

## **A Cramer-Rao Bound Characterization of the EM-Algorithm Mean Speed of Convergence**

Cédric Herzet, Valéry Ramon, Alexandre Renaux, Luc Vandendorpe

► **To cite this version:**

Cédric Herzet, Valéry Ramon, Alexandre Renaux, Luc Vandendorpe. A Cramer-Rao Bound Characterization of the EM-Algorithm Mean Speed of Convergence. IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers, 2008, 56 (6), pp.2218-2228. <inria-00444732>

**HAL Id: inria-00444732**

**<https://hal.inria.fr/inria-00444732>**

Submitted on 7 Jan 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Cramér-Rao Bound Characterization of the EM-Algorithm Mean Speed of Convergence

Cédric Herzet, Valéry Ramon, Alexandre Renaux, and Luc Vandendorpe, *Fellow, IEEE*

**Abstract**—This paper deals with the *mean speed of convergence* of the expectation-maximization (EM) algorithm. We show that the asymptotic behavior (in terms of the number of observations) of the EM algorithm can be characterized as a function of the Cramér-Rao bounds (CRBs) associated to the so-called *incomplete* and *complete* data sets defined within the EM-algorithm framework. We particularize our result to the case of a complete data set defined as the concatenation of the observation vector and a vector of nuisance parameters, independent of the parameter of interest. In this particular case, we show that the CRB associated to the complete data set is nothing but the well-known modified CRB. Finally, we show by simulation that the proposed expression enables to properly characterize the EM-algorithm mean speed of convergence from the CRB behavior when the size of the observation set is large enough.

**Index Terms**—Convergence of numerical methods, iterative methods, maximum-likelihood estimation.

## I. INTRODUCTION

SINCE its first statement by Dempster, Laird, and Rubin [1], the expectation-maximization (EM) algorithm has become a popular numerical method to compute maximum-likelihood (ML) estimates, see, e.g., [2] and references therein. Among the reasons of its success, its low complexity of implementation and its robustness are usually pointed out [2]. The main drawback of the EM algorithm is however its speed of convergence which, in some situations, may be extremely slow. In their seminal paper, Dempster, Laird, and Rubin [1] showed that the EM algorithm exhibits a linear speed of convergence, with a rate of convergence obtained from the information matrices associated to the *missing* and *complete* data sets. More recently, some authors [3], [4] have given further insights into the EM-algorithm convergence. In particular, in [3], [4], the authors emphasize that the EM algorithm may locally achieve quasi-Newton behavior in some specific situations.

In this paper, we address<sup>1</sup> the problem of the characterization of the EM-algorithm *mean* speed of convergence. We derive an

expression relating the *asymptotic* (in terms of number of observations) mean speed of convergence of the EM algorithm to the Cramér-Rao bounds (CRBs) associated to the *incomplete* and *complete* data sets. In particular, we emphasize that, as long as the number of observation is large, the proposed expression enables to have a good intuition of the EM-algorithm speed of convergence simply by looking at the CRB behavior. The paper is organized as follows. In Section II, we recall the basics of the EM algorithm and we give the general expressions of the CRB and modified CRB (MCRB) [6]. Section III is the core of the paper: we derive an expression relating the EM-algorithm mean speed of convergence to the CRBs associated to the *incomplete* and *complete* data estimation problems. In a first part, we show that the proposed CRB-based expression is asymptotically valid under some conditions. Then, we particularize the proposed expression to the case where the complete data set is the concatenation of the received observation and a vector of *nuisance* parameters independent of the parameter of interest. In this particular case, we show that the CRB associated to the complete data set is equal to the MCRB. Finally, we briefly discuss the complexity associated to the evaluation of the proposed expression. In Section IV, we illustrate the relevance of the proposed approach in two different examples. We consider the estimation of the mean in a Gaussian mixture problem and the estimation of the carrier phase offset in a digital communication system. In particular, we emphasize that the EM algorithm behavior may be well predicted from the knowledge of the CRB and MCRB when the number of observations is large.

## II. ML ESTIMATION, EM ALGORITHM, AND CRBS

In this section, we briefly review some notions which will be useful in the remainder of this paper. In Section II-A, we present the EM algorithm and discuss some of its features. In Section II-B, we recall the main equations of the standard and the modified CRBs.

### A. ML Estimation and the EM Algorithm

Let us consider an observation vector  $\mathbf{r}$  depending on an unknown deterministic scalar parameter  $b$ . The ML estimate of  $b$  given  $\mathbf{r}$  is the solution of the following maximization problem:

$$\hat{b}_{\text{ML}} = \arg \max_b p(\mathbf{r} | \tilde{b}) \quad (1)$$

where  $\tilde{b}$  is a trial value of  $b$ . The ML estimator enjoys very good asymptotic statistical properties but its evaluation is unfortunately quite complex in a number of practical problems [7].

In order to circumvent this issue, iterative ML search methods have been proposed in the literature. In particular, the EM algorithm, proposed by Dempster, Laird, and Rubin in [1], is a

Manuscript received April 3, 2007; revised December 6, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andreas Jakobsson.

C. Herzet is with the INRIA/IRISA, Campus Universitaire de Beaulieu, 35042 Rennes, France (e-mail: cherzet@inria.fr).

V. Ramon and L. Vandendorpe are with the Communications and Remote Sensing Laboratory, Université Catholique de Louvain, B1348 Louvain-la-Neuve, Belgium (e-mail: ramon@tele.ucl.ac.be; vandendorpe@tele.ucl.ac.be).

A. Renaux is with the Laboratoire des Signaux et Systemes, University Paris-Sud 11, Plateau de Moulon, 91192 Gif-sur-Yvette, France (e-mail: renaux@ieee.org).

Digital Object Identifier 10.1109/TSP.2008.917024

<sup>1</sup>This contribution is the extension of the following paper [5] by the same authors.

powerful iterative method which has been shown to provide a suitable solution to a number of problems encountered in the technical literature [2]. Formally, the EM algorithm is based on the following two steps:

$$\text{E-step: } \mathcal{Q}(\tilde{b}, \hat{b}^{(n)}) = \int_{\mathcal{Z}} p(\mathbf{z}|\mathbf{r}, \hat{b}^{(n)}) \log p(\mathbf{z}, \mathbf{r}|\tilde{b}) d\mathbf{z} \quad (2)$$

$$\text{M-step: } \hat{b}^{(n+1)} = \arg \max_{\tilde{b}} \mathcal{Q}(\tilde{b}, \hat{b}^{(n)}) \quad (3)$$

where  $\hat{b}^{(n)}$  is the estimate computed by the EM algorithm at iteration  $n$  and  $\mathbf{r} = f(\mathbf{z})$ , where  $f(\mathbf{z})$  is a *many-to-one* mapping.  $\mathcal{Z}$  is the set of values that  $\mathbf{z}$  can take on. Vectors  $\mathbf{r}$  and  $\mathbf{z}$  are often referred to as the *incomplete* and the *complete* data sets, respectively.

Since the EM algorithm is an iterative method, the question of its speed of convergence naturally arises. Dempster, Laird, and Rubin showed in their seminal paper [1] that the convergence of the EM algorithm is locally linear, i.e., we have in a neighborhood of  $\hat{b}_{\text{ML}}$  that

$$e^{(n+1)}(\mathbf{r}) = C(\mathbf{r}) e^{(n)}(\mathbf{r}) \quad (4)$$

where  $e^{(n)}(\mathbf{r}) = |\hat{b}^{(n)}(\mathbf{r}) - \hat{b}_{\text{ML}}(\mathbf{r})|$  and  $C(\mathbf{r})$  is the *rate of convergence* of the EM algorithm. The authors showed moreover that the rate of convergence of the EM algorithm may be related to the amount of *missing information*<sup>2</sup> in the considered problem, i.e.,

$$C(\mathbf{r}) = I_c^{-1}(\mathbf{r}) I_m(\mathbf{r}) \quad (5)$$

where  $I_c(\mathbf{r})$  and  $I_m(\mathbf{r})$  are, respectively, the information matrices associated to the complete and the missing data, i.e.,

$$I_c(\mathbf{r}) \triangleq - \left( \int_{\mathcal{Z}} p(\mathbf{z}|\mathbf{r}, \tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}, \mathbf{r}|\tilde{b}) d\mathbf{z} \right)_{\tilde{b}=\hat{b}_{\text{ML}}(\mathbf{r})} \quad (6)$$

$$I_m(\mathbf{r}) \triangleq - \left( \int_{\mathcal{Z}} p(\mathbf{z}|\mathbf{r}, \tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}|\mathbf{r}, \tilde{b}) d\mathbf{z} \right)_{\tilde{b}=\hat{b}_{\text{ML}}(\mathbf{r})}. \quad (7)$$

### B. The Standard and the Modified CRBs

The standard [7] and the modified [6] CRBs are lower bounds on the mean square estimation error of any unbiased estimator. In the case of the estimation of a scalar parameter  $b$  from a received observation vector  $\mathbf{r}$ , the (standard) CRB may be expressed as the inverse of the Fisher information matrix, i.e.,

$$\text{CRB}_{\mathbf{r}}(b) = - \left( E_{\mathbf{r}|b} \left[ \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r}|\tilde{b})_{\tilde{b}=b} \right] \right)^{-1} \quad (8)$$

$$= \left( E_{\mathbf{r}|b} \left[ \left( \frac{\partial}{\partial \tilde{b}} \log p(\mathbf{r}|\tilde{b})_{\tilde{b}=b} \right)^2 \right] \right)^{-1} \quad (9)$$

where  $\tilde{b}$  is a derivation variable.

<sup>2</sup>The missing information may actually be seen as the difference between the amount of information contained in the complete data set and the incomplete data set.

In some situations, the standard CRB may be quite tedious to evaluate. In particular, when the received observations  $\mathbf{r}$  also depend on a nuisance parameter vector  $\mathbf{a}$ , the evaluation of  $p(\mathbf{r}|b)$  may require a huge summation, see, e.g., [8]. In such situations, we may use the modified CRB (MCRB) [6], which is easier to compute but looser than the standard CRB, i.e.,  $\text{CRB}_{\mathbf{r}}(b) \geq \text{MCRB}(b)$ . Formally, the MCRB is defined as

$$\text{MCRB}(b) = - \left( E_{\mathbf{r}, \mathbf{a}|b} \left[ \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r}|\mathbf{a}, \tilde{b})_{\tilde{b}=b} \right] \right)^{-1}. \quad (10)$$

To conclude this subsection, let us mention that the MCRB has been shown in [9] to be the high-SNR asymptote of the standard CRB, i.e., the CRB and the MCRB coincide when the SNR tends to infinity.

### III. A CRB-BASED EXPRESSION OF THE EM MEAN SPEED OF CONVERGENCE

As discussed in the introduction, the local convergence of the EM algorithm given a *particular* observation vector  $\mathbf{r}$  has already been well studied in the literature. In some situations, however, one may be interested in the *average* speed of convergence of the EM algorithm; the average being taken over the distribution of the observations  $\mathbf{r}$ . In this section, we will focus on this problem. In particular, we will consider the evolution of the following quantity throughout the iterations:

$$E_{\mathbf{r}|b} [e^{(n+1)}(\mathbf{r})] = \int_{\mathcal{R}} e^{(n+1)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \quad (11)$$

where  $\mathcal{R}$  is the set of values that  $\mathbf{r}$  can take on.

In Section III-A, we will show that this quantity may be expressed as

$$E_{\mathbf{r}|b} [e^{(n+1)}(\mathbf{r})] = M_C E_{\mathbf{r}|b} [e^{(n)}(\mathbf{r})] \quad (12)$$

where

$$M_C = 1 - \frac{\text{CRB}_{\mathbf{z}}(b)}{\text{CRB}_{\mathbf{r}}(b)} \quad (13)$$

when the size of the observation vector  $\mathbf{r}$  tends to infinity and under some regularity conditions. In other words, we will emphasize that, in the asymptotic regime, the evolution of the mean absolute error  $E_{\mathbf{r}|b} [e^{(n)}(\mathbf{r})]$  is linear and that the factor of proportionality  $M_C$  is only a function of the ratio of two CRBs. In the sequel, we will refer to  $M_C$  as the *mean convergence rate* (MCR).

In Section III-B, we will emphasize that the MCR can be related to the well-known MCRB when the complete data set is made up of the concatenation of the observation vector  $\mathbf{r}$  and a vector of nuisance parameter  $\mathbf{a}$  independent of  $b$ . In such a case, we show that

$$M_C = 1 - \frac{\text{MCRB}(b)}{\text{CRB}_{\mathbf{r}}(b)}. \quad (14)$$

Finally, in Section III-C we discuss the practical evaluation of the MCR and the associated complexity.

### A. Asymptotic Speed of Convergence of the EM Algorithm

In this section, we show that (12)–(13) enable to characterize the EM-algorithm behavior in the asymptotic regime. In order to do so, we will proceed in two steps. First, we will show that,  $\forall \epsilon > 0$ ,  $E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})]$  may be lower and upper bounded as follows:

$$\begin{aligned} & \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^- \left( \int_{\mathcal{R}} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} - \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}} \right) \\ & \leq E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})] \\ & \leq \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^+ \int_{\mathcal{R}} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} + \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}}. \end{aligned} \quad (15)$$

where  $\sigma$  is the maximum of the standard deviations of  $I_m(\mathbf{r})$  and  $I_c(\mathbf{r})$ , and

$$\begin{aligned} \text{MSE}_{\text{EM}}^{(n)} & \triangleq \int_{\mathcal{R}} \left| \hat{b}_{\text{ML}}(\mathbf{r}) - \hat{b}^{(n)}(\mathbf{r}) \right|^2 p(\mathbf{r}|b) d\mathbf{r}, \quad (16) \\ K^- & \triangleq \frac{1 - \frac{\epsilon}{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}}{1 + \frac{\epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]}} \\ K^+ & \triangleq \frac{1 + \frac{\epsilon}{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}}{1 - \frac{\epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]}}. \end{aligned} \quad (17)$$

Then, in a second part, we will show that these bounds tend to  $M_C E_{\mathbf{r}|b}[e^{(n)}]$  when the size of the observation vector tends to infinity.

Let us first show (15). We will assume that  $\hat{b}^{(n)}$  is in a neighborhood of  $\hat{b}_{\text{ML}}$  with probability 1. Therefore, using (4) we have

$$E_{\mathbf{r}|b}[e^{(n+1)}] = \int_{\mathcal{R}} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r}. \quad (18)$$

Let us define,  $\forall \epsilon > 0$ , the following subspaces of the observation space  $\mathcal{R}$

$$\mathcal{R}_\epsilon = \{ \mathbf{r} | |I_m(\mathbf{r}) - E_{\mathbf{r}|b}[I_m(\mathbf{r})]| \leq \epsilon; |I_c(\mathbf{r}) - E_{\mathbf{r}|b}[I_c(\mathbf{r})]| \leq \epsilon \} \quad (19)$$

$$\bar{\mathcal{R}}_\epsilon = \mathcal{R} \setminus \mathcal{R}_\epsilon. \quad (20)$$

Since  $\mathcal{R}_\epsilon \cup \bar{\mathcal{R}}_\epsilon = \mathcal{R}$  and  $\mathcal{R}_\epsilon \cap \bar{\mathcal{R}}_\epsilon = \emptyset$ , we have

$$\begin{aligned} & \int_{\mathcal{R}} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & = \int_{\mathcal{R}_\epsilon} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & \quad + \int_{\bar{\mathcal{R}}_\epsilon} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r}. \end{aligned} \quad (21)$$

In order to find a lower and an upper bound on the left-hand side of (21), we will derive a lower and an upper bound on each term

of the right-hand side. Let us first consider the first term. Using the definition of  $\mathcal{R}_\epsilon$ , we can lower-bound and upper-bound  $C(\mathbf{r})$  as follows:

$$\begin{aligned} C(\mathbf{r}) & = \frac{I_m(\mathbf{r})}{I_c(\mathbf{r})} \\ & \leq \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})] + \epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})] - \epsilon} = \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} \frac{1 + \frac{\epsilon}{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}}{1 - \frac{\epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]}} \quad (22) \\ & \geq \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})] - \epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})] + \epsilon} = \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} \frac{1 - \frac{\epsilon}{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}}{1 + \frac{\epsilon}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]}}. \end{aligned} \quad (23)$$

Hence, using the definition of  $K^-$  and  $K^+$  in (17) we have

$$\begin{aligned} & \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^- \int_{\mathcal{R}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & \leq \int_{\mathcal{R}_\epsilon} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & \leq \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^+ \int_{\mathcal{R}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r}. \end{aligned} \quad (24)$$

Let us now derive a bound on the second term in (21). Since  $0 \leq C(\mathbf{r}) \leq 1$ ,  $e^{(n)}(\mathbf{r}) \geq 0$ , and  $p(\mathbf{r}|b) \geq 0$ , we have

$$\begin{aligned} 0 & \leq \int_{\bar{\mathcal{R}}_\epsilon} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & \leq \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r}. \end{aligned} \quad (25)$$

Moreover, by Cauchy-Schwarz we have that

$$\begin{aligned} & \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r}|b) d\mathbf{r} \\ & \leq \sqrt{\int_{\bar{\mathcal{R}}_\epsilon} (e^{(n)}(\mathbf{r}))^2 p(\mathbf{r}|b) d\mathbf{r}} \sqrt{\int_{\bar{\mathcal{R}}_\epsilon} p(\mathbf{r}|b) d\mathbf{r}}, \quad (26) \\ & \leq \sqrt{\int_{\bar{\mathcal{R}}_\epsilon} (e^{(n)}(\mathbf{r}))^2 p(\mathbf{r}|b) d\mathbf{r}} \sqrt{\int_{\bar{\mathcal{R}}_\epsilon} p(\mathbf{r}|b) d\mathbf{r}}. \end{aligned} \quad (27)$$

Now we have that

$$\int_{\bar{\mathcal{R}}_\epsilon} p(\mathbf{r}|b) d\mathbf{r} = \Pr\{\mathbf{r} \in \bar{\mathcal{R}}_\epsilon\}, \quad (28)$$

$$\begin{aligned} & \leq \max(\Pr\{|I_m(\mathbf{r}) - E_{\mathbf{r}|b}[I_m(\mathbf{r})]| > \epsilon\} \\ & \quad \Pr\{|I_c(\mathbf{r}) - E_{\mathbf{r}|b}[I_c(\mathbf{r})]| > \epsilon\}) \end{aligned} \quad (29)$$

since  $\mathbf{r} \in \bar{\mathcal{R}}_\epsilon$  is defined by the intersection of two events. Moreover, by Chebychev we have that the probability that  $I_m(\mathbf{r})$  and  $I_c(\mathbf{r})$  are “ $\epsilon$ -away” from their mean is upper bounded by

$$\Pr\{|I_m(\mathbf{r}) - E_{\mathbf{r}|b}[I_m(\mathbf{r})]| > \epsilon\} \leq \frac{\sigma_m^2}{\epsilon^2} \quad (30)$$

$$\Pr\{|I_c(\mathbf{r}) - E_{\mathbf{r}|b}[I_c(\mathbf{r})]| > \epsilon\} \leq \frac{\sigma_c^2}{\epsilon^2} \quad (31)$$

where  $\sigma_m^2$  (respectively,  $\sigma_c^2$ ) is the variance of  $I_m(\mathbf{r})$  (respectively,  $I_c(\mathbf{r})$ ). Defining

$$\sigma = \max(\sigma_m, \sigma_c) \quad (32)$$

we have, therefore

$$0 \leq \int_{\bar{\mathcal{R}}_\epsilon} C(\mathbf{r}) e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \leq \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}}. \quad (33)$$

Combining (24) and (33), we can, therefore, bound  $E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})]$  as follows:

$$\begin{aligned} & \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^- \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \\ & \leq E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})] \\ & \leq \frac{E_{\mathbf{r}|b}[I_m(\mathbf{r})]}{E_{\mathbf{r}|b}[I_c(\mathbf{r})]} K^+ \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} + \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}}. \end{aligned} \quad (34)$$

Furthermore, we have

$$\int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \leq \int_{\mathcal{R}} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \quad (35)$$

$$\begin{aligned} & \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \\ & = \int_{\mathcal{R}} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} - \int_{\bar{\mathcal{R}}_\epsilon} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} \end{aligned} \quad (36)$$

$$\geq \int_{\mathcal{R}} e^{(n)}(\mathbf{r}) p(\mathbf{r} | b) d\mathbf{r} - \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}}. \quad (37)$$

Inequality (35) follows from the nonnegativity of  $e^{(n)}(\mathbf{r})$  and  $p(\mathbf{r} | b)$ . Equations (36) and (37), respectively, follows from the fact that  $\bar{\mathcal{R}}_\epsilon \cup \bar{\mathcal{R}}_\epsilon = \mathcal{R}$  and from (33). Using (35) and (37), we finally end up with (15).

Let us now consider the bounds in (15) when the number of observations, say  $N$ , tends to infinity and under the following conditions:

$$\lim_{N \rightarrow \infty} E_{\mathbf{r}|b}[(\hat{b}_{\text{ML}} - b)^2] = 0 \quad (38)$$

$$\lim_{N \rightarrow \infty} \frac{\text{CRB}_{\mathbf{r}}}{\text{CRB}_{\mathbf{z}}} \neq 1 \quad (39)$$

$$\sigma < \infty. \quad (40)$$

First, taking (38) into account, we also have asymptotically that

$$E_{\mathbf{r}|b}[I_c(\mathbf{r})] = \text{CRB}_{\mathbf{z}}^{-1} \quad (41)$$

$$E_{\mathbf{r}|b}[I_m(\mathbf{r})] = \text{CRB}_{\mathbf{z}}^{-1} - \text{CRB}_{\mathbf{r}}^{-1}. \quad (42)$$

Relations (42) and (41) can be shown as follows. Using the definition of the complete-data information matrix (6) and taking the expectation with respect to  $p(\mathbf{r} | b)$ , we have

$$\begin{aligned} E_{\mathbf{r}|b}[I_c(\mathbf{r})] & = - \int_{\mathcal{R}} p(\mathbf{r} | b) \left[ \int_{\mathcal{Z}} p(\mathbf{z} | \mathbf{r}, \tilde{b}) \right. \\ & \quad \left. \times \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}, \mathbf{r} | \tilde{b}) dz \right]_{|\tilde{b}=\hat{b}_{\text{ML}}(\mathbf{r})} d\mathbf{r}. \end{aligned} \quad (43)$$

From (38), we have that  $\hat{b}_{\text{ML}} \simeq b$  asymptotically and therefore

$$\begin{aligned} & E_{\mathbf{r}|b}[I_c(\mathbf{r})] \\ & = \left[ - \int_{\mathcal{R}} \int_{\mathcal{Z}} p(\mathbf{z}, \mathbf{r} | \tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}, \mathbf{r} | \tilde{b}) dz d\mathbf{r} \right]_{|\tilde{b}=b} \\ & = \left[ - \int_{\mathcal{Z}} p(\mathbf{z} | \tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}, \mathbf{r} | \tilde{b}) dz \right]_{|\tilde{b}=b}. \end{aligned} \quad (44)$$

Now, we have that

$$\begin{aligned} & \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}, \mathbf{r} | \tilde{b}) \\ & = \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r} | \mathbf{z}, \tilde{b}) + \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z} | \tilde{b}) \end{aligned} \quad (45)$$

$$= \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z} | \tilde{b}) \quad (46)$$

since<sup>3</sup>  $p(\mathbf{r} | \mathbf{z}, \tilde{b}) = \mathbb{1}\{\mathbf{r} = f(\mathbf{z})\}$  and is, therefore, not a function of  $\tilde{b}$ . Plugging (46) in (44), we get

$$E_{\mathbf{r}|b}[I_c(\mathbf{r})] = \left[ - \int_{\mathcal{Z}} p(\mathbf{z} | \tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z} | \tilde{b}) dz \right]_{|\tilde{b}=b}. \quad (47)$$

Comparing (47) with (8), we see that the right-hand side (RHS) of (47) is nothing but the inverse of the CRB associated to  $b$  and based on the complete data set  $\mathbf{z}$ . This shows (41). Let us now consider (42). First, notice that  $I_m(\mathbf{r})$  and  $I_c(\mathbf{r})$  may be related [2] as

$$- \left( \frac{\partial}{\partial \tilde{b}^2} \log p(\mathbf{r} | \tilde{b}) \right)_{|\tilde{b}=\hat{b}_{\text{ML}}(\mathbf{r})} = I_c(\mathbf{r}) - I_m(\mathbf{r}). \quad (48)$$

Based on (48), we may write

$$E_{\mathbf{r}|b}[I_m(\mathbf{r})] = E_{\mathbf{r}|b} \left[ I_c(\mathbf{r}) + \left( \frac{\partial}{\partial \tilde{b}^2} \log p(\mathbf{r} | \tilde{b}) \right)_{|\tilde{b}=\hat{b}_{\text{ML}}(\mathbf{r})} \right]. \quad (49)$$

Using the fact that  $\hat{b}_{\text{ML}} \simeq b$  from (38), we finally have

$$\begin{aligned} & E_{\mathbf{r}|b}[I_m(\mathbf{r})] \\ & = \text{CRB}_{\mathbf{z}}^{-1}(b) + E_{\mathbf{r}|b} \left[ \left( \frac{\partial}{\partial \tilde{b}^2} \log p(\mathbf{r} | \tilde{b}) \right)_{|\tilde{b}=b} \right] \end{aligned} \quad (50)$$

$$= \text{CRB}_{\mathbf{z}}^{-1}(b) - \text{CRB}_{\mathbf{r}}^{-1}(b) \quad (51)$$

where (50) follows from (41), and (51) follows from the definition of the CRB (8). This shows (42).

Plugging (38), (39), and (40) into (15), we get

$$\begin{aligned} & M_C K_\infty^- \left( E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})] - \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}} \right) \\ & \leq E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})] \\ & \leq M_C K_\infty^+ E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})] + \frac{\sigma}{\epsilon} \sqrt{\text{MSE}_{\text{EM}}^{(n)}}. \end{aligned} \quad (52)$$

<sup>3</sup> $\mathbb{1}\{S\}$  is the indicator function which is equal to 1 if the statement  $S$  is true and 0 otherwise.

where

$$K_{\infty}^{-} \triangleq \frac{1 - \frac{\epsilon}{\text{CRB}_{\mathbf{r}}^{-1}}}{1 + \frac{\epsilon}{\text{CRB}_{\mathbf{z}}^{-1} - \text{CRB}_{\mathbf{r}}^{-1}}}$$

$$K_{\infty}^{+} \triangleq \frac{1 + \frac{\epsilon}{\text{CRB}_{\mathbf{r}}^{-1}}}{1 - \frac{\epsilon}{\text{CRB}_{\mathbf{z}}^{-1} - \text{CRB}_{\mathbf{r}}^{-1}}}. \quad (53)$$

Note that (38) also implies that  $\lim_{N \rightarrow \infty} \text{CRB}_{\mathbf{r}} = 0$  since the CRB is a lower bound on the mean square error. Therefore, as long as (39) holds we have for any  $\epsilon < \infty$  that

$$K_{\infty}^{-} = 1, \quad K_{\infty}^{+} = 1. \quad (54)$$

Finally, notice that our initial assumption that  $\hat{b}^{(n)}$  is (with probability 1) in a neighborhood of  $\hat{b}_{\text{ML}}$  ensures that  $\text{MSE}_{\text{EM}}^{(n)} < \infty$ . Therefore, since (52) is valid  $\forall \epsilon < \infty$ , we have

$$M_C E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})] - \epsilon'$$

$$\leq E_{\mathbf{r}|b}[e^{(n+1)}(\mathbf{r})] \leq M_C E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})] + \epsilon' \quad \forall \epsilon' > 0. \quad (55)$$

This proves (12).

Let us pause a moment to discuss this result. As far as our building assumptions are valid, we just showed that (12) and (13) establish a relationship between the rate of improvement of the mean absolute error  $E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})]$  and the CRBs associated to the complete and incomplete data sets. In particular, we see from (13) that the (mean) rate at which the EM algorithm converges to the ML estimate decreases a function of the ratio  $\text{CRB}_{\mathbf{z}}/\text{CRB}_{\mathbf{r}}$ . This ratio is actually a measure of the improvement of the estimation quality which can be achieved by observing the complete-data set instead of the incomplete-data set. Note that (see the Appendix),

$$0 \leq \text{CRB}_{\mathbf{z}} \leq \text{CRB}_{\mathbf{r}} \quad (56)$$

and, therefore, by (13)

$$0 \leq M_C \leq 1. \quad (57)$$

This implies that the mean absolute estimation error will be non-increasing at each iteration, which is in good accordance with the convergence properties of the EM algorithm [1], [10]. Note also from (13) that the larger the ratio  $\text{CRB}_{\mathbf{z}}/\text{CRB}_{\mathbf{r}}$ , the slower the mean convergence of the EM algorithm. Interestingly, this result is in good accordance with the existing convergence results [1], [2] according to which the speed of convergence of the EM algorithm depends on the amount of *missing* information between the incomplete and the complete data sets.

### B. Mean Convergence Rate: Nuisance Parameter Case

In this section, we particularize the MCR expression (13) to the case where the complete data set is defined as  $\mathbf{z} \triangleq [\mathbf{r}^T, \mathbf{a}^T]^T$ , where  $\mathbf{a}$  is a vector of *nuisance parameters* affecting  $\mathbf{r}$  and independent of  $b$ . This particular case often occurs in digital communication systems where we have to estimate some channel

parameter  $b$  independent of the transmitted symbol sequence  $\mathbf{a}$ . We show that, in this particular case, the MCR is written as

$$M_C = 1 - \frac{\text{MCRB}(b)}{\text{CRB}_{\mathbf{r}}(b)}. \quad (58)$$

To prove (58), we show that  $\text{CRB}_{\mathbf{z}}^{-1}(b) = \text{MCRB}^{-1}(b)$  when  $\mathbf{z} \triangleq [\mathbf{r}^T, \mathbf{a}^T]^T$ . Using the definition of both the CRB and the complete data set  $\mathbf{z}$ , we, respectively, have

$$\text{CRB}_{\mathbf{z}}^{-1}(b) = \left[ - \int_{\mathbf{z}} p(\mathbf{z}|\tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{z}|\tilde{b}) d\mathbf{z} \right]_{|\tilde{b}=b} \quad (59)$$

$$= \left[ - \int_{\mathcal{A}} \int_{\mathcal{R}} p(\mathbf{r}, \mathbf{a}|\tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r}, \mathbf{a}|\tilde{b}) d\mathbf{r} d\mathbf{a} \right]_{|\tilde{b}=b} \quad (60)$$

Using the Bayes rule and the independence between  $\mathbf{a}$  and  $b$ , we finally have

$$\text{CRB}_{\mathbf{z}}^{-1}(b) = \left[ - \int_{\mathcal{A}} \int_{\mathcal{R}} p(\mathbf{r}, \mathbf{a}|\tilde{b}) \times \frac{\partial^2}{\partial \tilde{b}^2} \left( \log p(\mathbf{r}|\mathbf{a}, \tilde{b}) + \log p(\mathbf{a}) \right) d\mathbf{r} d\mathbf{a} \right]_{|\tilde{b}=b} \quad (61)$$

$$= \left[ - \int_{\mathcal{A}} \int_{\mathcal{R}} p(\mathbf{r}, \mathbf{a}|\tilde{b}) \frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r}|\mathbf{a}, \tilde{b}) d\mathbf{r} d\mathbf{a} \right]_{|\tilde{b}=b}. \quad (62)$$

Comparing (62) with (10), we see that the RHS of (62) corresponds to the definition of  $\text{MCRB}^{-1}(b)$ . This shows (58).

### C. Practical Evaluation of the MCR

In this section, we briefly discuss the complexity associated to the evaluation of the MCR. In particular, it is interesting to compare the complexity associated to the computation of the MCR to the complexity pertaining to the evaluation of the standard convergence rate  $C(\mathbf{r})$  defined in (5).

On the one hand, the evaluation of  $C(\mathbf{r})$  implies: i) the computation of  $\hat{b}_{\text{ML}}$ ; ii) the evaluation of the information matrices  $I_m(\mathbf{r})$  and  $I_c(\mathbf{r})$ . Unfortunately, these two operations imply most of the time a large computational burden in scenarios of practical interest. First, the computation of  $\hat{b}_{\text{ML}}$  implicitly requires a complexity equivalent to running the EM algorithm until convergence. Indeed, if there exists an algorithm able to compute  $\hat{b}_{\text{ML}}$  with a complexity lower than the EM algorithm, considering the EM algorithm has only little interest. Moreover, the evaluation of  $I_c(\mathbf{r})$  and  $I_m(\mathbf{r})$  is also usually a complex task. Indeed, the EM algorithm is often used in situations where Newton-type algorithms (which requires the computation of the Hessian of the incomplete-data likelihood function) is too complex [2]. Now, if the evaluation of  $\frac{\partial^2}{\partial \tilde{b}^2} \log p(\mathbf{r}|\tilde{b})$  is complex, so is the evaluation of  $I_c(\mathbf{r})$  and  $I_m(\mathbf{r})$  since these quantities are related as in (48).

On the other hand, the computation of the MCR (13) usually exhibits a reasonable complexity. Indeed, the evaluation of (13) requires the evaluation of  $\text{CRB}_{\mathbf{z}}$  and  $\text{CRB}_{\mathbf{r}}$ . Now,  $\text{CRB}_{\mathbf{z}}$  has usually an explicit expression very easy to evaluate (see [6] and Section IV for some examples) and its complexity is, therefore,

negligible. Moreover,  $\text{CRB}_{\mathbf{r}}$  can for example be efficiently evaluated as follows [8].

- 1) Generate  $K$  independent realizations of  $\mathbf{r}_i$  according to  $p(\mathbf{r} | b)$ .
- 2) For each vector  $\mathbf{r}_i$ ,  $1 \leq i \leq K$ , evaluate  $\frac{\partial}{\partial \tilde{b}} \log p(\mathbf{r}_i | \tilde{b})$  as follows:

$$\begin{aligned} & \frac{\partial}{\partial \tilde{b}} \log p(\mathbf{r}_i | \tilde{b}) \\ &= \frac{\partial}{\partial \tilde{b}} \log \int_{\mathcal{Z}} p(\mathbf{z}, \mathbf{r}_i | \tilde{b}) d\mathbf{z} \end{aligned} \quad (63)$$

$$= \frac{\int_{\mathcal{Z}} \frac{\partial}{\partial \tilde{b}} p(\mathbf{z}, \mathbf{r}_i | \tilde{b}) d\mathbf{z}}{p(\mathbf{r}_i | \tilde{b})} \quad (64)$$

$$= \int_{\mathcal{Z}} \frac{p(\mathbf{z}, \mathbf{r}_i | \tilde{b})}{p(\mathbf{r}_i | \tilde{b})} \frac{\partial}{\partial \tilde{b}} \log p(\mathbf{z}, \mathbf{r}_i | \tilde{b}) d\mathbf{z} \quad (65)$$

$$= \int_{\mathcal{Z}} p(\mathbf{z} | \mathbf{r}_i, \tilde{b}) \frac{\partial}{\partial \tilde{b}} \log p(\mathbf{z}, \mathbf{r}_i | \tilde{b}) d\mathbf{z} \quad (66)$$

where we have used the fact that

$$\frac{\partial}{\partial \tilde{b}} \log f(\tilde{b}) = \frac{\frac{\partial}{\partial \tilde{b}} f(\tilde{b})}{f(\tilde{b})}$$

in (64) and (65).

- 3) Compute  $\text{CRB}_{\mathbf{r}}(b)$  as follows:

$$\text{CRB}_{\mathbf{r}}(b) \simeq \frac{1}{K} \sum_{i=1}^K \left( \frac{\partial}{\partial \tilde{b}} \log p(\mathbf{r}_i | \tilde{b}) \Big|_{\tilde{b}=b} \right)^2. \quad (67)$$

Comparing (66) with (2)–(3), we can notice that the evaluation of (66) has more or less the same complexity as *one* EM-algorithm iteration (indeed, maximizing a function is in general roughly as complex as computing its gradient). The number of realizations  $K$  depends on the accuracy we want to achieve on the speed of convergence of the EM algorithm. In our simulations (see Section IV), we have noticed that if  $\text{CRB}_{\mathbf{r}}$  is not very close to  $\text{CRB}_{\mathbf{z}}$ , we can already achieve a good accuracy for very small values of  $K$  (around 10–20). On the contrary, when  $\text{CRB}_{\mathbf{r}}$  is close to  $\text{CRB}_{\mathbf{z}}$ , one needs to increase  $K$  to achieve the required accuracy.

Let us conclude this section by mentioning that in a number of situations, the proposed expression (13) enables to have a good insight into the EM-algorithm without making any computation. Indeed, first it is worth noticing that the behavior of the CRBs in many estimation problems has already been studied and is available in the literature. We can therefore take benefit from these results to predict the EM-algorithm convergence via (13). Moreover, as it will become clear from our examples in Section IV, the behavior of the CRB is often predictable by some intuitive reasoning. In such cases, it is therefore easy from (13) to have a *qualitative* idea of the evolution of the EM-algorithm speed of convergence when some parameters of the problem at hand are modified.

#### IV. EXAMPLES

In this section, we illustrate by simulation that the proposed expression enables to properly predict the convergence of the EM algorithm when the size of the observation set is large

enough. In particular, we will consider the estimation of the mean in a Gaussian mixture scenario and the problem of estimating the carrier phase in a digital communication receiver.

##### A. Gaussian Mixture

As a first example, we consider the ML estimation of a parameter  $b$  in the following Gaussian mixture problem:

$$p(\mathbf{r} | b) = \prod_i p(r_i | a_i = 1, b) p(a_i = 1) + p(r_i | a_i = 2, b) p(a_i = 2) \quad (68)$$

where

$$p(r_i | a_i = 1, b) = \frac{1}{\sqrt{2\pi\sigma_1}} \exp \left\{ -\frac{(r_i + b)^2}{2\sigma_1^2} \right\} \quad (69)$$

$$p(r_i | a_i = 2, b) = \frac{1}{\sqrt{2\pi\sigma_2}} \exp \left\{ -\frac{(r_i - b)^2}{2\sigma_2^2} \right\}. \quad (70)$$

In words, the observation  $r_i$  is generated by first selecting either Gaussian distribution (69) or (70) with probability  $p(a_i = 1)$  and  $p(a_i = 2)$ , respectively, and then drawing the observation according to  $p(r_i | a_i, b)$ . Equation (68)–(70), therefore, defines a particular Gaussian mixture problem where the two Gaussian distributions have, up to a sign, the same mean  $b$  but can have different variances. In the sequel we will assume that the distribution  $p(a_i)$  is the same  $\forall i$  and equal to  $p(a)$ .

Let us consider the ML estimation of  $b$  by means of the EM algorithm. Defining the complete data set  $\mathbf{z} \triangleq [\mathbf{r}^T, \mathbf{a}^T]^T$ , where  $\mathbf{a}$  is the vector made up of all the  $a_i$ , and using the standard EM (2)–(3) we easily get the following update equation:

$$\hat{b}^{(n+1)} = \frac{\sum_i r_i \alpha_i^{(n)}}{\sum_i \beta_i^{(n)}} \quad (71)$$

where

$$\alpha_i^{(n)} = \frac{p(r_i | a_i = 2, \hat{b}^{(n)})}{\sigma_2^2} - \frac{p(r_i | a_i = 1, \hat{b}^{(n)})}{\sigma_1^2} \quad (72)$$

$$\beta_i^{(n)} = \frac{p(r_i | a_i = 2, \hat{b}^{(n)})}{\sigma_2^2} + \frac{p(r_i | a_i = 1, \hat{b}^{(n)})}{\sigma_1^2}. \quad (73)$$

According to our previous derivations, the EM-algorithm behavior should be asymptotically well characterized by means of the CRB and MCRB associated to the estimation problem [see (58)]. The CRB and MCRB can be computed as follows. On the one hand, using (10) one readily obtains that

$$\text{MCRB} = \frac{1}{N} \left( \frac{p(a=1)}{\sigma_1^2} + \frac{p(a=2)}{\sigma_2^2} \right)^{-1}. \quad (74)$$

In particular, if  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , we have that

$$\text{MCRB} = \frac{\sigma^2}{N} \quad (75)$$

i.e., the MCRB is equal to the CRB associated to the estimation of the mean of *one* single Gaussian of variance  $\sigma^2$  [7]. On the other hand, the CRB can be computed following the procedure described in Section III-C. The CRB and the MCRB are

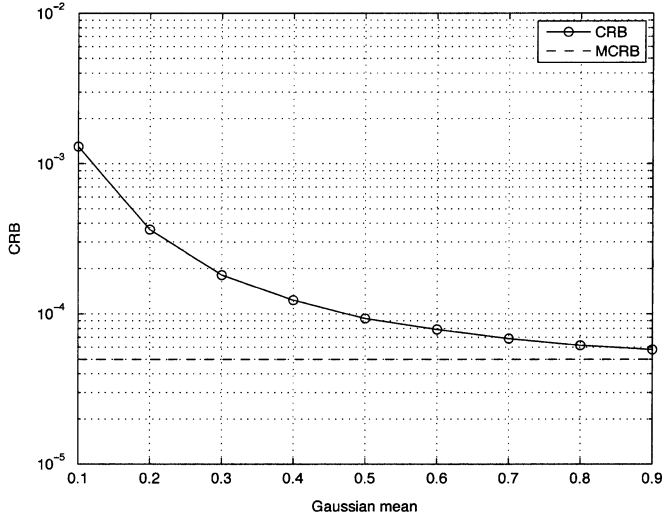


Fig. 1. Standard and modified CRBs versus the absolute mean of the Gaussians of the mixture.

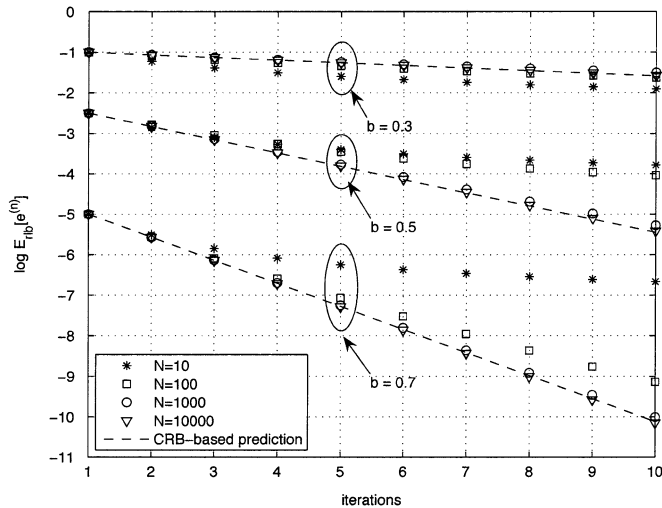


Fig. 2. Evolution of the mean absolute distance between the EM and the ML estimates for different number of observations.

represented in Fig. 1 for  $\sigma_1^2 = \sigma_2^2 = 0.5$ ,  $0.1 \leq b \leq 0.9$ , and  $p(a = 1) = p(a = 2) = 1/2$ .

We have represented<sup>4</sup> in Fig. 2 the evolution of  $\log E_{\mathbf{r}|b}[e^{(n)}(\mathbf{r})]$  versus the number of iterations. The dashed lines represent the performance predicted via (12) and (58) whereas the triangles, circles, squares, and stars represent the actual EM-algorithm performance<sup>5</sup> for different number of observations. We can notice that the simulation results are in good accordance with the proposed prediction when the number of observations increases. We can also notice that, as expected from the behavior of the CRB and MCRB in Fig. 1, the mean speed of convergence of the EM algorithm increases when the value of  $b$  increases.

<sup>4</sup>for the sake of clarity, the curves corresponding to different values of  $b$  have been vertically shifted.

<sup>5</sup>These curves have been computed by ‘‘Monte Carlo simulations,’’ i.e., by averaging  $e^{(n)}(\mathbf{r})$  for a large number of independent realizations of  $\mathbf{r}$  drawn according to  $p(\mathbf{r}|b)$ .

Note that the behavior of the CRB can often be easily predicted by a little thought. For example, in the considered scenario, we can easily predict that the CRB tends to the MCRB when  $b$  is sufficiently large. Indeed, on the one hand, if  $b$  increases, the overlapping between the two Gaussians decreases. On the other hand, the MCRB [see (75)] actually corresponds to the CRB associated to the estimation of the mean of one single Gaussian. Now, it is intuitively clear that we can achieve the same degree of accuracy when estimating the mean of one Gaussian or the mean of two nonoverlapping Gaussians having opposite means. From this simple reasoning and using the proposed CRB-based expression, it is therefore easy to have a *qualitative* idea of the EM-algorithm behavior in the limit of large  $N$ . For example, in the considered case, we can predict that the mean speed of convergence will increase when  $b$  increases.

In the next section, we will give more examples in which we can intuitively predict how the EM-algorithm speed of convergence will evolve when some parameters of the problem at hand are modified.

### B. Carrier Phase Estimation

In this section, we consider the practical example of the carrier-phase synchronization in a digital receiver. The observation model may be expressed as follows:

$$\mathbf{r} = \sqrt{\frac{E_s}{N_0}} \mathbf{a} e^{jb} + \mathbf{v} \quad (76)$$

where  $\mathbf{a}$  is a vector of data symbols belonging to constellation alphabet  $\mathcal{A}$ ,  $E_s$  is the mean energy per symbol,  $N_0$  is the noise spectral density,  $b$  is the carrier-phase offset, and  $\mathbf{v}$  is a vector of zero-mean white Gaussian noise samples with complex variance equal to 1.

The EM algorithm is applied to the problem of computing the ML estimate of the carrier phase offset. Defining the complete data set as  $\mathbf{z} \triangleq [\mathbf{r}^T, \mathbf{a}^T]^T$ , we get the following EM update equation [11]:

$$\hat{b}^{(n+1)} = \arg \left\{ \sum_{i=1}^N \eta_i^{(n)} r_i \right\} \quad (77)$$

where

$$\eta_i^{(n)} = \sum_{a \in \mathcal{A}} a p(a_k = a | \mathbf{r}, \hat{b}^{(n)}) \quad (78)$$

and  $\arg\{x\}$  denotes the argument of the complex number  $x$ . Note that the definition of the complete data set is similar to the one made in Section III-B. Hence, in the remainder of this section, we will compare the EM-algorithm speed of convergence to the one predicted using (58).

The MCR expression in (58) is only a function of the CRB/MCRB-ratio. On the other hand, note that the behavior of the CRBs and MCRB associated to carrier-phase estimation has been extensively studied in the literature, see, e.g., [12], [8], [6], and [9]. In [12], the authors derive the CRB expression for uncoded transmissions, i.e., assuming that all transmitted symbols are equiprobable. In [8], a semianalytical method is proposed



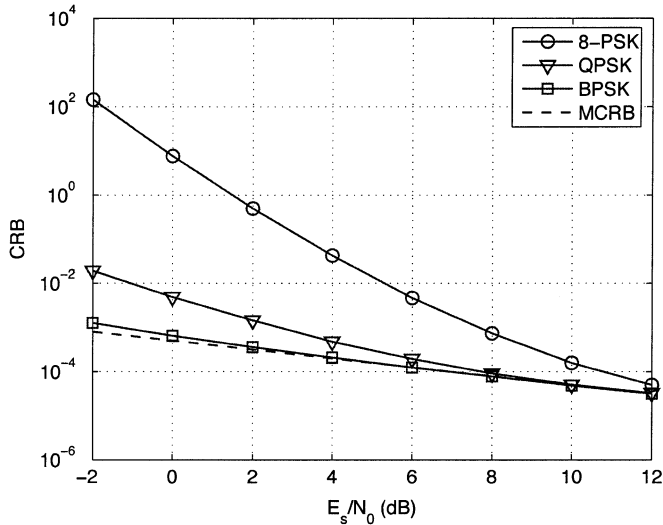


Fig. 3. CRBs versus  $E_s/N_0$ -ratio for uncoded PSK transmission with different constellation sizes.

for the evaluation of the CRB in coded scenarios. In [6], the authors derived the expression of the MCRB

$$\text{MCRB} = \frac{1}{2N} \left( \frac{E_s}{N_0} \right)^{-1}. \quad (79)$$

Finally, in [9], the author shows that the CRB is equal to the MCRB at high signal-to-noise ratios (SNRs).

In the remainder of this section, we will illustrate that (58) may help predicting the EM-algorithm behavior by taking benefit from the knowledge of the CRB and MCRB behavior. We will consider the sensitivity of the EM-algorithm convergence to, respectively, the symbol-constellation size, the SNR and the symbol-sequence *a priori* information  $p(\mathbf{a})$ . In each scenario, the EM-algorithm performance computed via Monte Carlo simulations will be compared to the one predicted by means of (58).

*a) Effect of the Constellation Size:* We first investigate the EM-algorithm behavior when the size of the symbol constellation alphabet varies. We consider the following setup. The transmitted frames consist of 1000 uncoded PSK symbols. The size of the constellation alphabet is set to either 2 (BPSK), 4 (QPSK) or 8 (8-PSK). We use a Gray mapping.

The CRBs and the MCRB<sup>6</sup> associated to this setup are represented versus the  $E_s/N_0$ -ratio in Fig. 3. We see that for a given  $E_s/N_0$ -ratio, the larger the constellation size the worse the achievable estimation quality. This behavior is predictable from some intuitive reasoning. Indeed, it is easy to show that the MCRB defined in (79) is equal to the CRB associated to the carrier phase estimation problem when the symbols  $\mathbf{a}$  are perfectly known at the receiver (i.e., there is no uncertainty about  $\mathbf{a}$ ). On the other hand, for a given  $E_s/N_0$ -ratio, it is clear that the more elements we have in the constellation alphabet  $\mathcal{A}$ , the more uncertainty we have about the transmitted symbols. As a consequence, this implies that increasing the constellation size can only decrease the estimation quality. Based on the behavior of the CRBs and MCRB and using (58), we can therefore expect the EM algorithm to exhibit (asymptotically) a faster mean speed of converge when the size of the constellation decreases.

<sup>6</sup>Note that the MCRB does not depend on the constellation size.

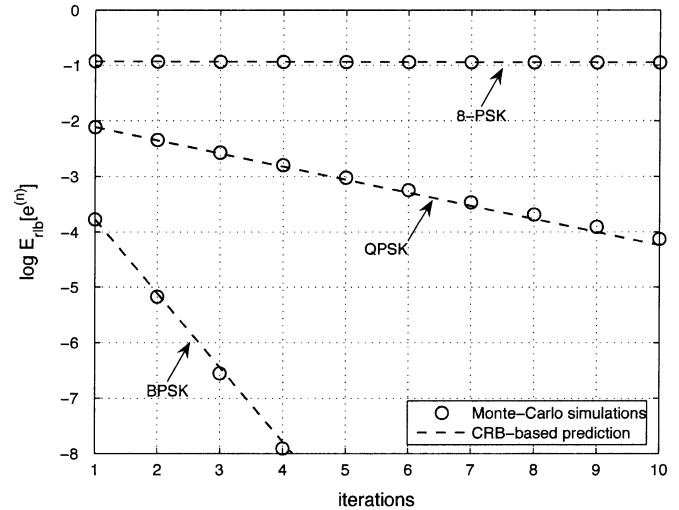


Fig. 4. Absolute mean distance between the ML estimate and the EM-algorithm estimate at a given iteration. Monte Carlo simulations are compared with the performance predicted by (58). The plot illustrates the sensitivity of the EM-algorithm speed of convergence to the choice of the constellation size.

Fig. 4 illustrates the validity of the proposed approach. We compare the EM-algorithm performance as predicted by (12) and (58) with actual performance computed via Monte Carlo simulations. More particularly, we have represented  $\log E_{r|b}[e^{(n)}(\mathbf{r})]$  versus the number of EM iterations. The  $E_s/N_0$ -ratio has been set to 4 dB. The EM algorithm has been initialized by means of a phase estimate computed by a Viterbi and Viterbi [13] synchronizer. The dashed curves correspond to the prediction computed from (12) and (58) whereas the circles correspond to the actual performance computed via Monte Carlo simulations. We see from Fig. 4 that the simulated points are very close to the performance predicted via (58). Therefore, as far as our simulation setup is concerned, the proposed CRB-based expression (58) of the MCR enables to accurately predict the EM algorithm convergence. In particular, we can easily predict from the CRB behavior in Fig. 3 that an increase of the constellation size will decrease the speed of convergence of the EM-based phase synchronizer.

*b) Effect of the SNR:* We now illustrate the sensitivity of the EM algorithm speed of convergence to the system operating SNR. We keep the same setup as in the previous point. The CRB and MCRB plotted in Fig. 3 are therefore still valid for computing the MCR via (58). We see that, as emphasized in [9], the CRBs tend to the MCRB at high SNR irrespective of the constellation size. From our previous derivations we can, therefore, conclude that, as long as  $N$  is large enough, the EM algorithm will exhibit a faster speed of convergence when the SNR increases.

In Fig. 5, we compare the performance predicted via (12) and (58) with the one computed by Monte Carlo simulations. The CRB-based predictions are plotted with dashed curves and the simulated points with circles. The constellation alphabet is a Gray-mapped 8-PSK and we have considered  $E_s/N_0$  equal to 4, 8, and 12 dB, respectively. We see from this figure that the behavior predicted by (12) and (58) is in good accordance with the results computed by Monte Carlo simulations. We also see that, as expected, the EM algorithm converges all the faster as the SNR is large.

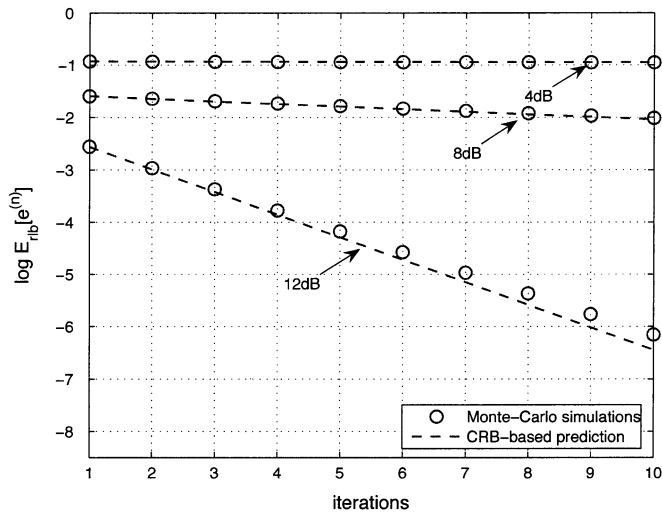


Fig. 5. Mean distance between the ML estimate and the EM-algorithm estimate at a given iteration. Monte Carlo simulations are compared with the performance predicted by (58). The plot illustrates the sensitivity of the EM-algorithm speed of convergence to the operating SNR.

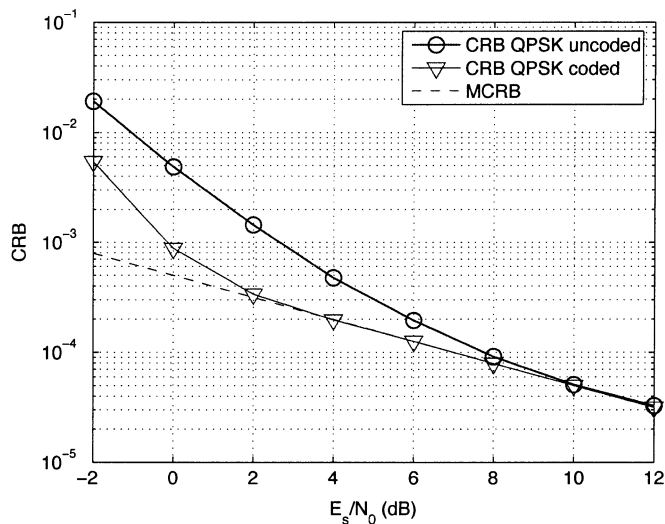


Fig. 6. CRBs versus  $E_s/N_0$ -ratio for QPSK uncoded and coded transmissions.

*c) Effect of Coding:* As a last example of the utility of (12) and (58), we consider the effect of an error-correcting code on the speed of convergence of the EM algorithm. Fig. 6 represents the CRB associated to a (particular) coded and an uncoded transmission, respectively. The transmitted frames consist of 1000 QPSK symbols with Gray mapping. For the coded transmission, we used a rate-1/2 convolutional code with constraint length equal to 3. Since the code structure provides some *a priori* information about the transmitted sequence, “coded” CRB is always lower than the “uncoded” one [8]. According to our previous reasoning, this means that the EM algorithm should exhibit a faster convergence for coded transmissions than for uncoded ones. This is illustrated in Figs. 7 and 8. In Fig. 7 we have represented the MCR predicted by (58) in the coded and uncoded cases. We see that, in good accordance with our intuition, the MCR is always lower for the coded than for the uncoded transmission. In Fig. 8, we compare the MCR approximation

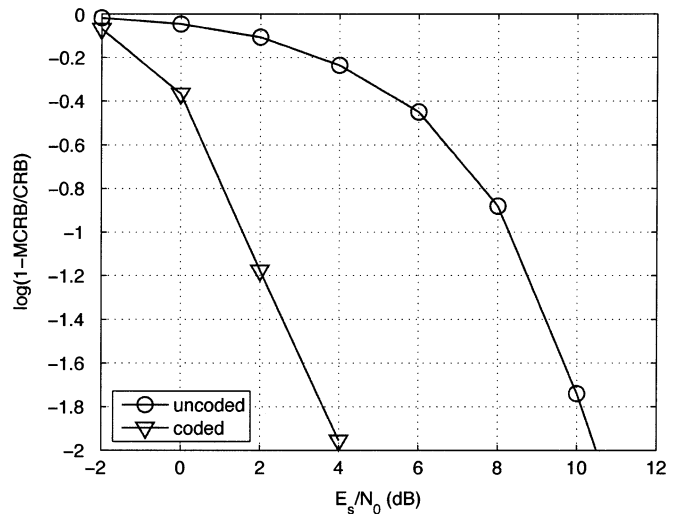


Fig. 7. EM-algorithm mean convergence rate versus  $E_s/N_0$ -ratio for QPSK uncoded and coded transmissions.

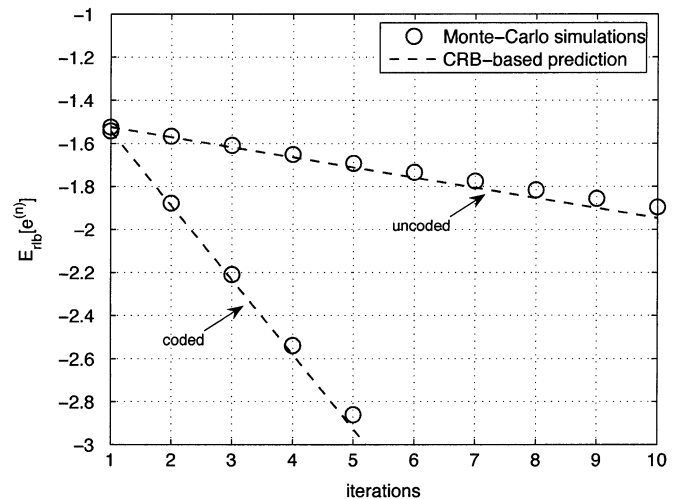


Fig. 8. Mean distance between the ML estimate and the EM-algorithm estimate at a given iteration. Monte Carlo simulations are compared with the performance predicted by (58). The plot illustrates the EM-algorithm speed of convergence for QPSK uncoded and coded transmissions.

given in (58) (dashed lines) to actual Monte Carlo simulation results (circles). We see that the EM speed of convergence is much faster in the coded case than in the uncoded one. We can note again the good accordance between the EM performance predicted via (12) and (58) and the simulation results.

## V. CONCLUSION

In this contribution, we addressed the problem of the characterization of the EM algorithm mean speed of convergence. Based on some building assumptions, we showed that the EM-algorithm *asymptotic* mean speed convergence can be related to the CRBs associated to the incomplete and the complete data sets, respectively. Using the derived expression, we showed that we can gain a good intuition into the EM-algorithm behavior: the further the incomplete-data CRB is from the complete-data CRB, the slower the EM-algorithm speed

of convergence. We then particularized our result to the case where the complete data set is defined as the concatenation of the incomplete data set and a nuisance-parameter vector. In such a case, we emphasized that the complete-data CRB is nothing but the well-known modified CRB (MCRB). Finally, we illustrated our derivations by simulation results. In particular, we showed that the performance predicted by our approach is in good accordance with Monte Carlo simulation results when the number of observations is large.

#### APPENDIX

Here, we show that (56) is valid. The first inequality in (56) directly derives from the definition of the CRB [7], [14]. The second inequality may be proved as follows. Using twice the Bayes rule on  $p(\mathbf{z}, \mathbf{r}|b)$ , we have

$$p(\mathbf{z} | \mathbf{r}, b) p(\mathbf{r} | b) = p(\mathbf{r} | \mathbf{z}, b) p(\mathbf{z} | b). \quad (80)$$

Since  $\mathbf{z}$  univocally defines  $\mathbf{r}$  through the many-to-one mapping  $\mathbf{r} = f(\mathbf{z})$  (see Section II) we have

$$p(\mathbf{z} | \mathbf{r}, b) p(\mathbf{r} | b) = \mathbb{1}\{\mathbf{r} = f(\mathbf{z})\} p(\mathbf{z} | b) \quad (81)$$

where  $\mathbb{1}\{S\}$  is the indicator function which is equal to 1 if  $S$  is true and 0 otherwise. Taking the logarithm and the second derivative of (81), we then get

$$\frac{\partial^2}{\partial b^2} \log p(\mathbf{r} | b) = -\frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | \mathbf{r}, b) + \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | b) \quad (82)$$

since  $\mathbb{1}\{\mathbf{r} = f(\mathbf{z})\}$  does not depend on  $b$ . Taking the expectation with respect to  $p(\mathbf{z} | \mathbf{r}, b)$  of both sides, we finally have

$$\begin{aligned} & \frac{\partial^2}{\partial b^2} \log p(\mathbf{r} | b) \\ &= - \int_{\mathbf{z}} p(\mathbf{z} | \mathbf{r}, b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | \mathbf{r}, b) d\mathbf{z} \\ & \quad + \int_{\mathbf{z}} p(\mathbf{z} | \mathbf{r}, b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | b) d\mathbf{z} \end{aligned} \quad (83)$$

since the LHS does not depend on  $\mathbf{z}$ . Considering the first term in the RHS of (83), it can readily be shown [7] that

$$\begin{aligned} & - \int_{\mathbf{z}} p(\mathbf{z} | \mathbf{r}, b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | \mathbf{r}, b) d\mathbf{z} \\ &= \int_{\mathbf{z}} p(\mathbf{z} | \mathbf{r}, b) \left( \frac{\partial}{\partial b} \log p(\mathbf{z} | \mathbf{r}, b) \right)^2 d\mathbf{z} \geq 0. \end{aligned} \quad (84)$$

Hence

$$\frac{\partial^2}{\partial b^2} \log p(\mathbf{r} | b) \geq \int_{\mathbf{z}} p(\mathbf{z} | \mathbf{r}, b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | b) d\mathbf{z}. \quad (85)$$

Taking the expectation with respect to  $p(\mathbf{r} | b)$  of both sides of (85), we have

$$\begin{aligned} & \int_{\mathcal{R}} p(\mathbf{r} | b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{r} | b) d\mathbf{r} \\ & \geq \int_{\mathcal{Z}} p(\mathbf{z} | b) \frac{\partial^2}{\partial b^2} \log p(\mathbf{z} | b) d\mathbf{z} \end{aligned} \quad (86)$$

where the RHS derives from the fact that

$$\int_{\mathcal{R}} p(\mathbf{z} | \mathbf{r}, b) p(\mathbf{r} | b) d\mathbf{r} = p(\mathbf{z} | b).$$

Taking the negative of (86) and inverting, we finally get (56).

#### REFERENCES

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, Jan. 1977.
- [2] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley Series in Probabil. Statist., 1997.
- [3] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Neural Computat.*, vol. 8, no. 1, pp. 129–151, 1996.
- [4] R. Salakhutdinov, S. Roweis, and Z. Ghahraman, On the Convergence of Bound Optimization Algorithms [Online]. Available: citeseer.ist.psu.edu/584732.html URL
- [5] C. Herzet and L. Vandencloper, "Prediction of the EM-algorithm speed of convergence with Cramer-Rao bounds," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, HI, Apr. 2007.
- [6] A. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1391–1399, Feb./Mar./Apr. 1994.
- [7] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing Communications and Control*. Englewood Cliffs, NJ: Prentice-Hall Signal Processing Series, 1995.
- [8] N. Noels, H. Steendam, and M. Moeneclaey, "The Cramer-Rao bound for phase estimation from coded linearly modulated signals," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 207–209, May 2003.
- [9] M. Moeneclaey, "On the true and the modified Cramer-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1536–1544, Nov. 1998.
- [10] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.
- [11] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "Turbo-synchronization: An EM algorithm approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Anchorage, AK, May 2003, pp. 2933–2937.
- [12] W. G. Cowley, "Phase and frequency estimation for PSK packets: Bounds and algorithms," *IEEE Trans. Commun.*, vol. 44, pp. 26–28, Jan. 1996.
- [13] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, pp. 543–551, Jul. 1983.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.



**Cédric Herzet** was born in Verviers, Belgium, in 1978. He received the electrical engineering and Ph.D. degrees in applied science from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2001 and 2006, respectively.

From August 2001 to April 2006, he was a Research Assistant with the Communications and Remote Sensing Laboratory, UCL. From August 2001 to December 2001, he worked on the topic of VDSL systems in collaboration with Alcatel Bell, Antwerp, Belgium. From January 2002 to April 2006, his research concerned synchronization and parameter estimation in iterative receivers. From May to August 2006, he was a Postdoctoral Researcher with the Ecole Normale Supérieure de Cachan, France. From October 2006 to October 2007, he was a Fulbright Postdoctoral Researcher with the University of California, Berkeley. He is currently a permanent researcher with the TEMICS Lab, IRISA/INRIA, Rennes, France. His research interests include detection, parameter estimation, coding theory, graphical models, and message-passing algorithms.



**Valéry Ramon** was born in Ath, Belgium, in 1978. He received the electrical engineering degree from the Université Catholique de Louvain (UCL), Louvain-la-Neuve, Belgium, in 2001. His graduation thesis concerned multiuser turbo receivers for wideband code division multiple access (WCDMA) transmissions.

Since August 2001, he has been a researcher with the Digital Communications Group (DIGICOM), Communications and Remote Sensing Laboratory, UCL. From August 2001 to December 2001, he worked on the topic of VDSL systems in collaboration with Alcatel Bell, Antwerp, Belgium. From January 2002 to January 2007, his research concerned channel estimation in iterative receivers. He is currently a Senior Researcher with the Wireless Group of the Inter-university Micro-Electronics Center (IMEC), Leuven, Belgium.



**Alexandre Renaux** was raised in a small French mountain town called "les Vosges." He received the Agrégation, the M.Sc., and the Ph.D. degrees in electrical engineering from the Ecole Normale Supérieure de Cachan, France, in 2002, 2003, and 2006, respectively.

From 2006 to 2007, he was a Postdoctoral Research Associate with the Department of Electrical and Systems Engineering, Washington University, St. Louis, MO. He is currently an Assistant Professor with the Department of Physics and a member of the Laboratory of Signals and Systems (L2S), University Paris 11, France. His research interests include detection and parameter estimation theory in statistical signal and array processing.



**Luc Vandendorpe** (F'06) was born in Mouscron, Belgium, in 1962. He received the electrical engineering degree (*summa cum laude*) and the Ph.D. degree from the Université Catholique de Louvain (UCL) Louvain-la-Neuve, Belgium, in 1985 and 1991, respectively.

Since 1985, he has been with the Communications and Remote Sensing Laboratory, UCL where he first worked in the field of bit-rate reduction techniques for video coding. From March 1992 to August 1992, he was a Visiting Scientist and Research Fellow with the Telecommunications and Traffic Control Systems Group of the Delft Technical University, The Netherlands, where he worked on Spread Spectrum Techniques for Personal Communications Systems. From October 1992 to August 1997, he was a Senior Research Associate of the Belgian NSF at UCL, and an Invited Assistant Professor. Presently, he is currently with the Université Catholique de Louvain (UCL), Belgium. He is the Belgian Delegate to COST 273 and 289. He is mainly interested in digital communication systems: equalization, joint detection/synchronization for CDMA, OFDM (multicarrier), MIMO and turbo-based communications systems (UMTS, xDSL, WLAN, etc.), and joint source/channel (de)coding.

Dr. Vandendorpe was a corecipient of the Biennial Alcatel-Bell Award from the Belgian NSF for a contribution in the field of image coding in 1990. In 2000, he was a corecipient (with J. Louveaux and F. Deryck) of the Biennial Siemens Award from the Belgian NSF for a contribution about filter bank-based multicarrier transmission. In 2004, he was a co-winner (with J. Czyz) of the Face Authentication Competition, FAC 2004. He is or has been a TPC member for the IEEE VTC Fall 1999, IEEE Globecom 2003 Communications Theory Symposium, the 2003 Turbo Symposium, IEEE VTC Fall 2003, and IEEE SPAWC 2005. He is a Technical Co-Chair (with P. Duhamel) for IEEE ICASSP 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a member of the Signal Processing Committee for Communications. He was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS FOR SYNCHRONIZATION AND EQUALIZATION between 2000 and 2002, and chair of the IEEE Benelux joint chapter on Communications and Vehicular Technology between 1999 and 2003.