only knows the symbol constellation. We explore various scenarios, including the case of a deterministic channel and the case of a random channel. Numerical evaluations of the CRB-based bounds are presented for different channels. In particular, the influence of the symbol-constellation size is highlighted. Finally, the performance of an estimator based on the expectation-maximization algorithm is shown to be very close to the proposed bounds.

1. INTRODUCTION

The Cramer-Rao bound (CRB) is a lower bound on the variance of any unbiased estimator. This bound is often used as a benchmark for assessing the performance of parameter estimators. In the context of channel estimation, the evaluation of the CRB is most of the time a computationally-demanding task. Therefore, most of the papers dealing with the CRB in the current literature focus on particular cases. In particular, the data-aided (DA) CRB may be straightforwardly calculated by applying the equations in [1]. Another particular case is the so-called "blind" CRB: in this context, the transmitted symbols are assumed to be continuous with uniform a priori distribution. Doing this assumption, several authors have derived an analytical expression of the CRB, see e.g. [2], [3]. In this paper, we consider a more general case: we assume that a priori probabilities on the transmitted symbols are available at the receiver. We propose a CRB for multiple-input multiple-output (MIMO) channel estimation which takes into account these symbol a priori probabilities. The proposed approach enables to continuously sweep all the cases from the non-data aided (NDA) to the DA case. To the best of the authors' knowledge, the NDA CRB had never been calculated before in the channel-estimation case. We compare the NDA CRB derived in this paper to the so-called "blind" CRB derived previously.

In the wireless context, the transmitted signal often faces channels with fading. In this paper we compare three different approaches for computing CRB's which take the channel fading into account. The sequel of this paper is organized as follows. In section 2, the system model is depicted. The CRB given a channel value and a symbol a priori distribution is derived in section 3. Three different bounds for random-channel estimation are compared in section 4. Finally in section 5, numerical results are presented: the different CRB's are computed and we compare the performance achieved by a channel estimator based on the expectation-maximization (EM) algorithm [4] to the CRB's computed in the previous sections. In particular, we show that the performance of this kind of estimator is very close to the CRB.

This research is partly funded by the Federal Office for Scientific, Technical and Cultural Affairs (OSTC, Belgium) through the IAP contract No P5/11.

2. SYSTEM MODEL

We consider a channel with n_T inputs and n_R outputs. At the transmitter, bursts of $n_T L_s$ complex symbols $s_k^{(i)}$ are sent, where i $(1 \leq i \leq n_T)$ denotes the channel input and k $(1 \leq k \leq L_s)$ is a time index. The signal received at time k at output j is therefore

$$r_k^{(j)} = \sum_{i=1}^{n_T} \sum_{l=0}^{L-1} h_l^{(i,j)} \, s_{k-l}^{(i)} + n_k^{(j)}, \tag{1}$$

where $h_l^{(i,j)}$ $(0 \leq l \leq L-1)$ is the l^{th} tap of the length-L channel impulse response from input i to output j, and $n^{(j)}(k)$ denotes the additive white Gaussian variance-σ_n^2 noise at time k at output j. To simplify the notations, vectors $\underline{s}$, $\underline{h}$ and $\underline{r}$ will be used. They respectively contain the transmitted symbols, the channel taps and the received samples.

3. CRB FOR DETERMINISTIC CHANNELS

In this section, we derive a CRB for the estimation of a deterministic channel which takes into account a priori probabilities on the transmitted symbols that are available at the receiver.

3.1. CRB with A Priori Probabilities on the Symbols

Let $\underline{h}_R$ denote the vector made up with the real part $\Re\{\underline{h}\}$ and the imaginary part $\Im\{\underline{h}\}$ of the channel taps i.e.

$$\underline{h}_R = [\Re\{\underline{h}\}^T \ \Im\{\underline{h}\}^T]^T. \tag{2}$$

Then, the error variance of any unbiased estimator $\hat{\underline{h}}_R$ of $\underline{h}_R$, i.e. any estimator $\hat{\underline{h}}_R$ which satisfies

$$E_{\underline{r}|\underline{h}}[\hat{\underline{h}}_R] = \underline{h}_R, \tag{3}$$

is lower-bounded by the Cramer-Rao bound i.e.

$$E_{\underline{r}|\underline{h}}[(\hat{\underline{h}}_R - \underline{h}_R)(\hat{\underline{h}}_R - \underline{h}_R)^T] \geq \text{CRB}(\underline{h}_R). \tag{4}$$

The Cramer-Rao bound relative to the estimation of parameter $\underline{h}_R$ may be evaluated as follows

$$\text{CRB}(\underline{h}_R) = \underline{\underline{J}}^{-1}(\underline{h}_R). \tag{5}$$

Matrix $\underline{\underline{J}}(\underline{h}_R)$ denotes the Fisher information matrix whose element at row l and column k is given by

$$\{\underline{\underline{J}}(\underline{h}_R)\}_{l,k} = E_{\underline{r}|\underline{h}_R}\left[\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}\,\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_k}\right]_{\big|\tilde{\underline{h}}_R=\underline{h}_R}, \tag{6}$$

where $\tilde{\underline{h}}_R$ is a derivation variable and $\{\tilde{\underline{h}}_R\}_k$ denotes the k^{th} element of vector $\tilde{\underline{h}}_R$. The evaluation of (6) may be performed in efficient way [5] by expressing factor $\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}$ as a function of the symbol a posteriori probability $p(\underline{s}|\underline{r},\underline{h}_R)$ i.e. (see the appendix)

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}_{\big|\tilde{\underline{h}}_R=h_R} = E_{\underline{s}|\underline{r},\underline{h}_R}\left[\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l}\right]. \tag{7}$$

Note that since the channel noise is white and Gaussian with variance σ_n^2, we have

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}},\underline{s})}{\partial\Re\{\tilde{h}_l^{(i,j)}\}}_{\big|\tilde{\underline{h}}=\underline{h}} = \frac{2}{\sigma_n^2}\sum_{k=1}^{L_s}\Re\Bigg\{s_{k-l}^{(i)*}r_k^{(j)} - \sum_{i'=1}^{n_T}\sum_{l'=0}^{L-1} h_{l'}^{(i',j)} s_{k-l'}^{(i')} s_{k-l}^{(i)*}\Bigg\}$$

and

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}},\underline{s})}{\partial\Im\{\tilde{h}_l^{(i,j)}\}}_{\big|\tilde{\underline{h}}=\underline{h}} = \frac{2}{\sigma_n^2}\sum_{k=1}^{L_s}\Im\Bigg\{s_{k-l}^{(i)*}r_k^{(j)} - \sum_{i'=1}^{n_T}\sum_{l'=0}^{L-1} h_{l'}^{(i',j)} s_{k-l'}^{(i')} s_{k-l}^{(i)*}\Bigg\}.$$

Introducing then the following notations:

$$\eta_k^{(i)} = E_{\underline{s}|\underline{r},\underline{h}_R}[s_k^{(i)}] \tag{8}$$

$$\rho_{k,k'}^{(i,i')} = E_{\underline{s}|\underline{r},\underline{h}_R}[s_k^{(i)}\ s_{k'}^{(i')*}], \tag{9}$$

we finally have

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}})}{\partial\Re\{\tilde{h}_l^{(i,j)}\}}_{\big|\tilde{\underline{h}}=\underline{h}} = \frac{2}{\sigma_n^2}\sum_{k=1}^{L_s}\Re\Bigg\{\eta_{k-l}^{(i)*}r_k^{(j)} - \sum_{i'=1}^{n_T}\sum_{l'=0}^{L-1} h_{l'}^{(i',j)} \rho_{k-l',k-l}^{(i',i)}\Bigg\} \tag{10}$$

and

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}})}{\partial\Im\{\tilde{h}_l^{(i,j)}\}}_{\big|\tilde{\underline{h}}=\underline{h}} = \frac{2}{\sigma_n^2}\sum_{k=1}^{L_s}\Im\Bigg\{\eta_{k-l}^{(i)*}r_k^{(j)} - \sum_{i'=1}^{n_T}\sum_{l'=0}^{L-1} h_{l'}^{(i',j)} \rho_{k-l',k-l}^{(i',i)}\Bigg\}. \tag{11}$$

The CRB may therefore be calculated by using (5), (6), (10) and (11). In most cases $\eta_k^{(i)}$ and $\rho_{k,l}^{(i,j)}$ cannot be calculated analytically. They can however be evaluated by means of numerical computations. A maximum a posteriori (MAP) equalizer for example is able to deliver $p(s_k^{(i)}|\underline{r},\underline{h}_R)$ based on given $\underline{r}$ and $p(\underline{s})$. This quantity is sufficient to compute $\eta_k^{(i)}$. For the computation of $\rho_{k,l}^{(i,j)}$, more details can be found in [4]. Since $\eta_k^{(i)}$ and $\rho_{k,l}^{(i,j)}$ depend on $\underline{r}$, the Fisher information matrix has to be averaged over many bursts of symbols to numerically perform the mathematical expectation with respect to $\underline{\text{r}}$ in (6).

Note that no assumption about the symbol knowledge have been made in the derivation of the CRB expressions. They are therefore valid regardless of the symbol a priori distribution $p(\underline{s})$. In particular, the NDA case may be computed by considering that all the transmitted sequences are a priori equiprobable. Note that this bound on the contrary to the blind ones [3], [2] consider that the receiver knows the symbol constellation used at the transmitter. To the best of the authors' knowledge, the NDA CRB with this side information has never been proposed although some guidelines for the calculation of this CRB using a totally different approach has been given in [2].

3.2. CRB vs. the Mutual Information

In the previous section, the CRB has been derived for a fixed a priori probabilities on the transmitted symbols $p(\underline{s})$. In order to better understand how the a priori information about the transmitted symbols modifies the CRB we propose another lower bound which does not depend on a particular a probability $p(\underline{s})$ but rather on the mutual information (MI) between these probabilities and the transmitted sequence $\underline{s}$. There are many probability sets $p(\underline{s})$ that correspond to a given quantity of mutual information. Once MI has been chosen, a way to find such probability sets is to generate them randomly as in [6] for the EXIT chart analysis of iterative decoding. The probabilities $p(\underline{s})$ can then be considered as random variables depending on MI. Note that MI $= 0$ corresponds to the NDA case whereas MI $= 1$ leads to the DA case. For a given mutual information MI, the variance of any unbiased channel estimator is then lower-bounded as

$$E_{\underline{r}|\underline{h},\mathrm{MI}}\left[\left(\hat{\underline{h}}_R-\underline{h}_R\right)\left(\hat{\underline{h}}_R-\underline{h}_R\right)^T\right]\geq E_{p(\underline{s})|\mathrm{MI}}\left[\underline{\underline{J}}^{-1}(\underline{h}_R)\right], \tag{12}$$

where the elements of $\underline{\underline{J}}^{-1}(\underline{h}_R)$ are defined in (6). Using the Jensen's inequality for matrices [7], we obtain

$$E_{\underline{r}|\underline{h},\mathrm{MI}}\left[\left(\hat{\underline{h}}_R-\underline{h}_R\right)\left(\hat{\underline{h}}_R-\underline{h}_R\right)^T\right]\geq\left(E_{p(\underline{s})|\mathrm{MI}}\left[\underline{\underline{J}}(\underline{h})\right]\right)^{-1}. \tag{13}$$

The expectation with respect to $p(\underline{s})$ may be approximated by an arithmetical average over many bursts. Not that the bound defined in (13) is not an actual CRB since it has been modified by the Jensen's inequality. However, in the sequel we will be referred to this bound as the CRB for a given amount of mutual information ($\mathrm{CRB}_{\mathrm{MI}}$).

4. CRB FOR RANDOM CHANNELS

The CRB derived in section 3.1 is valid for a given channel value $\underline{h}$. However, in a wireless environment channel is usually affected by fading and is therefore random. In this case, a bound for assessing the average performance of an estimator is more interesting than a bound only valid for a given channel value. In this section we present three different bounds which take the random nature of the channel into account.

4.1. Classical CRB for a Random Channel

In [1], the CRB for the estimation of a random parameter is derived. It requires the estimate to be unbiased on average, i.e.

$$E_{\underline{h}_R,\underline{r}}\left[\hat{\underline{h}}_R\right]=m_{\underline{h}_R}, \tag{14}$$

where $m_{\underline{h}_R}$ denotes the mean of distribution $p(\underline{h}_R)$. Under this constraint one may show [1] that the error variance of any estimator $\hat{\underline{h}}_R$ is lower-bounded by

$$E_{\underline{h},\underline{r}}\left[(\hat{\underline{h}}_R-\underline{h}_R)(\hat{\underline{h}}_R-\underline{h}_R)^T\right]\geq\mathrm{CRB}_{\mathrm{Rand}}. \tag{15}$$

$\mathrm{CRB}_{\mathrm{Rand}}$ is given by

$$\mathrm{CRB}_{\mathrm{Rand}}=\left(E_{\underline{h}_R}\left[\underline{\underline{J_2}}(\underline{h}_R)\right]\right)^{-1}, \tag{16}$$

where $\underline{\underline{J_2}}(\underline{h}_R)$ is a matrix whose elements are

$$\{\underline{\underline{J_2}}(\underline{h}_R)\}_{l,k}=E_{\underline{r}|\underline{h}_R}\left[\frac{\partial\ln p(\underline{r},\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}\,\frac{\partial\ln p(\underline{r},\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_k}\right]_{\Big|\tilde{\underline{h}}_R=\underline{h}_R}. \tag{17}$$

Matrix $\underline{\underline{J_2}}(\underline{h}_R)$ is different from the Fisher information matrix $\underline{\underline{J}}(\underline{h}_R)$ in (6). The derivative of the joint probability of the channel and the received signal appears in (17) instead of the conditional probability of the received signal in (6). $\mathrm{CRB}_{\mathrm{Rand}}$ is actually a lower bound for estimators which know the joint probability $p(\underline{r},\underline{h}_R)$ and thus the a priori channel distribution $p(\underline{h}_R)$. In the next section, we derive a bound for assessing the performance of estimators which do not know the a priori distribution of the channel.

4.2. CRB for a Conditionally Unbiased Estimator

Let us consider an estimator $\hat{\underline{h}}_R$ of $\underline{h}_R$ which does not have any a priori information about the channel distribution $p(\underline{h}_R)$. Assume that this estimator is conditionally unbiased i.e.

$$E_{\underline{r}|\underline{h}_R}[\hat{\underline{h}}_R]=\underline{h}_R. \tag{18}$$

Then, using results from section 3.1, it may easily be shown that the error variance of estimator $\hat{\underline{h}}_R$ is lower-bounded as follows

$$E_{\underline{r},\underline{h}_R}\left[\left(\hat{\underline{h}}_R-\underline{h}_R\right)\left(\hat{\underline{h}}_R-\underline{h}_R\right)^T\right]\geq\mathrm{CRB}_{\mathrm{CU}}, \tag{19}$$

where

$$\mathrm{CRB}_{\mathrm{CU}}=E_{\underline{h}_R}\left[\underline{\underline{J}}^{-1}(\underline{h}_R)\right]. \tag{20}$$

and $\underline{\underline{J}}(\underline{h}_R)$ is computed according to (6).

$\mathrm{CRB}_{\mathrm{CU}}$ is not very convenient to compute because of the double average over the channel and the received signal that is separated by a matrix inversion. Using the Jensen's inequality for matrices [7] on (20), we may show that

$$\mathrm{CRB}_{\mathrm{CU2}}=\left(E_{\underline{h}_R}\left[\underline{\underline{J}}(\underline{h}_R)\right]\right)^{-1} \tag{21}$$

is also a lower bound on the variance of unbiased estimators because

$$\mathrm{CRB}_{\mathrm{CU2}}\leq\mathrm{CRB}_{\mathrm{CU}}. \tag{22}$$

$\mathrm{CRB}_{\mathrm{CU2}}$ is much easier to compute than $\mathrm{CRB}_{\mathrm{CU}}$. The expression of $\mathrm{CRB}_{\mathrm{CU2}}$ is close to the one of $\mathrm{CRB}_{\mathrm{Rand}}$ except that matrix $\underline{\underline{J}}(\underline{h}_R)$ is averaged in (21) instead of $\underline{\underline{J_2}}(\underline{h}_R)$ in (15).

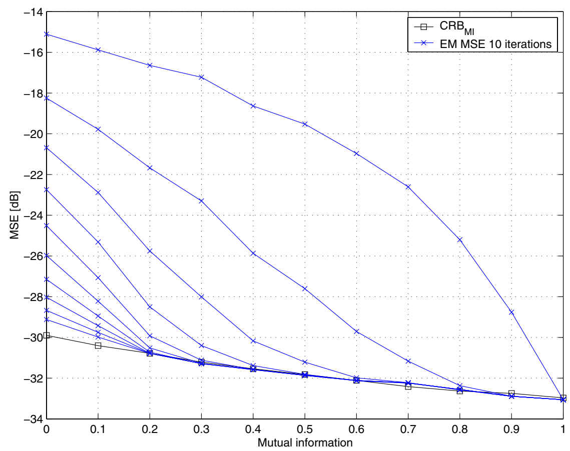

Fig. 1. Comparison of the CRB and the MSE of the EM channel estimates. Porat channel. $Es/N_0 = 0$ dB. 2004 BPSK symbols.

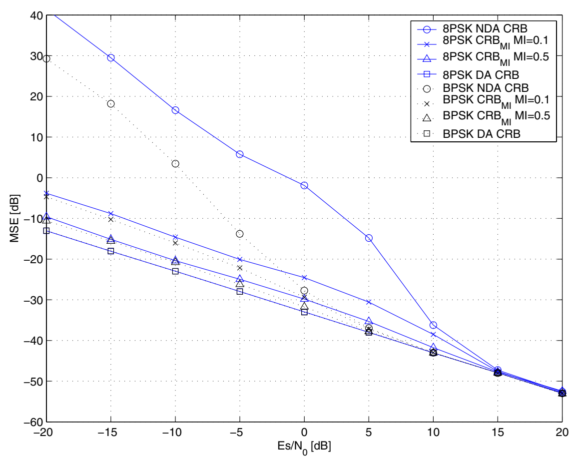

Fig. 2. Influence of the constellation size on the CRB. Proakis B channel. 2004 symbols.

5. NUMERICAL RESULTS

Numerical evaluation of the various lower bounds are presented in this section. A simulation environment is used to compute the required quantities like $\eta_k^{(i)}$ and $\rho_{k,k'}^{(i,i')}$. Random symbol bursts of $n_T \times 2004$ symbols are sent over the MIMO channel and $\eta_k^{(i)}$ and $\rho_{k,k'}^{(i,i')}$ are computed by a MAP equalizer. Unless otherwise stated, the modulation is BPSK. The averaging over $\underline{r}$ is approximated by sending 5000 bursts. When the results have to be averaged over $\underline{h}$, the channel changes at each burst except for the CRB_{CU}.

In Fig. 1, the CRB_{MI} is plotted versus the mutual information. The considered channel is the single-input single-output Porat channel whose complex channel impulse response is $[2 - 0.4j; 1.5 + 1.8j; 1; 1.2 - 1.3j; 0.8 + 1.6j]$.

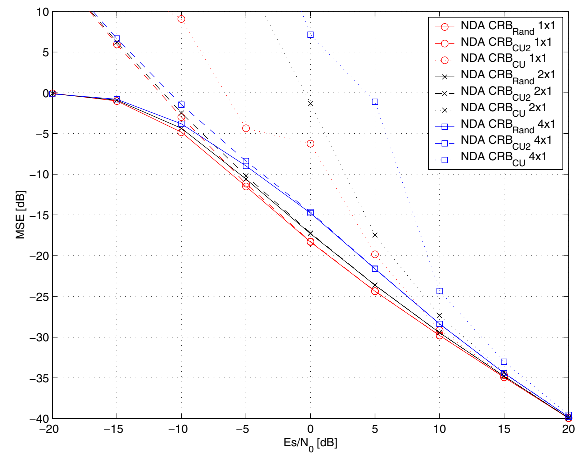

Fig. 3. Comparison of the NDA CRB_{Rand}, the NDA CRB_{CU} and the NDA CRB_{CU2} over flat Rayleigh-fading channels with {1,2,4} inputs and one output. 100 BPSK symbols.

The symbol constellation is BPSK and the Es/N_0 is equal to 0 dB. Fig. 1 shows that the CRB_{MI} decreases as the MI increases. The mean square error (MSE) of the EM channel estimator [4] is compared to this lower bound. The quality of the initial estimate provided to the EM estimator has been fixed to -10 dB. Despite this quite bad initialization, the EM estimator variance is very close to the CRB_{MI} after 10 iterations.

The influence of the constellation size on the CRB is shown in Fig. 2. The bounds are computed for the single-input single-output Proakis B channel whose real coefficients are $[0.407; 0.815; 0.407]$. It can be noticed that all the bounds, regardless of the symbol mutual information and of the constellation size, converge to the DA CRB at high Es/N_0 ratios. The NDA CRB and the CRB_{MI} with $\text{MI} = 0.1$ benefit from a smaller constellation size at low to medium Es/N_0. The knowledge of the constellation points helps the channel estimator and this knowledge brings a bigger amount of information in the BPSK case than in the 8PSK case. When MI is larger, the relative amount of information brought by the knowledge of the constellation points is smaller. It is why the CRB_{MI} with $\text{MI} = 0.5$ is not affected by the constellation size. Of course, the DA CRB is identical regardless of the constellation.

In Fig. 3, the NDA CRB_{Rand}, the NDA CRB_{CU} and the NDA CRB_{CU2} for random channels are compared. The considered channel is a flat Rayleigh-fading channel with 1, 2 or 4 inputs and one output. The CRB_{CU} is averaged over 5000 different channels and 5000 bursts of 100 BPSK symbols have been sent over each channel. For low Es/N_0, the knowledge of the channel distribution is helpful to the channel estimator. The CRB_{Rand} is always smaller than the two others. Note that in the DA case, both bounds are equal.

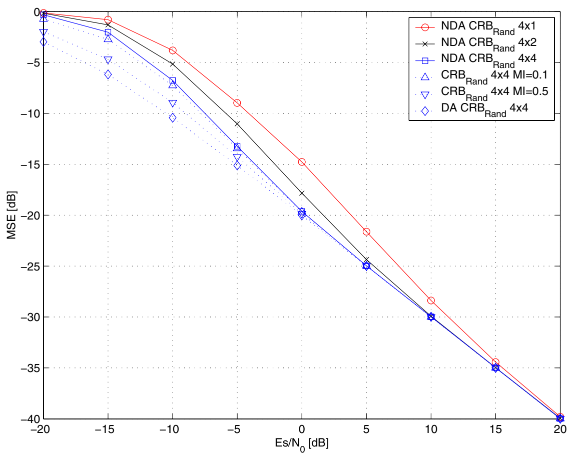

Fig. 4. Influence of the number of channel outputs on the $\mathrm{CRB}_{\mathrm{Rand}}$ over flat Rayleigh-fading channels with 4 inputs and $\{1,2,4\}$ outputs. 100 BPSK symbols.

Fig. 4 shows that the $\mathrm{CRB}_{\mathrm{Rand}}$ decreases with the number of output for a flat MIMO Rayleigh-fading channel. This is probably due to the receive diversity which enables to better estimate the transmitted symbols when the number of outputs increases. Channel estimators can then benefit from these better symbol estimates to improve the channel estimates. As for the MI, it is not very useful in the 4-input 4-output channel case except for very low Es/N_0.

6. CONCLUSION

In this paper we have derived CRB's for assessing the performance of channel estimators when some a priori information about the transmitted sequence is available at the receiver. Several cases have been considered. We have first considered the calculation of the CRB when the channel is deterministic. Expressions have been given both for the case of fixed symbol a posteriori probabilities and the case of fixed mutual information. Secondly, the problem of computing a lower bound in the case of random channel has been tackled. Finally, numerical results have been presented. The different CRB-based lower bounds have been computed in different scenarios and the performance of a channel estimator based on the EM algorithm have been shown to be very close to the proposed lower bounds.

7. REFERENCES

[1] J. M. Mendel, "Lessons in estimation theory for signal processing, communications, and control," Prentice Hall Signal Processing Series, Englewood Cliffs (NJ), USA, 1995.

[2] B. M. Sadler, R. J. Kozick, T. Moore, "Bounds on bearing and symbol estimation with side information," *IEEE Trans. Signal Processing,* vol. 49, pp. 822-834, Apr. 2001.

[3] E. De Carvalho, D.T.M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.,* pp. 129-132, Paris, France, Apr. 16-18, 1997.

[4] G. K. Kaleh, R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.,* vol. 42, pp. 2406-2413, July 1994.

[5] N. Noels, H. Steendam, M. Moeneclaey, "On the Cramer-Rao lower bound and the performance of synchronizers for (turbo) encoded systems," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.,* Lisbon, Portugal, July 11-14, 2004.

[6] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.,* vol. 49, pp. 1727-1737, Oct. 2001.

[7] R. Ravikanth, S.P. Meyn, "Bounds on achievable performance in the identification and adaptive control of time-varying systems," *IEEE Trans. Automatic Control,* vol. 44, pp. 670-682, Apr. 1999.

A. APPENDIX

As in [5] in the synchronization-parameter case, the CRB for the channel estimates can be expressed in terms of the a posteriori probabilities on the symbols $p(\underline{s}|\underline{r},\underline{h}_R)$. Let $\mathcal{S}$ be the set of all the possible symbol bursts, then

$$\begin{aligned}\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l} &= \frac{1}{p(\underline{r}|\tilde{\underline{h}}_R)}\frac{\partial p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}\\ &= \sum_{\underline{s}\in\mathcal{S}}\frac{\partial p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l}\frac{p(\underline{s})}{p(\underline{r}|\tilde{\underline{h}}_R)}. \end{aligned} \quad (23)$$

Knowing that

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l} = \frac{1}{p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}\frac{\partial p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l}, \quad (24)$$

(23) becomes:

$$\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l} = \sum_{\underline{s}\in\mathcal{S}}\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l}\frac{p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})p(\underline{s})}{p(\underline{r}|\tilde{\underline{h}}_R)}. \quad (25)$$

Using the Bayes' rule, we get:

$$\left.\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R)}{\partial\{\tilde{\underline{h}}_R\}_l}\right|_{\tilde{\underline{h}}_R=\underline{h}_R} = E_{\underline{s}|\underline{r},\underline{h}_R}\left[\frac{\partial \ln p(\underline{r}|\tilde{\underline{h}}_R,\underline{s})}{\partial\{\tilde{\underline{h}}_R\}_l}\right] \quad (26)$$