PERFORMANCE LOWER BOUNDS OF BLIND SYSTEM IDENTIFICATION TECHNIQUES IN THE PRESENCE OF CHANNEL ORDER ESTIMATION ERROR

Le Trung Thanh[⋆,†], Karim Abed-Meraim[⋆], Nguyen Linh Trung[†]

[⋆]PRISME Laboratory, University of Orléans, Orléans, France
[†]AVITECH Institute, VNU University of Engineering and Technology, Hanoi, Vietnam

***Abstract*—In this paper, we derive two performance lower bounds for blind system identification in the presence of channel order estimation error. The first bound deals with models where both the channel and unknown symbols are deterministic, and obtained via the constrained misspecified Cramer-Rao bound (MCRB). When transmitted symbols are unknown random variables i.i.d. drawn from a stochastic Gaussian process, variance of any unbiased estimators is always higher than the second MCRB bound. Both proposed MCRB bounds reduce to the classical Cramer-Rao bounds when the channel order is known or accurately estimated. Besides, the stochastic MCRB is lower than the deterministic bound, especially at high SNRs.**

***Index Terms*—Peformance lower bounds, Constrained Cramer-Rao bound, Misspecification, MIMO, Channel order.**

I. INTRODUCTION

Channel estimation is one of the most fundamental problems in wireless communications. Many efficient methods have been proposed for channel estimation so far. One can categorize them into two main classes: data-aided and blind estimation [1]. Data-aided estimators exploit pilot symbols at the receiver to estimate the channel. Blind estimators, on the other hand, can identify the channel parameters directly from output observations without the need for pilots. Accordingly, blind channel estimation is a promising candidate to solve the pilot overhead and increase the spectral efficiency of communication systems. However, their accuracy is often less than that of data-aided estimators, in which case one might rely on semi-blind approaches [1].

For many channel estimators, accurate estimation of the channel order is of great importance and determines the performance of the channel estimation. There are several methods designed for channel order estimation in the literature such as [2]–[5]. However, these methods only work well under certain conditions and do not always result in the true order in practice. This work focuses on evaluating performance lower bounds for channel estimators in the presence of channel order estimation error.

It is well known that the Cramer-Rao bound (CRB) provides a lower bound on the variance of any unbiased estimators and is often used as a benchmark for parameter estimators [6]. Many studies have been conducted to derive analytical expressions of the CRB for channel estimators in general and for blind estimators in particular, for examples [7]–[13]. These CRB bounds, however, are appropriate only for perfect specification models, i.e., the true channel order is either known in advance or accurately estimated. This limitation motivates us to look for new performance bounds dealing with the imperfect knowledge of channel order information.

Contribution: We propose to use the misspecified CRB (MCRB), which is an extended version of the CRB for misspecfication models [14]–[16], in order to analyze the theoretical performance limit of blind estimators when the channel order is misspecified. In particular, a new interpretation of MCRB via the Moore–Penrose inverse, called generalized MCRB (GMCRB), is proposed to deal with the inherent ambiguity in blind identification. The proposed GMCRB is the tightest constrained MCRB among all choices of parametric constraints to regularize the singular problem. The GMCRB is not only identical to the usual MCRB for regular problems but also consistent with the classical CRB under well-specified models. Two closed-form expressions of the GMCRB are then derived for unbiased channel estimators when unknown transmitted symbols are (i) deterministic ($\mathrm{GMCRB}^{\mathtt{Det}}$) and (ii) stochastic ($\mathrm{GMCRB}^{\mathtt{Stoch}}$). The two proposed GMCRBs, for the first time, provide performance lower bounds for blind channel estimation techniques under model order misspecification.

II. SYSTEM MODEL

We consider a convolutive MIMO system with N_t transmit antennas and N_r receive antennas whose individual channels are modeled as finite impulse responses (FIR). The output vector $\boldsymbol{y}[t] = \big[y_1[t], y_2[t], \dots, y_{N_r}[t]\big]^\top \in \mathbb{C}^{N_r \times 1}$ received at N_r receive antennas is formulated by

$$\boldsymbol{y}[t] = \sum_{i=0}^{L-1} \boldsymbol{H}[i]\boldsymbol{x}[t-i] + \boldsymbol{n}[t], \quad t = 0, 1, \dots, N-1. \tag{1}$$

where $\boldsymbol{H}$ is the overall $N_r \times N_t$ MIMO channel of order $L-1$, $\{\boldsymbol{x}[t] \in \mathbb{C}^{N_t \times 1}\}_{t \in \mathcal{Z}}$ represent the transmitted symbols, and $\boldsymbol{n}[t]$ is an $N_r \times 1$ additive noise vector drawn from an i.i.d. circular complex Guassian distribution $\mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}_{N_r})$, $\mathbb{E}\{\boldsymbol{n}[t]\boldsymbol{n}[t]^\top\} = \boldsymbol{0}$. For simplicity, it is assumed that a preamble block of zero samples is added to avoid intersymbol interference from two successive blocks, i.e. $\boldsymbol{x}[t] = \boldsymbol{0}$ for $t < 0$. One often stacks the N output samples $\{\boldsymbol{y}[t]\}_{t=0}^{N-1}$ into a single vector $\boldsymbol{y} \in \mathbb{C}^{NN_r \times 1}$ as

$$\boldsymbol{y} = \big[\boldsymbol{y}[0]^\top, \boldsymbol{y}[1]^\top, \dots, \boldsymbol{y}[N-1]^\top\big]^\top. \tag{2}$$

Accordingly, we can recast the convolution (1) into the following standard expression

$$\boldsymbol{y} = \mathcal{T}(\boldsymbol{h})\boldsymbol{x} + \boldsymbol{n}, \tag{3}$$

where the input vector $\boldsymbol{x}$ and the noise vector $\boldsymbol{n}$ are given by

$$\boldsymbol{x} = \big[\boldsymbol{x}[-L+1]^\top, \dots, \boldsymbol{x}[0]^\top, \dots, \boldsymbol{x}[N-1]\big]^\top, \tag{4}$$

$$\boldsymbol{n} = \big[\boldsymbol{n}[0]^\top, \boldsymbol{n}[1]^\top, \dots, \boldsymbol{n}[N-1]^\top\big]^\top, \tag{5}$$

and $\mathcal{T}(\boldsymbol{h})$ is a $N_rN \times N_t(L+N-1)$ block matrix representing the convolution:

$$\mathcal{T}(\boldsymbol{h}) = \left[\begin{array}{ccc|cccc} \boldsymbol{H}[L-1] & \dots & \boldsymbol{H}[1] & \boldsymbol{H}[0] & & & \\ & \ddots & \dots & \boldsymbol{H}[1] & \boldsymbol{H}[0] & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \boldsymbol{H}[L-1] & \dots & \boldsymbol{H}[0] \end{array}\right]. \quad (6)$$

$\mathcal{T}(\boldsymbol{h})$ is the block Toeplitz matrix depending on the channel coefficients $\boldsymbol{h} = \left[\boldsymbol{h}_{L-1}^\top, \dots, \boldsymbol{h}_0^\top\right]^\top$, with $\boldsymbol{h}_i = \text{vec}(\boldsymbol{H}_i)$. For well-definedness [17], [18], we suppose that $\mathcal{T}(\boldsymbol{h})$ is of full column-rank and has more rows than columns, i.e., $N_rN > N_t(L+N)$. For short, we denote $\mathcal{T}(\boldsymbol{h})$ by $\boldsymbol{\mathcal{H}}$.

Thanks to the "vec trick" in [19, Lemma 4.3.1], we can express (3) as a linear operation on the channel coefficient vector $\boldsymbol{h}$:

$$\boldsymbol{y} = \boldsymbol{\mathcal{X}}\boldsymbol{h} + \boldsymbol{n} = \left(\boldsymbol{X}^\top \otimes \boldsymbol{I}_{N_r}\right)\boldsymbol{h} + \boldsymbol{n}, \quad (7)$$

where $\boldsymbol{X} \in \mathbb{C}^{LN_t \times N}$ is the matrix formed by input samples

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{x}[N-L] \\ \vdots & \vdots & \iddots & \vdots \\ \vdots & \mathbf{0} & \iddots & \vdots \\ \mathbf{0} & \boldsymbol{x}[0] & \dots & \boldsymbol{x}[N-2] \\ \boldsymbol{x}[0] & \boldsymbol{x}[1] & \dots & \boldsymbol{x}[N-1] \end{bmatrix}, \quad (8)$$

and operator "$\otimes$" denotes the Kronecker product.

III. Misspecified Cramer-Rao Bound

In this section, we briefly review an extended version of the CRB when dealing with misspecification models, called misspecified CRB (MCRB) [16]. For further information about its derivation, properties, and applications, we refer the reader to [14]–[16] and references therein.

Assume that data samples are i.i.d drawn from the true distribution $f_{\mathbf{y}}$. However, users adopt a different distribution $g(\boldsymbol{y}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}$ to characterize statistics of $\boldsymbol{y}$ instead, where $g(\boldsymbol{y}|\boldsymbol{\theta}) \neq f(\boldsymbol{y}) \ \forall \boldsymbol{\theta}$ is allowed. For the users, the problem of interest is now to estimate $\boldsymbol{\theta}$[1].

A. Pseudo-true Parameter $\boldsymbol{\theta}_{pt}$

In this context, Kullback-Leibler (KL) divergence is used to determine the "best" performance unbiased estimators might attain in the presence of misspecification models [16].

The KL divergence measures the amount of information loss when we use the assumed $g_{\mathbf{y}|\boldsymbol{\theta}}$ to approximate the true $f_{\mathbf{y}}$, which is defined by

$$\text{KL}(f_{\mathbf{y}} \parallel g_{\mathbf{y}|\boldsymbol{\theta}}) \triangleq \mathbb{E}_f\left\{\log\left(\frac{f(\boldsymbol{y})}{g(\boldsymbol{y}|\boldsymbol{\theta})}\right)\right\}. \quad (9)$$

The unique parameter minimizing $\text{KL}(f_{\mathbf{y}} \parallel g_{\mathbf{y}|\boldsymbol{\theta}})$ is so-called the *pseudo-true* parameter, $\boldsymbol{\theta}_{pt}$. In practice, $\text{KL}(f_{\mathbf{y}} \parallel g_{\mathbf{y}|\boldsymbol{\theta}})$ cannot be obtained since the true density $f_{\mathbf{y}}$ is generally unknown, so we can estimate the maximum likelihood (MLE) for the density instead. Particularly, minimizing (9) is equivalent to maximizing the expectation of the misspecified log-likelihood function $\ell(\boldsymbol{y}|\boldsymbol{\theta}) \triangleq \log g(\boldsymbol{y}|\boldsymbol{\theta})$, i.e.,

$$\boldsymbol{\theta}_{pt} \triangleq \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmin}}\, \text{KL}(f_{\mathbf{y}} \parallel g_{\mathbf{y}|\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmax}}\, \mathbb{E}_f\left\{\ell(\boldsymbol{y}|\boldsymbol{\theta})\right\}. \quad (10)$$

Accordingly, it is shown in [14], [21] that the "quasi" MLE $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ converges in probability to $\boldsymbol{\theta}_{pt}$.

In the following, we always assume the existence and the uniqueness of the *pseudo-true* parameter and provide the closed-form expression of $\boldsymbol{\theta}_{pt}$ if possible.

B. Unconstrained MCRB

Let $\hat{\boldsymbol{\theta}}$ be an estimator derived under the misspecified model $g(\boldsymbol{y}|\boldsymbol{\theta})$ from the output samples. We call $\hat{\boldsymbol{\theta}}$ misspecified (MS)-unbiased estimator if and only if

$$\mathbb{E}_f\{\hat{\boldsymbol{\theta}}(\boldsymbol{y})\} = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y}) f(\boldsymbol{y}) d\boldsymbol{y} = \boldsymbol{\theta}_{pt}. \quad (11)$$

We define the two matrices $\boldsymbol{J}_{\boldsymbol{\theta}}$ and $\boldsymbol{A}_{\boldsymbol{\theta}}$ as follows[2]

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \mathbb{E}_f\left\{\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\right)^H\right\}, \quad \boldsymbol{A}_{\boldsymbol{\theta}} = \mathbb{E}_f\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^H}\right\}. \quad (12)$$

When $\boldsymbol{A}_{\boldsymbol{\theta}}$ is non-singular at $\boldsymbol{\theta} = \boldsymbol{\theta}_{pt}$, the total covariance of any MS-unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ is bounded by MCRB [16]

$$\text{VAR}(\hat{\boldsymbol{\theta}}(\boldsymbol{y})) \geq \text{MCRB}(\boldsymbol{\theta}_{pt}) \triangleq \boldsymbol{A}_{\boldsymbol{\theta}_{pt}}^{-1} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}_{\boldsymbol{\theta}_{pt}}^{-1}. \quad (13)$$

C. Constrained MCRB

When additional constraints are imposed on $\boldsymbol{\theta}$, the constrained version of the MCRB, called constrained MCRB (CMCRB), has been recently introduced by Stefano *et al.* in [22], [23].

Suppose that $\boldsymbol{\theta}$ is required to satisfy the following constraint $u(\boldsymbol{\theta}) = \mathbf{0}$. If the Jacobian matrix $\nabla_u(\boldsymbol{\theta}) \triangleq \frac{\partial u(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}$ is of full rank for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and there exists $\boldsymbol{U}$ spanning its null space,

$$\nabla_u(\boldsymbol{\theta})\boldsymbol{U} = \mathbf{0} \ \text{ and } \ \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}, \quad (14)$$

then the following expression holds for the CMCRB

$$\begin{aligned} \text{VAR}(\hat{\boldsymbol{\theta}}(\boldsymbol{y})) \geq\, & \boldsymbol{U}(\boldsymbol{U}^H\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U})^{-1}(\boldsymbol{U}^H\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U}) \times \\ & \times (\boldsymbol{U}^H\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U})^{-1}\boldsymbol{U}^H \triangleq \text{CMCRB}(\boldsymbol{\theta}_{pt}), \end{aligned} \quad (15)$$

under the assumption that $\boldsymbol{U}^H\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U}$ is nonsingular.

IV. Proposed MCRB for Blind Channel Estimation

Blind techniques consider the estimation of channel parameters from only the channel outputs. Due to the inherent matrix ambiguity, the matrix $\boldsymbol{A}_{\boldsymbol{\theta}}$ is singular, the usual MCRB may not exist and its properties cannot be applied directly. In this section, we propose a new interpretation of the MCRB, called generalized MCRB (GMCRB), which is able to deal with blind channel estimation in particular and singular problems

[1]When dealing with a mixture of both real parameters ($\boldsymbol{\theta}_r$) and complex parameters ($\boldsymbol{\theta}_c$), we consider the following augmented representation: $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_c^\top, \boldsymbol{\theta}_c^H, \boldsymbol{\theta}_r^\top\right]^\top$. The real (dual) representation $\underline{\boldsymbol{\theta}} = [\text{Re}(\boldsymbol{\theta}_c)^\top, \text{Im}(\boldsymbol{\theta}_c)^\top, \boldsymbol{\theta}_r^\top]^\top$ is another option, but there always exists a linear and invertible map $\mathcal{L}$ such that $\underline{\boldsymbol{\theta}} \mapsto \boldsymbol{\theta} = \mathcal{L}(\bar{\boldsymbol{\theta}}) = \boldsymbol{L}\underline{\boldsymbol{\theta}}$, where $\boldsymbol{L}$ is the matrix representation of $\mathcal{L}$. Therefore, the two forms are interchangeable [20].

[2]If $\boldsymbol{\theta} = \left[\theta_1, \theta_2, \dots, \theta_n\right]^\top$, then $\frac{\partial}{\partial \boldsymbol{\theta}} = \left[\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_n}\right]^\top$, $\frac{\partial}{\partial \boldsymbol{\theta}^*} = \left[\frac{\partial}{\partial \theta_1^*}, \frac{\partial}{\partial \theta_2^*}, \dots, \frac{\partial}{\partial \theta_n^*}\right]^\top$, $\frac{\partial}{\partial \boldsymbol{\theta}^\top} = \left[\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2}, \dots, \frac{\partial}{\partial \theta_n}\right]$, and $\frac{\partial}{\partial \boldsymbol{\theta}^H} = \left[\frac{\partial}{\partial \theta_1^*}, \frac{\partial}{\partial \theta_2^*}, \dots, \frac{\partial}{\partial \theta_n^*}\right]$.

in general. Our main result is stated in the following lemma whose detailed proof is omitted here due to the space limit.

Lemma 1. *Let $\hat{\theta}(\boldsymbol{y})$ be an MS-unbiased estimator derived under model misspecification from observed data. The total variance of $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ is lower bounded according to*

$$\mathrm{var}_f\left(\hat{\boldsymbol{\theta}}(\boldsymbol{y})\right) \geq \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \triangleq \mathrm{GMCRB}(\boldsymbol{\theta}_{pt}), \tag{16}$$

where $(\cdot)^{\#}$ denotes the Moore–Penrose inverse operator and the two matrices $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ and $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ are defined as in (12).

Proof Sketch. When $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ is nonsingular, i.e., $\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} = \boldsymbol{A}^{-1}_{\boldsymbol{\theta}_{pt}}$, the proposed GMCRB reduces to the usual MCRB in (13).
When $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ is singular with rank r, there is no MS-unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ with finite variance. In this case, additional constraints should be imposed on $\boldsymbol{\theta}$ to insure its uniqueness. For a given constraint $u(\boldsymbol{\theta})$ with a full-rank $\nabla_u(\boldsymbol{\theta})$, we prove

$$\mathrm{CMCRB}(\boldsymbol{\theta}_{pt}) \geq \mathrm{GMCRB}(\boldsymbol{\theta}_{pt}). \tag{17}$$

We first exploit the fact that the pseudo-inverse matrix of $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ can be expressed as follows

$$\begin{aligned}\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} &= \boldsymbol{U}_r \boldsymbol{\Sigma}_r^{-1} \boldsymbol{U}_r^H \\ &= \boldsymbol{U}_r \big(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\big)^{-1} \boldsymbol{U}_r^H,\end{aligned} \tag{18}$$

where $\boldsymbol{U}_r$ and $\boldsymbol{\Sigma}_r$ are the eigenvector matrix and the matrix of non-zero eigenvalues of $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ respectively. Accordingly, (16) becomes

$$\begin{aligned}\mathrm{GMCRB}(\boldsymbol{\theta}_{pt}) &= \boldsymbol{U}_r \big(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\big)^{-1} \times \\ &\times \big(\boldsymbol{U}_r^H \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\big)\big(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\big)^{-1} \boldsymbol{U}_r^H,\end{aligned} \tag{19}$$

which is identical to the CMCRB as in (15). Next, it is easy to find a constraint $u(\boldsymbol{\theta})$ satisfying $\nabla_u(\boldsymbol{\theta})\boldsymbol{U}_r = \boldsymbol{0}$, thus the proposed GMCRB holds for the CMCRB.
Now, for any orthogonal matrix $\boldsymbol{U}$ such that $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ is nonsingular, we can show that

$$\lambda_i\big[\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}}\big] \leq \lambda_i\big[\boldsymbol{U}\big(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big)^{-1} \boldsymbol{U}^H\big], i = 1, 2, \ldots, r, \tag{20}$$

where $\lambda_i[\boldsymbol{M}]$ is the i-th largest eigenvalue of $\boldsymbol{M}$. In parallel, we show that given three positive-semidefinite Hermitian matrices of the same rank $\boldsymbol{M}, \boldsymbol{N}$, and $\boldsymbol{X}$, if $\boldsymbol{M} \geq \boldsymbol{N}$ then $\boldsymbol{M}\boldsymbol{X}\boldsymbol{M} \geq \boldsymbol{N}\boldsymbol{X}\boldsymbol{N}$. Accordingly, we can conclude that

$$\begin{aligned}\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} &\leq \boldsymbol{U}\big(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big)^{-1} \boldsymbol{U}^H \times \\ &\times \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big)^{-1} \boldsymbol{U}^H = \mathrm{CMCRB}(\boldsymbol{\theta}_{pt}).\end{aligned} \tag{21}$$

It ends the proof. □

In the following, we propose to use the generalized interpretation (16) of the MCRB to determine the performance limit of unbiased blind estimators. Particularly, we focus on the deterministic model ($\mathrm{GMCRB}^{\mathtt{Det}}$) and the stochastic model ($\mathrm{GMCRB}^{\mathtt{Stoch}}$). In the former case, the unknown transmitted symbols are assumed to be deterministic, whereas in the latter case, we assume that they are unknown random variables i.i.d. drawn from a Gaussian distribution.

A. *Deterministic* $\mathrm{GMCRB}^{\mathtt{Det}}$

In this model, the output vector $\boldsymbol{y}$ is drawn from the Gaussian distribution $\mathcal{CN}\big(\boldsymbol{\mathcal{H}}\boldsymbol{x}, \sigma_n^2 \boldsymbol{I}_{N_r N}\big)$[3] and the vector of unknown parameters is $\boldsymbol{\phi} = \big[\boldsymbol{h}^\top, \boldsymbol{x}^\top, \boldsymbol{h}^H, \boldsymbol{x}^H, \sigma_n^2\big]^\top$.
In the presence of channel order estimation error (i.e., $\tilde{L} \neq L$), the users will fit the assumed $g(\boldsymbol{y}|\boldsymbol{\theta})$ to $\boldsymbol{y}$

$$g(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{(\pi\sigma^2)^{N_r N}} \exp\bigg(-\frac{1}{\sigma^2}\big\|\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{x}}\big\|_2^2\bigg), \tag{22}$$

where $\tilde{\boldsymbol{x}} = \big[\boldsymbol{0}^\top_{N_t(\tilde{L}-1)}, \ \boldsymbol{x}[0]^\top, \boldsymbol{x}[1]^\top, \ldots, \boldsymbol{x}[N-1]^\top\big]^\top \in \mathbb{C}^{N_t(\tilde{L}+N-1)\times 1}$ and $\tilde{\boldsymbol{\mathcal{H}}}$ is assumed to be formed by

$$\tilde{\boldsymbol{\mathcal{H}}} = \big[\tilde{\boldsymbol{\mathcal{H}}}_1 | \tilde{\boldsymbol{\mathcal{H}}}_2\big] = \begin{bmatrix} \boldsymbol{H}[\tilde{L}-1] & \ldots & \boldsymbol{H}[0] & & \\ & \ddots & & \ddots & \\ & & \boldsymbol{H}[\tilde{L}-1] & \ldots & \boldsymbol{H}[0] \end{bmatrix}, \tag{23}$$

where the sub-matrix $\tilde{\boldsymbol{\mathcal{H}}}_1$ contains the first $N_t(\tilde{L}-1)$ columns of $\tilde{\boldsymbol{\mathcal{H}}}$, $\tilde{\boldsymbol{\mathcal{H}}}_2$ contains the remaining columns. Instead of $\boldsymbol{\phi}$, the parameter of interest now becomes

$$\boldsymbol{\theta} = \big[\tilde{\boldsymbol{h}}^\top, \boldsymbol{x}^\top, \tilde{\boldsymbol{h}}^H, \boldsymbol{x}^H, \sigma^2\big]^\top. \tag{24}$$

The misspecified log-likehood function is given by

$$\ell(\boldsymbol{y}|\boldsymbol{\theta}) = \mathrm{const} - N_r N \log(\sigma^2) - \frac{1}{\sigma^2}\big\|\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{x}}\big\|_2^2. \tag{25}$$

The pseudo-true parameter $\boldsymbol{\theta}_{pt}$ is derived from minimizing $\mathrm{KL}(f_{\mathbf{y}} \| g_{\mathbf{y}|\boldsymbol{\theta}})$ or maximizing the expectation of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ over the true distribution $f_{\mathbf{y}}$, i.e.,

$$\boldsymbol{\theta}_{pt} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathrm{argmin}} \left\{ N_r N \log(\sigma^2) + \frac{\sigma_n^2 N_r N + \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{x}}\|_2^2}{\sigma^2} \right\}, \tag{26}$$

where $\boldsymbol{\mu} \triangleq \mathbb{E}_f\{\boldsymbol{y}\} = \boldsymbol{\mathcal{H}}\boldsymbol{x}$ is the true mean of $\boldsymbol{y}$. The minimizer of (26) can be obtained directly by applying the maximum likelihood estimation (MLE) or elegant methods reviewed in [17], e.g.

$$\tilde{\boldsymbol{\mathcal{H}}}_{pt} = \underset{\tilde{\boldsymbol{\mathcal{H}}}}{\mathrm{argmin}} \big\|\big(\boldsymbol{I} - \boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}}\big)\boldsymbol{\mu}\big\|_2^2, \tag{27}$$

$$\boldsymbol{x}_{pt} = \big(\tilde{\boldsymbol{\mathcal{H}}}^H_{2,pt} \tilde{\boldsymbol{\mathcal{H}}}_{pt}\big)^{\#} \tilde{\boldsymbol{\mathcal{H}}}^H_{2,pt} \boldsymbol{\mu}, \tag{28}$$

$$\sigma^2_{pt} = \sigma_n^2 + \frac{\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mathcal{H}}}_{pt}\tilde{\boldsymbol{x}}_{pt}\|_2^2}{N_r N}, \tag{29}$$

where $\boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}} \triangleq \tilde{\boldsymbol{\mathcal{H}}}(\tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}})^{-1}\tilde{\boldsymbol{\mathcal{H}}}$.
The first partial derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ is given by

$$\frac{\partial \ell}{\partial \tilde{\boldsymbol{h}}^*} = \frac{1}{\sigma^2}\tilde{\boldsymbol{\mathcal{X}}}^H\big(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}}\big), \qquad \frac{\partial \ell}{\partial \tilde{\boldsymbol{h}}} = \frac{1}{\sigma^2}\tilde{\boldsymbol{\mathcal{X}}}^\top\big(\boldsymbol{y}^* - \tilde{\boldsymbol{\mathcal{X}}}^*\tilde{\boldsymbol{h}}^*\big),$$
$$\frac{\partial \ell}{\partial \boldsymbol{x}^*} = \frac{1}{\sigma^2}\tilde{\boldsymbol{\mathcal{H}}}_2^H\big(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}}\big), \qquad \frac{\partial \ell}{\partial \boldsymbol{x}} = \frac{1}{\sigma^2}\tilde{\boldsymbol{\mathcal{H}}}_2^\top\big(\boldsymbol{y}^* - \tilde{\boldsymbol{\mathcal{X}}}^*\tilde{\boldsymbol{h}}^*\big),$$
$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{\sigma^4}\big\|\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}}\big\|_2^2 - \frac{N_r N_p}{\sigma^2}.$$

Let us denote $\boldsymbol{e}$ the error mean, i.e., $\boldsymbol{e} = \boldsymbol{\mu} - \tilde{\boldsymbol{\mathcal{X}}}_{pt}\tilde{\boldsymbol{h}}_{pt}$. Accordingly, we have

$$\mathbb{E}_f\big\{(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}})(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}})^H\big\} = \sigma^2 \boldsymbol{I}_{N_r N} + \boldsymbol{e}\boldsymbol{e}^H,$$
$$\mathbb{E}_f\big\{(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}})(\boldsymbol{y} - \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}})^\top\big\} = \boldsymbol{e}\boldsymbol{e}^\top.$$

[3]Due to the commutativity of convolution, i.e., $\boldsymbol{\mathcal{H}}\boldsymbol{x} = \boldsymbol{\mathcal{X}}\boldsymbol{h}$, we can state $\boldsymbol{y} \sim \mathcal{CN}\big(\boldsymbol{\mathcal{H}}\boldsymbol{x}, \sigma_n^2 \boldsymbol{I}_{N_r N}\big)$ or $\boldsymbol{y} \sim \mathcal{CN}\big(\boldsymbol{\mathcal{X}}\boldsymbol{h}, \sigma_n^2 \boldsymbol{I}_{N_r N}\big)$.

From (**??**), $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ is derived from $\mathbb{E}_f\left\{\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\right)^H\right\}$ at $\boldsymbol{\theta}=\boldsymbol{\theta}_{pt}$

$$\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}=\frac{1}{\sigma_{pt}^4}\begin{bmatrix}\boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}^*} & \boldsymbol{0}\\ \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}^*} & \boldsymbol{0}\\ \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}^*} & \boldsymbol{0}\\ \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}^*} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & N_rN\end{bmatrix}, \quad (30)$$

where

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}} &= (\boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}^*})^* &&= \tilde{\boldsymbol{\mathcal{X}}}^H(\sigma_n^2\boldsymbol{I}+\boldsymbol{e}\boldsymbol{e}^H)\tilde{\boldsymbol{\mathcal{X}}},\\
\boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}} &= (\boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}})^H &&= \tilde{\boldsymbol{\mathcal{X}}}^H(\sigma_n^2\boldsymbol{I}+\boldsymbol{e}\boldsymbol{e}^H)\tilde{\boldsymbol{\mathcal{H}}}_2,\\
\boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}^*} &= (\boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}})^H &&= \tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{e}\boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{X}}}^*,\\
\boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}^*} &= (\boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}})^H &&= \tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{e}\boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{H}}}_2^*,\\
\boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}} &= (\boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}^*})^H &&= \tilde{\boldsymbol{\mathcal{X}}}^\top\boldsymbol{e}^*\boldsymbol{e}^H\tilde{\boldsymbol{\mathcal{H}}}_2,\\
\boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}^*} &= (\boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}^*})^H &&= \tilde{\boldsymbol{\mathcal{X}}}^\top(\sigma_n^2\boldsymbol{I}+\boldsymbol{e}^*\boldsymbol{e}^\top)\tilde{\boldsymbol{\mathcal{H}}}_2^*,\\
\boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}} &= (\boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}^*})^* &&= \tilde{\boldsymbol{\mathcal{H}}}_2^H(\sigma_n^2\boldsymbol{I}+\boldsymbol{e}\boldsymbol{e}^H)\tilde{\boldsymbol{\mathcal{H}}}_2,\\
\boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}^*} &= (\boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}})^H &&= \tilde{\boldsymbol{\mathcal{H}}}_2^H\boldsymbol{e}\boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{H}}}_2^*.
\end{aligned}$$

Taking the expectation of the second partial derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ over $f_{\mathbf{y}}$ yields $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ as

$$\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}=\frac{-1}{\sigma_{pt}^2}\begin{bmatrix}\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}} & \tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{H}}}_2 & \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{e}\\ \tilde{\boldsymbol{\mathcal{H}}}_2^H\tilde{\boldsymbol{\mathcal{X}}} & \tilde{\boldsymbol{\mathcal{H}}}_2^H\tilde{\boldsymbol{\mathcal{H}}}_2 & \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{H}}}_2^H\boldsymbol{e}\\ \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{X}}}^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \tilde{\boldsymbol{\mathcal{X}}}^\top\tilde{\boldsymbol{\mathcal{H}}}_2^* & \tilde{\boldsymbol{\mathcal{X}}}^\top\boldsymbol{e}^*\\ \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{H}}}_2^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \tilde{\boldsymbol{\mathcal{H}}}_2^\top\tilde{\boldsymbol{\mathcal{H}}}_2^* & \tilde{\boldsymbol{\mathcal{H}}}_2^\top\boldsymbol{e}^*\\ \boldsymbol{e}^H\tilde{\boldsymbol{\mathcal{X}}} & \boldsymbol{e}^H\tilde{\boldsymbol{\mathcal{H}}}_2 & \boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{H}}}_2^* & \frac{N_rN}{\sigma_{pt}^2}\end{bmatrix}. \quad (31)$$

B. *Stochastic* GMCRB^{Stoch}

Suppose that the unknown transmitted symbols are circular Gaussian random variables i.i.d. drawn from $\mathcal{CN}(\boldsymbol{0},\sigma_x^2\boldsymbol{I}_{N_t})$[4]. Accordingly, the vector of true unknown parameters is $\boldsymbol{\phi}=[\boldsymbol{h}^\top,\boldsymbol{h}^H,\sigma_x^2,\sigma_n^2]^\top$ and the received signal $\boldsymbol{y}$ is a circular Gaussian variable with zero-mean and covariance $\boldsymbol{C}$ which is given by $\boldsymbol{C}=\sigma_x^2\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H+\sigma_n^2\boldsymbol{I}_{N_rN}$.

Due to the imperfect knowledge of L, the following distribution function $g_{\mathbf{y}|\boldsymbol{\theta}}$ is used instead

$$g(\boldsymbol{y}|\boldsymbol{\theta})=\frac{1}{\pi^{N_rN}\det(\boldsymbol{R})}\exp\big(-\boldsymbol{y}^H\boldsymbol{R}^{-1}\boldsymbol{y}\big), \quad (32)$$

with the misspecified covariance $\boldsymbol{R}=\sigma_x^2\tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{\mathcal{H}}}^H+\sigma^2\boldsymbol{I}_{N_rN}$, while the zero-mean $\boldsymbol{\mu}$ is correctly specified, thanks to $\mathbb{E}_f\{\boldsymbol{x}\}=\boldsymbol{0}$. The parameter of interest now becomes

$$\boldsymbol{\theta}=\big[\tilde{\boldsymbol{h}}^\top,\tilde{\boldsymbol{h}}^H,\sigma_x^2,\sigma^2\big]^\top. \quad (33)$$

We obtain that $\mathrm{KL}(f_{\mathbf{y}}\|g_{\mathbf{y}|\boldsymbol{\theta}})$ has the closed-form expression

$$\mathrm{KL}(f_{\mathbf{y}}\|g_{\mathbf{y}|\boldsymbol{\theta}})=\log\big(\det\big(\boldsymbol{R}\boldsymbol{C}^{-1}\big)\big)+\mathrm{tr}\left\{\boldsymbol{R}^{-1}\boldsymbol{C}\right\}-1. \quad (34)$$

Unfortunately, minimization (34) w.r.t. $\boldsymbol{\theta}$ is not an easy problem to solve. As mentioned in Section III-A, we can apply the MLE to estimate the covariance $\boldsymbol{R}$ and $\boldsymbol{\theta}_{pt}$.

[4]For simplicity, we assume that the sources are with equal power, $\sigma_{x,k}^2=\sigma_x^2, k=1,2,\dots,N_t$.

The partial derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})=\log g(\boldsymbol{y}|\boldsymbol{\theta})$ is

$$\frac{\partial \ell}{\partial \theta_i^*}=-\,\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\right\}+\boldsymbol{y}^H\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{y}. \quad (35)$$

where $\frac{\partial \boldsymbol{R}}{\partial \tilde{h}_i^*}=\sigma_x^2\mathcal{T}(\tilde{\boldsymbol{h}})\mathcal{T}\big(\frac{\partial \tilde{\boldsymbol{h}}}{\partial \tilde{h}_i}\big)^H$, $\frac{\partial \boldsymbol{R}}{\partial \tilde{h}_i}=\big(\frac{\partial \boldsymbol{R}}{\partial \tilde{h}_i^*}\big)^*$, $\frac{\partial \boldsymbol{R}}{\partial \sigma_x^2}=\tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{\mathcal{H}}}^H$, and $\frac{\partial \boldsymbol{R}}{\partial \sigma^2}=\boldsymbol{I}_{N_rN}$.

The misspecified FIM $\boldsymbol{J}_{\boldsymbol{\theta}}$ is derived from $\mathbb{E}_f\left\{\frac{\partial \ell}{\partial \theta_i^*}\frac{\partial \ell}{\partial \theta_j}\right\}$

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\theta}}(i,j)&=\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{C}\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_j}\boldsymbol{R}^{-1}\boldsymbol{C}\right\}\\&+\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_j}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}.\end{aligned} \quad (36)$$

Taking $\mathbb{E}_f\left\{\frac{\partial^2 \ell}{\partial \theta_j\partial \theta_i^*}\right\}$, we obtain $\boldsymbol{A}_{\boldsymbol{\theta}}$ as follows

$$\begin{aligned}\boldsymbol{A}_{\boldsymbol{\theta}}(i,j)&=-\,\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_j}\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\} \quad (37)\\&+\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial^2 \boldsymbol{R}}{\partial \theta_j\partial \theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}-\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_i^*}\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \theta_j}\boldsymbol{R}^{-1}\boldsymbol{C}\right\}.\end{aligned}$$

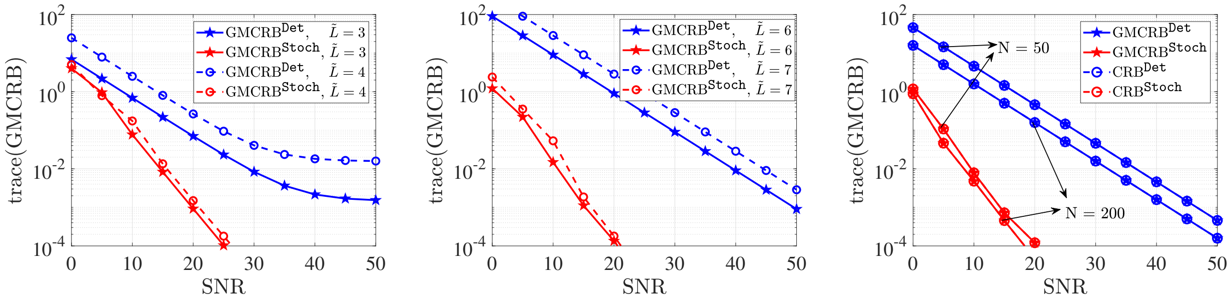
V. Examples & Discussions

The following simulations correspond to the convolutive MIMO system: the number of receive antennas $N_r=3$, of transmit antennas $N_t=2$, the true channel order $L_{tr}=5$ and the number of data samples $N=50$. Experimental results are averaged over 10 independent runs.

Fig. 1(a) and Fig. 1(b) plot the trace of GMCRB bounds (w.r.t. the channel parameters) versus $\mathrm{SNR}=10\log_{10}(\sigma_x^2/\sigma_n^2)$ in the presence of channel order underestimations and overestimations, respectively. We can see that the GMCRB$^{\mathtt{Stoch}}$ bounds are much lower than the GMCRB$^{\mathtt{Det}}$ in both cases, especially at high SNRs. In the former case, the GMCRB$^{\mathtt{Det}}$ tends to converge towards an error level as SNR increases, whereas the GMCRB$^{\mathtt{Stoch}}$ may be inversely proportional to SNR. Probably because the error mean $\boldsymbol{e}=\boldsymbol{\mu}-\tilde{\boldsymbol{\mathcal{H}}}_{pt}\tilde{\boldsymbol{x}}_{pt}$ is independent of the noise and hence $\sigma_{pt}^2\approx\|\boldsymbol{e}\|_2^2/(N_rN)\gg\sigma_n^2$ at high SNRs, while the GMCRB$^{\mathtt{Det}}$ is proportional to σ_{pt}^2. In the Gaussian stochastic models, we, however, do not misspecify the mean $\boldsymbol{\mu}$, but the covariance $\boldsymbol{C}$.

When the channel order is overestimated, the GMCRB$^{\mathtt{Det}}$ and the GMCRB$^{\mathtt{Stoch}}$ are both proportional to the noise variance, as shown in Fig. 1(b). In this case, the pseudo-true channel is estimated as $\tilde{\boldsymbol{h}}_{pt}=[\boldsymbol{h}_{tr}^\top,\boldsymbol{0}]^\top$ and hence the mean and covariance are perfectly specified even if the number of parameters of interest is ill determined. The lesser number of unknown parameters needed to be estimated, the lower the performance bound provided by the GMCRB.

Fig. 1(c) shows that the proposed GMCRB bounds are identical to the classical CRB bounds when $\tilde{L}=L_{tr}$. Indeed, the pseudo-true $\boldsymbol{\theta}_{pt}$ is equal to the true parameter of interest $\boldsymbol{\phi}$ and $f(\boldsymbol{y}|\boldsymbol{\phi})=g(\boldsymbol{y}|\boldsymbol{\theta})$, i.e., the model is correctly specified. Accordingly, we obtain the error mean $\boldsymbol{e}=\boldsymbol{0}$ and the covari-



(a) Underestimation $\tilde{L} < L_{tr} = 5$. (b) Overestimation $\tilde{L} > L_{tr} = 5$. (c) Perfect estimation $\tilde{L} = L_{tr} = 5$.

Fig. 1: Proposed GMCRB bounds for blind channel estimation.

ance $\boldsymbol{R} = \boldsymbol{C}$, hence $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}} = -\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ in both GMCRB bounds. As a result, the $\mathrm{GMCRB}^{\mathtt{Det}}$ becomes

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \frac{1}{\sigma_n^2}\begin{bmatrix} \boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{H}}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\mathcal{H}}_2^H\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{H}}_2^H\boldsymbol{\mathcal{H}}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{X}}^\top\boldsymbol{\mathcal{X}}^H & \boldsymbol{\mathcal{X}}^\top\boldsymbol{\mathcal{H}}_2^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{H}}_2^\top\boldsymbol{\mathcal{X}}^* & \boldsymbol{\mathcal{H}}_2^\top\boldsymbol{\mathcal{H}}_2^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{N_r N}{\sigma_n^2} \end{bmatrix}, \quad (38)$$

and $\mathrm{GMCRB}^{\mathtt{Stoch}}$ turns out to be the well-known formula [6]

$$\boldsymbol{J}_{\boldsymbol{\theta}}(i,j) = \mathrm{tr}\left\{\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_i^*}\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_j}\right\}. \quad (39)$$

VI. Conclusions

In this paper, we addressed the problem of analyzing the theoretical performance limit of blind system identification techniques when the channel order is misspecified. Two closed-form expressions of the misspecified CRB were presented for the class of unbiased blind estimators when unknown symbols are (i) deterministic ($\mathrm{GMCRB}^{\mathtt{Det}}$) and (ii) stochastic ($\mathrm{GMCRB}^{\mathtt{Stoch}}$). Numerical experiments were provided to illustrate the validity of the two proposed bounds. Future works will derive the misspecified CRB bound for semi-blind system identification.

VII. Acknowledgment

This work was funded by the National Foundation for Science and Technology Development of Vietnam (Grant No. 01/2019/TN).

References

[1] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Commun. Surv. Tutor.*, vol. 9, no. 2, pp. 18–48, 2007.
[2] A. P. Liavas, P. A. Regalia, and J. Delmas, "Blind channel approximation: Effective channel order determination," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3336–3344, 1999.
[3] X. Wang, H. Wu, S. Y. Chang, Y. Wu, and J. Chouinard, "Efficient non-pilot-aided channel length estimation for digital broadcasting receivers," *IEEE Trans. Broadcast.*, vol. 55, no. 3, pp. 633–641, 2009.
[4] S. Karakütük and T. E. Tuncer, "Channel matrix recursion for blind effective channel order estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1642–1653, 2011.
[5] J. Tian, T. Zhou, T. Xu, H. Hu, and M. Li, "Blind estimation of channel order and SNR for OFDM systems," *IEEE Access*, vol. 6, pp. 12 656–12 664, 2018.
[6] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.
[7] E. de Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *IEEE Works. Signal Process. Adv. Wirel. Commun.*, 1997, pp. 129–132.
[8] E. de Carvalho, J. Cioffi, and D. Slock, "Cramer-Rao bounds for blind multichannel estimation," in *IEEE Global Telecommun. Conf.*, 2000, pp. 1036–1040.
[9] L. Berriche, K. Abed-Meraim, and J. C. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2004, pp. 397–400.
[10] F. Gao and A. Nallanathan, "Blind channel estimation for OFDM systems via a generalized precoding," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1155–1164, 2007.
[11] B. Su and K. Tseng, "Cramér-Rao bound for blind channel estimation in cyclic prefixed MIMO-OFDM systems with few received symbols," in *Asilomar Conf. Signal Syst. Comput.*, 2014, pp. 966–970.
[12] A. Ladaycia, A. Mokraoui, K. Abed-Meraim, and A. Belouchrani, "Performance bounds analysis for semi-blind channel estimation in MIMO-OFDM communications systems," *IEEE Trans. Wirel. Commun.*, vol. 16, no. 9, pp. 5925–5938, 2017.
[13] M. Nait-Meziane, K. Abed-Meraim, Z. Zhao, and N. L. Trung, "On the Gaussian Cramér-Rao bound for blind single-input multiple-output system identification: Fast and asymptotic computations," *IEEE Access*, vol. 8, pp. 166 503–166 512, 2020.
[14] Q. H. Vuong, "Cramer-Rao bounds for misspecified models," *Caltech*, 1986.
[15] C. D. Richmond and L. L. Horowitz, "Parameter bounds on estimation accuracy under model misspecification," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2263–2278, 2015.
[16] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 142–157, 2017.
[17] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug 1997.
[18] E. de Carvalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (Global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, 2004.
[19] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*. Cambridge University Press, 1994.
[20] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér–Rao bounds for real and complex parameters," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1032–1049, 2012.
[21] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, pp. 1–25, 1982.
[22] S. Fortunati, F. Gini, and M. S. Greco, "The constrained misspecified Cramér–Rao bound," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 718–721, 2016.
[23] S. Fortunati, "Misspecified Cramér-Rao bounds for complex unconstrained and constrained parameters," in *European Signal Process. Conf.*, 2017, pp. 1644–1648.