Misspecified Cramer-Rao Bounds for Blind Channel Estimation under Channel Order Misspecification

Le Trung Thanh, Karim Abed-Meraim, *Fellow, IEEE*, Nguyen Linh Trung, *Senior Member, IEEE*

***Abstract*—In estimation, the misspecified Cramer–Rao bound (MCRB), which is an extension of the well-known Cramer–Rao bound (CRB) when the underlying system model is misspecified, has recently attracted much attention. In this paper, we introduce a new interpretation of the MCRB, called the generalized MCRB (GMCRB), via the Moore–Penrose inverse operator. This bound is useful for singular problems and particularly blind channel estimation problems in which the Hessian matrix is noninvertible. Two closed-form expressions of the GMCRB are derived for unbiased blind estimators when the channel order is misspecified. The first bound deals with deterministic models where both the channel and unknown symbols are deterministic. The second one is devoted to stochastic models where we assume that transmitted symbols are unknown random variables i.i.d. drawn from a Gaussian distribution. Two case studies of channel order misspecification are investigated to demonstrate the effectiveness of the proposed GMCRBs over the classical CRBs. When the channel order is known or accurately estimated, both generalized bounds reduce to the classical bounds. Besides, the stochastic GMCRB is lower than the deterministic one, especially at high SNR.**

***Index Terms*—Peformance lower bounds, constrained Cramer-Rao bound, misspecification, MIMO, channel order.**

I. Introduction

Channel estimation is one of the most fundamental and essential problems in wireless communications. Methods for channel estimation can be categorized 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 that case one might rely on semi-blind estimation, such as [1].

For many estimators, accurate information about the channel order is of great importance, and it strongly impacts their estimation performance. It is well-known that most blind channel estimation techniques are sensitive to the channel order estimation errors [2]–[4]. Underestimation (underspecification) of the channel order degrades their performance significantly. When the channel order is overestimated (overspecified), some blind channel estimation techniques are robust to overspecification, such as [5]–[7]. However, their estimation accuracy is not as good as that of the well-known blind techniques (e.g., the subspace-based algorithm in [8] and the least-square algorithm in [9]) when the channel order is accurately estimated.

There are various methods for channel order estimation, for examples [10]–[18] to name just a few. Among them, the two most well-known algorithms are based on the information-theoretic criteria, they are: minimum description length (MDL) [10] and Akaike information criteria (AIC) [11]. These algorithms, however, only work well under certain conditions and do not always result in the true order in practice. For example, MDL inclines towards overestimation of the channel order at high SNR, while AIC aims to estimate order of the most significant part of the channel called effective channel order. By contrast, the method of Liavas *et al.* in [12] tend to underestimate the channel order [14]. This paper focuses on evaluating performance lower bounds for blind channel estimators in the presence of channel order estimation error, that is under channel order *misspecification*.

It is well known that the Cramer-Rao bound (CRB) provides a lower bound on the variance of any unbiased estimator and is often used as a benchmark for parameter estimators [19]. Various analytical expressions of CRB have been derived for channel estimation in general and blind estimation in particular, e.g. [20]–[26]. These bounds, however, are appropriate only for perfect specification models, i.e., the true channel order is already known or accurately estimated. This limitation motivates us to look for new performance bounds in order to deal with imperfect knowledge of channel order. Under channel order misspecification, *misspecified Cramer–Rao bound* (MCRB), considered as a generalization of the classical CRB [27]–[29], has been introduced as an attractive alternative, due to the fact that the true model which generates the data is generally different from the assumed model used to estimate parameters of interest. Under certain conditions, the MCRB results in the Huber sandwich covariance which provides the lowest bound on the variance of unbiased estimators sharing the same mean with the maximum likelihood estimator (MLE) when the underlying model is misspecified [30].

In the literature, several interpretations of the MCRB have been investigated for unconstrained, constrained, and Bayesian estimation under model misspecification [31]–[34]. The MCRB has already been applied in several signal pro-

Le Trung Thanh is with PRISME Laboratory, University of Orléans, France, and the Advanced Institute of Engineering and Technology (AVITECH), University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam. Email: trung-thanh.le@univ-orleans.fr, thanhletrung@vnu.edu.vn.

Karim Abed-Meraim is with PRISME Laboratory, University of Orléans, France. Email: karim.abed-meraim@univ-orleans.fr.

Nguyen Linh Trung (corresponding author) is with the Advanced Institute of Engineering and Technology (AVITECH), University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam. E-mail: linhtrung@vnu.edu.vn.

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

cessing and communications applications such as direction of arrival estimation [28], time of arrival estimation [35], and harmonic approximations of inharmonic signals [36].

In the MCRB context, it is always assumed that the Hessian matrix of the log-likelihood function, often referred to as the misspecified Fisher information matrix (MFIM), is nonsingular [29]. Otherwise, some parameters to be estimated can be expressed in terms of others which means they cannot be completely identified from the observed data [37]. Intuitively, the singularity of the Hessian matrix indicates that the estimation problem is irregular or singular and the variance of estimators could be infinite. It often happens in estimation problems that involve a large number of parameters such as (deep) neural networks [38]–[40], complex biological/biomedical systems [41]–[43] as well as blind channel estimation [44]–[46]. In such problems, the usual MCRB may be nonexistent and its properties cannot be applied directly.

In the literature, there exist several approaches to rectifying and regularizing estimation problems with a singular Hessian matrix. The first approach is to introduce prior information for the estimation model to deal with the existence and uniqueness of the solution [47]. In this way, the estimation becomes Bayesian inference, and the performance lower bounds turn out to be the Bayesian ones. However, incorporating the prior information may not be sufficient to ensure estimation success from real data [41], [48]. Moreover, such prior information is not always available in practice. The second approach is to perform biased estimation [49], [50]. It is shown that introducing an appropriate bias can handle the irregular estimation problems with singular Fisher information matrices [51]. The resulting bounds, however, may not be useful in applications that prefer unbiased estimates. The third approach is to impose additional constraints on the unknown parameters. The parametric constraints lead to a transformed estimation problem in which unbiased estimators with finite variance may exist. In general, the value and computational complexity of the constrained bounds (e.g., the constrained CRB) depend on the choice of constraints. However, the constrained bound is often under restrictive unbiasedness conditions in some situations; thus, it may not be a lower bound for unbiased constrained estimators [52], [53]. Another good alternative is to use the pseudo-inverse matrix [54], [55]. Mathematically, the pseudo-inverse can be derived from the nonzero eigenvalues and corresponding eigenvectors of the Hessian matrix, so the result is well-conditioned in most cases. It is also shown in [55] that the pseudo-inverse-based approach actually belongs to the class of constrained problems. In this study, we focus on the third approach.

Constrained MCRB (CMCRB) for unbiased estimators has recently been introduced by Fortunati *et al.* in [32], [33]. It is indicated that the CMCRB is consistent with the classical constrained CRB in [56], [57] and its existence is derived from the identifiability of the constrained pseudo-true parameter [32]. However, the expression of CMCRB may not be readily computable in some applications when the constraints are too complicated. The underlying motivation of this paper is to fill in this gap by providing a *generalized interpretation* of the MCRB so that we are able to determine the performance limit of estimators derived from singular problems in a simple but rigorous way.

The paper has three main contributions as follows. Based on the preliminary result in [58], we first introduce the new generalized interpretation of MCRB (referred to as GMCRB) via the Moore–Penrose inverse operator. Compared to the state-of-the-art bounds, the proposed GMCRB offers several appealing advantages. Among them is that GMCRB is mathematically much simpler and more accessible than CMCRB. In singular problems where prior constraints imposed on parameters of interest are too complicated and numerous, the computation of CMCRB may be intractable. By contrast, the proposed GMCRB is always existent, and its interpretation is not only intuitive and formally sound but also easy to compute. Moreover, GMCRB provides the tightest bound for CMCRB among all choices of parametric constraints to regularize the singular problem. Also, the generalized form of GMCRB holds for both singular and regular estimation problems. Specifically, GMCRB is identical to the usual MCRB for regular problems but also consistent with the classical CRB under well-specified models. In addition to a performance lower bound, we also establish a sufficient condition for parametric constraints to achieve the proposed bound. This result supports future studies of making elaborate designs of parametric constraints and deriving efficient constrained estimators for singular estimation problems.

Second, in the context of blind channel estimation, we derive two closed-form expressions of the GMCRB for analyzing the theoretical performance limit of blind estimators when the channel order is misspecified. One is devoted to deterministic models where unknown transmitted symbols are assumed to be deterministic. The other is for stochastic models where transmitted symbols are unknown random variables i.i.d. drawn from a stochastic Gaussian process. The variance of any misspecified-unbiased estimator is always higher than the stochastic MCRB. The two proposed MCRBs, for the first time, provide performance lower bounds for blind channel estimation techniques under model misspecification.

And third, two case studies of channel order misspecification are investigated in detail. We show that the channel order underspecification leads to more problems than the overspecification. While the overspecification may give rise to the inefficiency of channel order estimation only, the underspecification introduces bias and hence several misleading results, e.g. gain in efficiency from excluding some channel parameters. The case studies suggest that we should always prefer overspecification to underspecification in practice.

The rest of the paper is structured as follows. In Section II, the usual MCRB under model misspecification is briefly reviewed. In Section III, our new generalized interpretation of the MCRB is established, via the Moore-Penrose inverse operator. In Sections IV and V, the problem of blind MIMO channel estimation is formulated and then the two closed-formed expressions of the MCRB are derived, for unbiased blind estimators when the channel order is misspecified. In Section VI, two case studies of channel order misspecification are investigated. Conclusions are given in Section VII.

II. Misspecified Cramer-Rao Bound

We briefly reviews the MCRB, which is an extension of the CRB for dealing with misspecified models [29]. For further information of its derivation, properties and applications, we refer the reader to [27]–[29] and references therein. Assume that the sample data follow the true distribution $f(\boldsymbol{y})$. However, users adopt a different distribution $g(\boldsymbol{y}|\boldsymbol{\theta}), \boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_n]^\top \in \boldsymbol{\Theta}$, to characterize statistics of $\boldsymbol{y}$ instead, where $g(\boldsymbol{y}|\boldsymbol{\theta}) \neq f(\boldsymbol{y})$, for all $\boldsymbol{\theta}$. For the users, the problem of interest is 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 [29].

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

$$\mathrm{KL}\left(f(\boldsymbol{y}) \| g(\boldsymbol{y}|\boldsymbol{\theta})\right) \triangleq \mathbb{E}_f\left\{\log\left(\frac{f(\boldsymbol{y})}{g(\boldsymbol{y}|\boldsymbol{\theta})}\right)\right\}. \tag{1}$$

The unique parameter minimizing $\mathrm{KL}(f(\boldsymbol{y})\|g(\boldsymbol{y}|\boldsymbol{\theta}))$ is the so-called *pseudo-true* parameter, $\boldsymbol{\theta}_{pt}$. In practice, $\mathrm{KL}(f(\boldsymbol{y})\|g(\boldsymbol{y}|\boldsymbol{\theta}))$ 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 (1) 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.,

$$\boldsymbol{\theta}_{pt} \triangleq \operatorname*{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{KL}(f(\boldsymbol{y})\|g(\boldsymbol{y}|\boldsymbol{\theta})) = \operatorname*{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{E}_f\left\{\ell(\boldsymbol{y}|\boldsymbol{\theta})\right\}. \tag{2}$$

Accordingly, it is shown in [27], [30] that the MLE $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ converges in probability to $\boldsymbol{\theta}_{pt}$. Moreover, its asymptotic covariance matrix is equal to the MCRB which will be later defined in next subsection.

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

$$\mathbb{E}_f\left\{\hat{\boldsymbol{\theta}}(\boldsymbol{y})\right\} = \int \hat{\boldsymbol{\theta}}(\boldsymbol{y}) f(\boldsymbol{y}) d\boldsymbol{y} = \boldsymbol{\theta}_{pt}. \tag{3}$$

We define the following matrices $\boldsymbol{J}_{\boldsymbol{\theta}}$ and $\boldsymbol{A}_{\boldsymbol{\theta}}$:[2]

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \mathbb{E}_f\left\{\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\right)^H\right\} \text{ and } \boldsymbol{A}_{\boldsymbol{\theta}} = \mathbb{E}_f\left\{\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^H}\right\}, \tag{4}$$

where $\ell \triangleq \ell(\boldsymbol{y}|\boldsymbol{\theta})$. When $\boldsymbol{A}_{\boldsymbol{\theta}}$ is nonsingular at $\boldsymbol{\theta} = \boldsymbol{\theta}_{pt}$, the error covariance matrix of any MS-unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ is bounded by MCRB [29]

$$\boldsymbol{C}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{\theta}_{pt}\big) \succeq \mathrm{MCRB}(\boldsymbol{\theta}_{pt}) \triangleq \boldsymbol{A}^{-1}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{-1}_{\boldsymbol{\theta}_{pt}}, \tag{5}$$

where

$$\boldsymbol{C}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{\theta}_{pt}\big) = \mathbb{E}_f\Big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta}_{pt}\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}) - \boldsymbol{\theta}_{pt}\big)^H\Big\}, \tag{6}$$

and " $\succeq$ " denotes the positive semidefinite order.

C. Constrained MCRB

When additional constraints are imposed on $\boldsymbol{\theta}$, the constrained version of MCRB has recently been introduced by Fortunati *et al.* in [32], [33], hereafter referred to as CMCRB.

Suppose that $\boldsymbol{\theta}$ is required to satisfy the following constraint: $\boldsymbol{u}(\boldsymbol{\theta}) = \boldsymbol{0}$.[3] If the Jacobian matrix $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}$ is full row-rank for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and there exists a matrix $\boldsymbol{U}$ spanning its null space,

$$\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})\boldsymbol{U} = \boldsymbol{0} \;\text{ and }\; \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}, \tag{7}$$

then the following holds for the CMRCB:

$$\begin{aligned}\boldsymbol{C}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{\theta}_{pt}\big) \succeq \boldsymbol{U}\big(\boldsymbol{U}^H\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U}\big)^{-1}\big(\boldsymbol{U}^H\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U}\big)\times \\ \times\big(\boldsymbol{U}^H\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}\boldsymbol{U}\big)^{-1}\boldsymbol{U}^H \triangleq \mathrm{CMCRB}_{\boldsymbol{u}}(\boldsymbol{\theta}_{pt}),\end{aligned} \tag{8}$$

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

III. Generalized Misspecified Cramer-Rao Bound

In this section, we introduce the new generalized interpretation of the MCRB, hereafter referred to as GMCRB. Our main result is stated in the following lemma.

Lemma 1. *Let $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ 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*

$$\boldsymbol{C}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{\theta}_{pt}\big) \succeq \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \triangleq \mathrm{GMCRB}(\boldsymbol{\theta}_{pt}), \tag{9}$$

where $(\cdot)^{\#}$ denotes the Moore–Penrose inverse operator and the two matrices $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ and $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ are defined as in (4).

Proof Sketch. The proof consists of two cases: (a) full rank $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ (i.e., regular estimation) and (b) singular $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ (i.e., singular estimation). For case (a), we indicate that GMCRB boils down to the usual MCRB, hence (9) is proved. For case (b), we prove that the proposed GMCRB is the tightest bound for CMCRB under any constraints imposed to regularizing the estimation. It results in $\boldsymbol{C}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{\theta}_{pt}) \succeq \mathrm{CMCRB}_{\boldsymbol{u}}(\boldsymbol{\theta}_{pt}) \succeq \mathrm{GMCRB}(\boldsymbol{\theta}_{pt})\ \forall \boldsymbol{u}(\boldsymbol{\theta})$. □

In what follows, we provide the detailed proof of Lemma 1 as well as present the connection to the existing MCRBs.

[1] When dealing with a mixture of both real parameters ($\boldsymbol{\theta}_r \in \mathbb{R}^{n_r}$) and complex parameters ($\boldsymbol{\theta}_c \in \mathbb{C}^{n_c}, \boldsymbol{\theta}_c \notin \mathbb{R}^{n_c}$), we consider the augmented representation: $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_c^\top, \boldsymbol{\theta}_c^H, \boldsymbol{\theta}_r^\top\right]^\top \in \mathbb{C}^{2n_c} \times \mathbb{R}^{n_r}$.

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

[3] In some applications, parameters of interest are constrained to lie in a proper subset of the original parameter space. Some examples are positivity, atomicity and bandwidth constraints in phase retrieval [59], array geometry constraints in direction-of-arrival estimation [60], and prior channel state information in wireless channel estimation [61]. The function $\boldsymbol{u}(\boldsymbol{\theta})$ aims to represent imposed constraints on unknown parameters in such applications. Moreover, $\boldsymbol{u}(\boldsymbol{\theta})$ may contain additional constraints to regularizing singular estimation problems.

First, when $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ is nonsingular, the proposed GMCRB reduces to the unconstrained MCRB in (5), i.e.,

$$\mathrm{GMCRB}(\boldsymbol{\theta}_{pt}) = \boldsymbol{A}^{-1}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{-1}_{\boldsymbol{\theta}_{pt}} = \mathrm{MCRB}(\boldsymbol{\theta}_{pt}). \tag{10}$$

When $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ is singular, 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 $\boldsymbol{u}(\boldsymbol{\theta})$ with a full rank $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})$, we will prove

$$\mathrm{CMCRB}_{\boldsymbol{u}}(\boldsymbol{\theta}_{pt}) \geq \mathrm{GMCRB}(\boldsymbol{\theta}_{pt}). \tag{11}$$

***Step I**: The link between GMCRB and CMCRB*

Performing eigenvalue decomposition (EVD) of $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \in \mathbb{C}^{n\times n}$ of rank r results in

$$\boldsymbol{A}_{\boldsymbol{\theta}_{pt}} = \begin{bmatrix} \boldsymbol{U}_r & \boldsymbol{V}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_r & \\ & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_r^H \\ \boldsymbol{V}_r^H \end{bmatrix}, \tag{12}$$

where $\boldsymbol{V}_r^H \boldsymbol{U}_r = \boldsymbol{0}$. Accordingly, we have

$$\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} = \boldsymbol{U}_r \boldsymbol{\Sigma}_r^{-1} \boldsymbol{U}_r^H = \boldsymbol{U}_r \left(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\right)^{-1} \boldsymbol{U}_r^H. \tag{13}$$

The expression (9) becomes

$$\begin{aligned} \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} &= \boldsymbol{U}_r \left(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\right)^{-1} \times \\ &\times \left(\boldsymbol{U}_r^H \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\right) \left(\boldsymbol{U}_r^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\right)^{-1} \boldsymbol{U}_r^H, \end{aligned} \tag{14}$$

which is identical to the CMCRB in (8). Accordingly, the proposed GMCRB holds for the CMCRB under the constraint function $\boldsymbol{u}(\boldsymbol{\theta})$ satisfying $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta}) = \boldsymbol{V}_r^H$.

***Stage II:** GMCRB is the minimum CMCRB*

The result follows immediately the next three propositions.

Proposition 1. *For any constraint $\boldsymbol{u}(\boldsymbol{\theta})$ such that $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})$ is full row-rank, $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})\boldsymbol{U} = \boldsymbol{0}$, $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_{r'}$, and $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ is nonsingular, we have*

$$\min_{\boldsymbol{u}(\boldsymbol{\theta})} \ \mathrm{rank}\left(\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})\right) = n - r. \tag{15}$$

Proof. We have $\mathrm{rank}(\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})) = n - r', 0 < r' < n$, and $\mathrm{rank}(\boldsymbol{U}) = r'$ by definition (7).

Under the assumption that $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ is nonsingular, we have $\mathrm{rank}(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}) = r'$ because $\boldsymbol{U}$ is full column-rank.

In parallel, we know that

$$\mathrm{rank}(\boldsymbol{M}\boldsymbol{N}) \leq \min\left(\mathrm{rank}(\boldsymbol{M}), \mathrm{rank}(\boldsymbol{N})\right) \ \forall \boldsymbol{M}, \boldsymbol{N},$$

hence

$$\begin{aligned} r' = \mathrm{rank}\left(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\right) &\leq \min\left(\mathrm{rank}(\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}), \mathrm{rank}(\boldsymbol{U})\right) \\ &= \min(r, r'). \end{aligned} \tag{16}$$

Accordingly, we obtain $r' \leq r$ and

$$\mathrm{rank}(\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})) = n - r' \geq n - r.$$

Equality is achieved when the constraint function $\boldsymbol{u}(\boldsymbol{\theta})$ satisfies $\nabla_{\boldsymbol{u}}(\boldsymbol{\theta}) = \boldsymbol{V}_r^H$, i.e., $\mathrm{rank}(\nabla_{\boldsymbol{u}}(\boldsymbol{\theta})) = n - r$. □

As a consequence, we may need at least $n - r$ constraints to insure the identifiability of unknown parameters when $\mathrm{rank}(\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}) = r < n$. The next proposition will indicate that the eigenvector matrix $\boldsymbol{U}_r$ is linked to the optimal minimum set of constraints to regularizing the singular estimation problem.

Proposition 2. *For any orthogonal matrix $\boldsymbol{U} \in \mathbb{C}^{n\times r}$, (i.e., $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_r$) and $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ is nonsingular, we have*

$$\lambda_i\left[\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}}\right] \leq \lambda_i\left[\boldsymbol{U}\left(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^H\right], i = 1, 2, \ldots, r, \tag{17}$$

where $\lambda_i[\boldsymbol{M}]$ is the i-th largest eigenvalue of $\boldsymbol{M}$.

Proof. By definition, $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \geq \boldsymbol{0}$, i.e., there exists a matrix $\boldsymbol{L}_A$ such that $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}} = \boldsymbol{L}_A^H \boldsymbol{L}_A$. Thus, we have

$$\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U} = \boldsymbol{U}^H \boldsymbol{L}_A^H \boldsymbol{L}_A \boldsymbol{U} = (\boldsymbol{L}_A\boldsymbol{U})^H (\boldsymbol{L}_A\boldsymbol{U}) \geq \boldsymbol{0}. \tag{18}$$

Since $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ is nonsingular, $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ and its inverse are positive definite.

For a given $\boldsymbol{B} > \boldsymbol{0}$, we can express

$$\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H = \boldsymbol{U}\boldsymbol{U}_B \boldsymbol{\Sigma}_B \boldsymbol{U}_B^H \boldsymbol{U}^H = \boldsymbol{V}\boldsymbol{\Sigma}_B \boldsymbol{V}^H, \tag{19}$$

where $\boldsymbol{B} \overset{\mathrm{EVD}}{=} \boldsymbol{U}_B \boldsymbol{\Sigma}_B \boldsymbol{U}_B^H$ and $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{U}_B$ is orthogonal. Therefore the eigenvalues of $\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^H$ are identical to those of $\boldsymbol{B}$. Accordingly, we obtain

$$\lambda_i\left[\boldsymbol{U}(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U})^{-1} \boldsymbol{U}^H\right] = \lambda_i\left[(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U})^{-1}\right]. \tag{20}$$

The inequality of (17) becomes

$$\lambda_i\left[\boldsymbol{\Sigma}_r^{-1}\right] \leq \lambda_i\left[(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U})^{-1}\right], i = 1, 2, \ldots, r. \tag{21}$$

We also know that $\boldsymbol{M} \geq \boldsymbol{N} > \boldsymbol{0}$ if and only if $\boldsymbol{M}^{-1} \leq \boldsymbol{N}^{-1}$ [62, Theorem 24, page 24], thus (21) is equivalent to

$$\lambda_i\left[\boldsymbol{\Sigma}_r\right] \geq \lambda_i\left[\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\right], \tag{22}$$

since $\boldsymbol{\Sigma}_r$ and $\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}$ are both positive-definite matrices. Thanks to the Poincare separation theorem [62, page 236], the inequality of (22) is always true. It ends the proof. □

Therefore, we can conclude that

$$\boldsymbol{A}^{\#}_{\boldsymbol{\theta}_{pt}} \leq \boldsymbol{U}\left(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^H \text{ when } \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_r. \tag{23}$$

Proposition 3. *Given three positive-semidefinite Hermitian matrices of the same rank $\boldsymbol{M}, \boldsymbol{N}$, and $\boldsymbol{X}$, if $\boldsymbol{M} \geq \boldsymbol{N}$ then*

$$\boldsymbol{M}\boldsymbol{X}\boldsymbol{M} \geq \boldsymbol{N}\boldsymbol{X}\boldsymbol{N}. \tag{24}$$

Proof. Similar to (18), we also obtain

$$\boldsymbol{E}\boldsymbol{X}\boldsymbol{E} = \boldsymbol{E}^H \boldsymbol{L}^H \boldsymbol{L}\boldsymbol{E} = (\boldsymbol{L}\boldsymbol{E})^H (\boldsymbol{L}\boldsymbol{E}) \geq \boldsymbol{0}, \tag{25}$$

for all positive-semidefinite matrix $\boldsymbol{E}$ ($\boldsymbol{L}$ being a square root matrix of $\boldsymbol{X}$).

Now, let us denote $\boldsymbol{E} = \boldsymbol{M} - \boldsymbol{N} \geq \boldsymbol{0}$, we have

$$\begin{aligned} &\boldsymbol{M}\boldsymbol{X}\boldsymbol{M} - \boldsymbol{N}\boldsymbol{X}\boldsymbol{N} = (\boldsymbol{M} - \boldsymbol{N})\boldsymbol{X}\boldsymbol{M} + \boldsymbol{N}\boldsymbol{X}(\boldsymbol{M} - \boldsymbol{N}) \\ &= (\boldsymbol{M} - \boldsymbol{N})\boldsymbol{X}(\boldsymbol{M} - \boldsymbol{N}) + (\boldsymbol{M} - \boldsymbol{N})\boldsymbol{X}\boldsymbol{N} + \boldsymbol{N}\boldsymbol{X}(\boldsymbol{M} - \boldsymbol{N}) \\ &= \boldsymbol{E}\boldsymbol{X}\boldsymbol{E} + \boldsymbol{E}\boldsymbol{X}\boldsymbol{N} + \boldsymbol{N}\boldsymbol{X}\boldsymbol{E}, \end{aligned} \tag{26}$$

which has the form of the generalized continuous-time Lyapunov equation (GCTLE) [63], [64].

Since $\boldsymbol{E}$ and $\boldsymbol{N}$ are both positive semidefinite, there exists a nonsingular Hermitian matrix $\boldsymbol{Q}$ such that

$$\boldsymbol{E} = \boldsymbol{Q}\,\mathrm{diag}(\boldsymbol{\alpha})\boldsymbol{Q}^H \text{ and } \boldsymbol{N} = \boldsymbol{Q}\,\mathrm{diag}(\boldsymbol{\beta})\boldsymbol{Q}^H, \tag{27}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}$, thanks to [65, Theorem 7.6.4, page 487] and [66, Theorem 39, page 103]. Therefore, all generalized eigenvalues of the matrix pencil $(-\boldsymbol{E}, \boldsymbol{N})$ are nonpositive. Thanks to Lyapunov theorem [64, Theorem 4.4, page 47], we can conclude that the GCTLE of (26) is positive semidefinite. It ends the proof. $\square$

Accordingly, thanks to Propositions 2 and 3, we obtain

$$\begin{aligned} \boldsymbol{A}_{\boldsymbol{\theta}_{pt}}^{\#} \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{A}_{\boldsymbol{\theta}_{pt}}^{\#} \preceq \boldsymbol{U}\big(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}_r\big)^{-1} \boldsymbol{U}^H \times \\ \times \boldsymbol{J}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big(\boldsymbol{U}^H \boldsymbol{A}_{\boldsymbol{\theta}_{pt}} \boldsymbol{U}\big)^{-1} \boldsymbol{U}^H = \mathrm{CMCRB}_{\boldsymbol{u}}(\boldsymbol{\theta}_{pt}). \quad (28) \end{aligned}$$

Remark 1. The proposed GMCRB is simpler and more accessible than the CMCRB. In some applications, we have to deal with complicated (multiple) constraints on $\boldsymbol{\theta}$ to regularize the estimation problem, e.g., the channel state information or side information (such as power, bandwidth, and delay constraints) in wireless communications. Therefore, the expression of $\boldsymbol{U}$ in (7) may not be readily computable. Moreover, different constraints $\boldsymbol{u}(\boldsymbol{\theta})$ may result in different values of the CMCRB. Our GMCRB provides the tightest bound for the CMCRB.

Remark 2. The proposed GMCRB is consistent with the usual MCRB for regular estimation problems, as shown in (10).

Remark 3. When the model is perfectly specified, the GMCRB reduces to its counterpart in [55]. Particularly, $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ and $-\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ become the Fisher information matrix, that is, $\mathrm{GMCRB}(\boldsymbol{\theta}_{pt}) = \boldsymbol{J}_{\boldsymbol{\theta}_{pt}}^{\#}$.

IV. MIMO System Model

We consider a convolutive MIMO system with N_t transmit antennas and N_r receive antennas whose channels are modelled to be finite impulse response (FIR). Suppose that the MIMO channel remains constant during the transmission period. Each transmitted block includes N data symbols and a cyclic prefix (CP) of appropriate length, used a guard interval to avoid the intersymbol interference from two successive blocks.

The output vector $\boldsymbol{y}[t] = \big[y_1[t], y_2[t], \ldots, y_{N_r}[t]\big]^\top \in \mathbb{C}^{N_r \times 1}$ received at N_r receive antennas is given by

$$\boldsymbol{y}[t] = \sum_{i=0}^{L-1} \boldsymbol{H}[i]\boldsymbol{x}[t-i] + \boldsymbol{n}[t], \quad (29)$$

where $\boldsymbol{H}$ is the $N_r \times N_t$ 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 a $N_r \times 1$ additive noise vector drawn from an i.i.d. circular complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_{N_r})$, $\mathbb{E}\{\boldsymbol{n}[t]\boldsymbol{n}[t]^\top\} = \mathbf{0}$.

After removing the CP, we often stack N output samples $\{\boldsymbol{y}[t]\}_{t=0}^{N-1}$ into a single vector $\boldsymbol{y} \in \mathbb{C}^{NN_r \times 1}$ as

$$\boldsymbol{y} = \big[\boldsymbol{y}[0]^\top, \boldsymbol{y}[1]^\top, \ldots, \boldsymbol{y}[N-1]^\top\big]^\top. \quad (30)$$

Accordingly, we can recast the convolution (29) into the following standard linear expression:

$$\boldsymbol{y} = \boldsymbol{\mathcal{X}}\boldsymbol{h} + \boldsymbol{n} = \big(\boldsymbol{X}^\top \otimes \boldsymbol{I}_{N_r}\big)\boldsymbol{h} + \boldsymbol{n}, \quad (31)$$

where the vector of channel parameters $\boldsymbol{h}$ and the noise vector $\boldsymbol{n}$ are given by

$$\boldsymbol{h} = \big[\boldsymbol{h}_{L-1}^\top, \boldsymbol{h}_{L-2}^\top, \ldots, \boldsymbol{h}_0^\top\big]^\top, \quad (32)$$

$$\boldsymbol{n} = \big[\boldsymbol{n}[0]^\top, \boldsymbol{n}[1]^\top, \ldots, \boldsymbol{n}[N-1]^\top\big]^\top, \quad (33)$$

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

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}[N-L+1] & \boldsymbol{x}[N-L+2] & \ldots & \boldsymbol{x}[N-L] \\ \vdots & \vdots & \iddots & \vdots \\ \vdots & \boldsymbol{x}[N-1] & \iddots & \vdots \\ \boldsymbol{x}[N-1] & \boldsymbol{x}[0] & \ldots & \boldsymbol{x}[N-2] \\ \boldsymbol{x}[0] & \boldsymbol{x}[1] & \ldots & \boldsymbol{x}[N-1] \end{bmatrix}, \quad (34)$$

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

Due to the commutativity of convolution, we also have

$$\boldsymbol{\mathcal{X}}\boldsymbol{h} = \mathcal{T}(\boldsymbol{h})\boldsymbol{x}, \quad (35)$$

where $\boldsymbol{\mathcal{H}} = \mathcal{T}(\boldsymbol{h})$ is the $N_r N \times N_t N$ channel Toeplitz matrix whose the first block column is $\big[\boldsymbol{H}[0]^\top, \ldots, \boldsymbol{H}[L-1]^\top, \mathbf{0}\big]^\top$

$$\mathcal{T}(\boldsymbol{h}) = \begin{bmatrix} \boldsymbol{H}[0] & & \mathbf{0} & & \boldsymbol{H}[1] \\ \boldsymbol{H}[1] & \ddots & \vdots & & \boldsymbol{H}[2] \\ \vdots & \ddots & \boldsymbol{H}[0] & & \vdots \\ \vdots & \ddots & \boldsymbol{H}[1] & \ddots & \boldsymbol{H}[L-1] \\ \boldsymbol{H}[L-1] & \ddots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \boldsymbol{H}[L-1] & \ddots & \vdots \\ \vdots & & \mathbf{0} & \ddots & \vdots \\ \vdots & & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \ldots & \mathbf{0} & & \boldsymbol{H}[0] \end{bmatrix}, \quad (36)$$

and the data vector $\boldsymbol{x}$ is given by

$$\boldsymbol{x} = \big[\boldsymbol{x}[0]^\top, \boldsymbol{x}[1]^\top, \ldots, \boldsymbol{x}[N-1]^\top\big]^\top \in \mathbb{C}^{NN_t \times 1}. \quad (37)$$

Generally, necessary and sufficient conditions are required for blind identifiability of the MIMO system [67]. These identifiability conditions are often formulated in terms of the concepts of "zeros" and "modes". Particularly, the channel transfer function should be irreducible, the number of input modes and samples must be larger than the channel order. We refer the reader to [67]–[69] for more details.

V. GMCRB for Blind Channel Estimation

Blind techniques consider the estimation of channel parameters from outputs only. Without prior information or additional constraints, it is well-known that the blind estimation is singular due to the inherent matrix ambiguity [67], so the usual MCRB may not exist. Therefore, we propose to use the proposed GMCRB in (9), instead, to determine the performance limit of unbiased blind estimators when the channel order is misspecified. Particularly, we focus on the deterministic model and the stochastic model. In the former case, the unknown transmitted symbols are assumed to be deterministic, whereas in the latter case, they are assumed unknown random variables i.i.d. drawn from a Gaussian distribution.

A. Deterministic GMCRB

In this model, the output vector $\boldsymbol{y}$ is drawn from the Gaussian distribution $\mathcal{CN}\big(\boldsymbol{\mathcal{H}}\boldsymbol{x}, \sigma_n^2 \boldsymbol{I}_{N_r N}\big)$ and the vector of unknown parameters is $\boldsymbol{\phi} = \big[\boldsymbol{h}^\top, \boldsymbol{x}^\top, \boldsymbol{h}^H, \boldsymbol{x}^H, \sigma_n^2\big]^\top$.

Due to the imperfect knowledge of the channel order ($\tilde{L} \neq L$), the users will fit the assumed $g(\boldsymbol{y}|\boldsymbol{\theta})$ to $\boldsymbol{y}$

$$g(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{(\pi\sigma^2)^{N_r N}} \exp\bigg(-\frac{1}{\sigma^2}\big\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}\big\|_2^2\bigg), \tag{38}$$

with a new noise variance σ^2 and a new mean $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}} = \tilde{\boldsymbol{\mathcal{H}}}\boldsymbol{x}$. Here, $\tilde{\boldsymbol{\mathcal{X}}}$ is supposed to be formed as $\tilde{\boldsymbol{\mathcal{X}}} = \tilde{\boldsymbol{X}}^\top \otimes \boldsymbol{I}_{N_r}$ with

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}[N-\tilde{L}+1] & \boldsymbol{x}[N-\tilde{L}+2] & \dots & \boldsymbol{x}[N-\tilde{L}] \\ \vdots & \vdots & \iddots & \vdots \\ \vdots & \boldsymbol{x}[N-1] & \iddots & \vdots \\ \boldsymbol{x}[N-1] & \boldsymbol{x}[0] & \dots & \boldsymbol{x}[N-2] \\ \boldsymbol{x}[0] & \boldsymbol{x}[1] & \dots & \boldsymbol{x}[N-1] \end{bmatrix}, \tag{39}$$

and $\tilde{\boldsymbol{\mathcal{H}}} = \mathcal{T}(\tilde{\boldsymbol{h}})$ is the block circulant matrix whose the first block column $\big[\boldsymbol{H}[0]^\top, \boldsymbol{H}[1]^\top, \dots, \boldsymbol{H}[\tilde{L}-1]^\top, \boldsymbol{0}\big]^\top$. Instead of $\boldsymbol{\phi}$, the parameters of interest now become

$$\boldsymbol{\theta} = \big[\tilde{\boldsymbol{h}}^\top, \boldsymbol{x}^\top, \tilde{\boldsymbol{h}}^H, \boldsymbol{x}^H, \sigma^2\big]^\top. \tag{40}$$

The misspecified log-likelihood function is given by

$$\ell(\boldsymbol{y}|\boldsymbol{\theta}) = \text{const} - N_r N \log(\sigma^2) - \frac{1}{\sigma^2}\big\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}\big\|_2^2. \tag{41}$$

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

$$\boldsymbol{\theta}_{pt} = \operatorname*{argmin}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bigg\{ N_r N \log(\sigma^2) + \frac{\sigma_n^2 N_r N + \|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|_2^2}{\sigma^2} \bigg\}, \tag{42}$$

where $\boldsymbol{\mu} \triangleq \mathbb{E}_f\{\boldsymbol{y}\} = \boldsymbol{\mathcal{X}}\boldsymbol{h}$ is the true mean of $\boldsymbol{y}$. The minimizer of (42) can be obtained directly by applying the MLE (or elegant methods surveyed in [67]), e.g.,

$$\tilde{\boldsymbol{h}}_{pt} = \operatorname*{argmin}_{\tilde{\boldsymbol{h}}} \big\|\big(\boldsymbol{I} - \boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}}\big)\boldsymbol{\mu}\big\|_2^2, \tag{43}$$

$$\boldsymbol{x}_{pt} = \big(\tilde{\boldsymbol{\mathcal{H}}}_{pt}^H \tilde{\boldsymbol{\mathcal{H}}}_{pt}\big)^{\#} \tilde{\boldsymbol{\mathcal{H}}}_{pt}^H \boldsymbol{\mu}, \tag{44}$$

$$\sigma_{pt}^2 = \sigma_n^2 + \frac{\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mathcal{H}}}_{pt}\boldsymbol{x}_{pt}\|_2^2}{N_r N}, \tag{45}$$

where $\boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}} \triangleq \tilde{\boldsymbol{\mathcal{H}}}(\tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}})^{-1}\tilde{\boldsymbol{\mathcal{H}}}^H$.[4]

The first partial derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ is given by

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

[4]It is well known that the projection matrix $\boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}}$ is invariant to any scale on $\tilde{\boldsymbol{\mathcal{H}}}$, so the solution of (43) and (44) are not exactly unique, i.e. $\tilde{\boldsymbol{h}}$ and $\boldsymbol{x}$ are unique up to an unknown scalar. This scalar can be estimated by acquiring further knowledge about the estimation model, e.g. we can determine it completely with only one known sample of the input sequence $\boldsymbol{x}$ [67].

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

$$\mathbb{E}_f\big\{\boldsymbol{y} - \tilde{\boldsymbol{\mu}}\big\} = \boldsymbol{e},$$
$$\mathbb{E}_f\big\{\|\boldsymbol{y} - \tilde{\boldsymbol{\mu}}\|_2^2\big\} = \sigma_n^2 N_r N + \|\boldsymbol{e}\|_2^2,$$
$$\mathbb{E}_f\big\{(\boldsymbol{y} - \tilde{\boldsymbol{\mu}})(\boldsymbol{y} - \tilde{\boldsymbol{\mu}})^\top\big\} = \boldsymbol{e}\boldsymbol{e}^\top,$$
$$\mathbb{E}_f\big\{(\boldsymbol{y} - \tilde{\boldsymbol{\mu}})(\boldsymbol{y} - \tilde{\boldsymbol{\mu}})^H\big\} = \sigma_n^2 \boldsymbol{I}_{N_r N} + \boldsymbol{e}\boldsymbol{e}^H.$$

From (4), $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ is derived from $\mathbb{E}_f\Big\{\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\big(\frac{\partial \ell}{\partial \boldsymbol{\theta}^*}\big)^H\Big\}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_{pt}$

$$\boldsymbol{J}_{\boldsymbol{\theta}_{pt}} = \frac{1}{\sigma_{pt}^4} \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{h},\boldsymbol{x}^*} & \boldsymbol{0} \\ \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{x},\boldsymbol{x}^*} & \boldsymbol{0} \\ \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{h}^*,\boldsymbol{x}^*} & \boldsymbol{0} \\ \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{h}^*} & \boldsymbol{J}_{\boldsymbol{x}^*,\boldsymbol{x}^*} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & N_r N \end{bmatrix}, \tag{46}$$

where

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

Taking the expectation of the second partial derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ over $f(\boldsymbol{y})$ results in $\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ as

$$\boldsymbol{A}_{\boldsymbol{\theta}_{pt}} = \frac{-1}{\sigma_{pt}^2} \begin{bmatrix} \tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}} & \tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{H}}} & \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{e} \\ \tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{X}}} & \tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}} & \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{H}}}^H\boldsymbol{e} \\ \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{X}}}^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \tilde{\boldsymbol{\mathcal{X}}}^\top\tilde{\boldsymbol{\mathcal{H}}}^* & \tilde{\boldsymbol{\mathcal{X}}}^\top\boldsymbol{e}^* \\ \boldsymbol{0} & \boldsymbol{0} & \tilde{\boldsymbol{\mathcal{H}}}^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \tilde{\boldsymbol{\mathcal{H}}}^\top\tilde{\boldsymbol{\mathcal{H}}}^* & \tilde{\boldsymbol{\mathcal{H}}}^\top\boldsymbol{e}^* \\ \boldsymbol{e}^H\tilde{\boldsymbol{\mathcal{X}}} & \boldsymbol{e}^H\tilde{\boldsymbol{\mathcal{H}}} & \boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{X}}}^* & \boldsymbol{e}^\top\tilde{\boldsymbol{\mathcal{H}}}^* & \frac{N_r N}{\sigma_{pt}^2} \end{bmatrix}. \tag{47}$$

B. Stochastic GMCRB

Suppose that the unknown transmitted symbols are circular Gaussian random variables i.i.d. drawn from $\mathcal{CN}(\boldsymbol{0}, \sigma_x^2\boldsymbol{I}_{N_t})$. For simplicity, we assume that the sources are with equal power, $\sigma_{x,k}^2 = \sigma_x^2, k = 1, 2, \dots, N_t$. 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 $\boldsymbol{C} = \sigma_x^2\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H + \sigma_n^2\boldsymbol{I}_{N_r N}$.

In the presence of channel order estimation error, the following distribution function $g(\boldsymbol{y}|\boldsymbol{\theta})$ is used instead

$$g(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{\pi^{N_r N}\det\boldsymbol{R}} \exp\Big(-\boldsymbol{y}^H\boldsymbol{R}^{-1}\boldsymbol{y}\Big), \tag{48}$$

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

$$\boldsymbol{\theta} = \big[\tilde{\boldsymbol{h}}^\top, \tilde{\boldsymbol{h}}^H, \sigma_x^2, \sigma^2\big]^\top. \tag{49}$$

In this case, $\mathrm{KL}\big(f(\boldsymbol{y})\|g(\boldsymbol{y}|\boldsymbol{\theta})\big)$ is given by

$$\mathrm{KL}\big(f(\boldsymbol{y})\|g(\boldsymbol{y}|\boldsymbol{\theta})\big) = \log\left(\frac{\det \boldsymbol{R}}{\det \boldsymbol{C}}\right) + \mathrm{tr}\left\{\boldsymbol{R}^{-1}\boldsymbol{C}\right\} - N_r N. \tag{50}$$

Accordingly, the *pseudo-true* parameter $\boldsymbol{\theta}_{pt}$ is derived from

$$\boldsymbol{\theta}_{pt} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\left\{\log\big(\det \boldsymbol{R}\big) + \mathrm{tr}\left\{\boldsymbol{R}^{-1}\boldsymbol{C}\right\}\right\}. \tag{51}$$

The minimizer of (51) can be achieved by using the iterative procedure in [70], [71]. In particular, at each iteration step, we first optimize (51) w.r.t. σ_x^2 for a fixed $\tilde{\boldsymbol{h}}$ and σ^2. Given σ_x^2 and $\tilde{\boldsymbol{h}}$, we then estimate the noise power σ^2. Finally, we update $\tilde{\boldsymbol{h}}$ based on the recent estimation of σ_x^2 and σ^2.

Following the same derivation in [70], we exploit that the first stage is equivalent to find σ_x^2 satisfying

$$\tilde{\boldsymbol{\mathcal{H}}}^H\boldsymbol{R}^{-1}\big(\boldsymbol{C}-\boldsymbol{R}\big)\boldsymbol{R}^{-1}\tilde{\boldsymbol{\mathcal{H}}} = \boldsymbol{0}. \tag{52}$$

Therefore, we can obtain an estimate of σ_x^2 as

$$\sigma_x^2 = \frac{1}{N_t N}\mathrm{tr}\left\{\hat{\boldsymbol{R}}_x\right\}, \text{ where} \tag{53}$$

$$\hat{\boldsymbol{R}}_x = \big(\tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}}\big)^{-1}\big(\tilde{\boldsymbol{\mathcal{H}}}^H\boldsymbol{C}\tilde{\boldsymbol{\mathcal{H}}} - \sigma^2\tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}}\big)\big(\tilde{\boldsymbol{\mathcal{H}}}^H\tilde{\boldsymbol{\mathcal{H}}}\big)^{-1}. \tag{54}$$

The noise power σ_n^2 is updated by

$$\begin{aligned}\sigma^2 &= \underset{\xi^2}{\operatorname{argmin}}\left\{\frac{1}{\xi^2}\mathrm{tr}\left\{(\boldsymbol{I}-\boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}})\boldsymbol{C}\right\} + (N_rN - N_tN)\log\xi^2\right\}\\ &= \frac{1}{N_rN - N_tN}\mathrm{tr}\left\{\big(\boldsymbol{I}_{N_rN} - \boldsymbol{P}_{\tilde{\boldsymbol{\mathcal{H}}}}\big)\boldsymbol{C}\right\}.\end{aligned} \tag{55}$$

The channel parameter $\tilde{\boldsymbol{h}}$ can be derived directly from the matrix $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{h}}} = \sigma_x^2\tilde{\boldsymbol{\mathcal{H}}}\tilde{\boldsymbol{\mathcal{H}}}^H$. More concretely, $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{h}}}$ is obtained by minimizing the following optimization

$$\begin{aligned}\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{h}}} &= \underset{\boldsymbol{R}_{\tilde{\boldsymbol{h}}}\succeq 0}{\operatorname{argmin}}\left\{\log(\det\boldsymbol{R}) + \frac{N_tN}{N_rN}\sum_{i=1}^{N_rN}\frac{\boldsymbol{v}_i^H\boldsymbol{R}^{-1}\boldsymbol{v}_i}{\boldsymbol{v}_i^H\boldsymbol{R}_{\text{old}}^{-1}\boldsymbol{v}_i}\right\}\\ \text{s.t. } \boldsymbol{R} &= \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{h}}} + \sigma^2\boldsymbol{I}_{N_rN}.\end{aligned} \tag{56}$$

where $\boldsymbol{R}_{\text{old}}$ is the old estimation of $\boldsymbol{R}$ at the previous step and $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots, \boldsymbol{v}_{N_rN}]$ is the root square of the true $\boldsymbol{C}$ [71]. The closed-form solution of (56) is given by

$$\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{h}}} = \sum_{i=1}^{N_tN}(\lambda_i - \sigma^2)\boldsymbol{u}_i\boldsymbol{u}_i^H, \tag{57}$$

where λ_i is the i-th top eigenvalue of $\sum_{i=1}^{N_rN}\frac{\boldsymbol{v}_i\boldsymbol{v}_i^H}{\boldsymbol{v}_i^H\boldsymbol{R}_{\text{old}}^{-1}\boldsymbol{v}_i}$ and $\boldsymbol{u}_i$ is the corresponding eigenvector.

Now, we turn to the derivation of the stochastic GMCRB. The first derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta}) = \log g(\boldsymbol{y}|\boldsymbol{\theta})$ is

$$\frac{\partial\ell}{\partial\theta_i^*} = -\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\right\} + \boldsymbol{y}^H\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{y}, \tag{58}$$

where

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

At $\boldsymbol{\theta} = \boldsymbol{\theta}_{pt}$, we have $\mathbb{E}_f\left\{\frac{\partial\ell(\boldsymbol{y}|\boldsymbol{\theta})}{\partial\theta_i^*}\right\} = \boldsymbol{0}$, i.e.,

$$\mathbb{E}_f\left\{\boldsymbol{y}^H\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{y}\right\} = \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\right\}. \tag{59}$$

Accordingly, we obtain

$$\begin{aligned}\mathbb{E}_f\left\{\frac{\partial\ell}{\partial\theta_i^*}\frac{\partial\ell}{\partial\theta_j}\right\}\Big|_{\boldsymbol{\theta}_{pt}} &= -\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\right\}\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\right\}\\ &+ \mathbb{E}_f\left\{\boldsymbol{y}^H\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\boldsymbol{y}\right\}.\end{aligned} \tag{60}$$

The matrix $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ is particularly derived from $\mathbb{E}_f\left\{\frac{\partial\ell}{\partial\theta_i^*}\frac{\partial\ell}{\partial\theta_j}\right\}\Big|_{\boldsymbol{\theta}_{pt}}$

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}(i,j) &= \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\boldsymbol{C}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\boldsymbol{C}\right\}\\ &+ \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}.\end{aligned} \tag{61}$$

The second derivative of $\ell(\boldsymbol{y}|\boldsymbol{\theta})$ is

$$\begin{aligned}\frac{\partial^2\ell}{\partial\theta_j\partial\theta_i^*} &= \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*} - \boldsymbol{R}^{-1}\frac{\partial^2\boldsymbol{R}}{\partial\theta_j\partial\theta_i^*}\right\}\\ &+ \mathrm{tr}\left\{\left(\boldsymbol{R}^{-1}\frac{\partial^2\boldsymbol{R}}{\partial\theta_j\partial\theta_i^*}\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\right.\right.\\ &\left.\left. - \boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\right)\boldsymbol{y}\boldsymbol{y}^H\right\}.\end{aligned} \tag{62}$$

Taking the expectation of (62) over the true distribution f results in

$$\begin{aligned}\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}(i,j) &= -\mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\}\\ &+ \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial^2\boldsymbol{R}}{\partial\theta_j\partial\theta_i^*}\Big(\boldsymbol{R}^{-1}\boldsymbol{C}-\boldsymbol{I}\Big)\right\} - \mathrm{tr}\left\{\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_i^*}\boldsymbol{R}^{-1}\frac{\partial\boldsymbol{R}}{\partial\theta_j}\boldsymbol{R}^{-1}\boldsymbol{C}\right\}.\end{aligned} \tag{63}$$

Remark 4. The MCRB for misspecified Gaussian models is a special case of Slepian-Bangs-type formulas for Complex Elliptically Symmetric distributions under model misspecification, we refer the reader to [72] for more details.

VI. Case Studies

In this section, we investigate two cases of the channel order misspecification (overspecification and underspecification) to demonstrate the effectiveness of the proposed GMCRBs over the usual CRBs. A direct comparison between GMCRB and CRB can be done if and only if the assumed model and true model share the same parameter space. It is known that the space of signal parameters is not changed in spite of the channel order misspecification. We here conduct "indirect" comparisons of GMCRB against CRBs w.r.t. both $\tilde{\boldsymbol{h}}$ and $\boldsymbol{x}$ in which the CRBs are derived directly from the assumed model without being aware of the misspecification. Moreover, we show that the proposed MCRBs reduce to the usual CRBs when the channel order is accurately specified, i.e., the system model is correctly specified.

A. Overspecification ($L_{pt} > L_{tr}$)

When the channel order is overspecified, we include some irrelevant parameters in the system model. Specifically, the users estimate the channel using the assumed model

$$y = \tilde{\mathcal{X}}\tilde{h} + n = \left[\mathcal{X}_{\text{add}}\ \mathcal{X}\right]\begin{bmatrix} h_{\text{add}} \\ h \end{bmatrix} + n, \quad (64)$$

while the original model is

$$y = \mathcal{X}h + n = \left[\mathcal{X}_{\text{add}}\ \mathcal{X}\right]\begin{bmatrix} \mathbf{0} \\ h \end{bmatrix} + n. \quad (65)$$

Clearly, the model of (65) can be seen as a special case of (64) with $h_{\text{add}} = \mathbf{0}$, so this overspecification is not really a case of misspecification.

Under the assumption that the noise $n \sim \mathcal{CN}(0, \sigma_n^2 I_{N_r N})$ is independent of the data x (fixed or stochastic), it is easy to verify the homoscedasticity, independence and normality of the linear model (64):

- $\mathbb{E}_f\{n|x\} = \mathbf{0}$, $\mathbb{E}_f\{n[i]n[j]^\top|x\} = \mathbf{0}\ \forall i \neq j$,
- $\text{var}_f\{n|x\} = \sigma_n^2 I_{N_r N}$,
- $n|x \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 I_{N_r N})$.

Accordingly, the MLE can be derived from the ordinary least-square estimator (OLS) which is the best linear unbiased estimator of (64) [73, Theorems 12.3f-g]. In particular, the OLS/MLE of $\tilde{h}$ conditional on x can be given by

$$\widehat{\tilde{h}}(y) = \left[\mathcal{X}_{\text{add}}\ \mathcal{X}\right]^{\#}\left(\left[\mathcal{X}_{\text{add}}\ \mathcal{X}\right]\begin{bmatrix} \mathbf{0} \\ h \end{bmatrix} + n\right), \quad (66)$$

$$\mathbb{E}_f\{\widehat{\tilde{h}}(y)\} = \begin{bmatrix}\mathbf{0} & h^\top\end{bmatrix}^\top = \tilde{h}_{pt}. \quad (67)$$

The estimate from (64) while (65) is true is then unbiased.

In the following, we show that the inclusion of h_{add} can lead to less accurate estimate for h. As a corollary, there is no efficient unbiased estimator from (64) achieving the lower bound provided by the classical CRB.

Proposition 4. *The inclusion of h_{add} does not lead to bias, but increases the variance and mean square error of the best linear estimator.*

Proof. Thanks to the Frisch–Waugh–Lovell theorem [74, Section 2.4], estimators of h based on the assumed model (64) are similar to the ones from the following modified model

$$Z_{\text{add}}y = Z_{\text{add}}\mathcal{X}h + Z_{\text{add}}n, \quad (68)$$

where $Z_{\text{add}} = I - P_{\text{add}}$ with $P_{\text{add}} = \mathcal{X}_{\text{add}}\left(\mathcal{X}_{\text{add}}^H\mathcal{X}_{\text{add}}\right)^{-1}\mathcal{X}_{\text{add}}^H$.

The OLS $\widehat{h}(y)$ conditional on x takes the form

$$\widehat{h}(y) = \left(\mathcal{X}^H Z_{\text{add}}\mathcal{X}\right)^{-1}\mathcal{X}^H Z_{\text{add}}y. \quad (69)$$

Taking the expectation of $\widehat{h}(y)$ results in

$$\mathbb{E}_f\{\widehat{h}(y)\} = \mathbb{E}_f\left\{\left(\mathcal{X}^H Z_{\text{add}}\mathcal{X}\right)^{-1}\mathcal{X}^H Z_{\text{add}}(\mathcal{X}h + n)\right\}$$
$$= h + \mathbb{E}_f\left\{\left(\mathcal{X}^H Z_{\text{add}}\mathcal{X}\right)^{-1}\mathcal{X}^H Z_{\text{add}}n\right\} = h, \quad (70)$$

because $\mathbb{E}_f\{n\} = \mathbf{0}$ and n is independent of x.

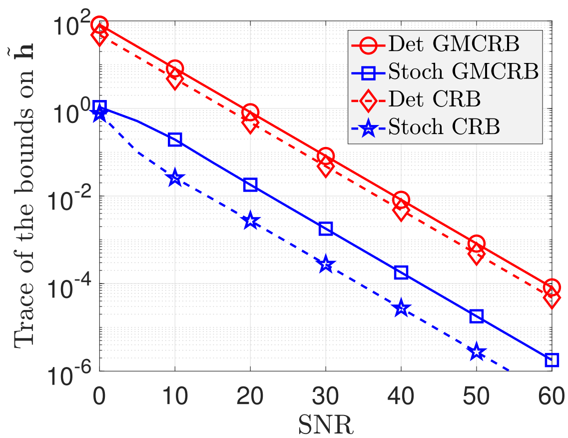


Fig. 1: Overspecification: $L_{pt} = 7 > L_{tr} = 5$.

Since $\widehat{h}(y)$ is unbiased, its mean square error (MSE) matrix is equal to its variance

$$\text{MSE}\left(\widehat{h}(y)\right) = C_f\left(\widehat{h}(y), h\right) = \mathbb{E}_f\left\{\left(\widehat{h}(y) - h\right)\left(\widehat{h}(y) - h\right)^H\right\}$$
$$= \left(\mathcal{X}^H Z_{\text{add}}\mathcal{X}\right)^{-1}\mathcal{X}^H Z_{\text{add}}\mathbb{E}_f\{nn^H\}Z_{\text{add}}\mathcal{X}(\mathcal{X}^H Z_{\text{add}}\mathcal{X})^{-1}$$
$$= \sigma_n^2\left(\mathcal{X}^H Z_{\text{add}}\mathcal{X}\right)^{-1} \geq \sigma_n^2\left(\mathcal{X}^H\mathcal{X}\right)^{-1}, \quad (71)$$

because of $Z_{\text{add}} = Z_{\text{add}}Z_{\text{add}}$, $P_{\text{add}} = P_{\text{add}}P_{\text{add}}$ and

$$\mathcal{X}^H\mathcal{X} - \mathcal{X}^H Z_{\text{add}}\mathcal{X} = \mathcal{X}^H(I - Z_{\text{add}})\mathcal{X} = \mathcal{X}^H P_{\text{add}}\mathcal{X}$$
$$= (P_{\text{add}}\mathcal{X})^H P_{\text{add}}\mathcal{X} \geq \mathbf{0}. \quad (72)$$

From the true model (65), we obtain the best unbiased estimator $\widehat{h}_{tr}(y)$ of h as

$$\widehat{h}_{tr}(y) = \left(\mathcal{X}^H\mathcal{X}\right)^{-1}\mathcal{X}^H y, \quad (73)$$

$$\text{MSE}\left(\widehat{h}_{tr}(y)\right) = \sigma_n^2\left(\mathcal{X}^H\mathcal{X}\right)^{-1}. \quad (74)$$

Accordingly, we can conclude $\text{MSE}\left(\widehat{h}(y)\right) \geq \text{MSE}\left(\widehat{h}_{tr}(y)\right)$, thanks to (71). It ends the proof. □

For illustrative purpose, we consider the following convolutive MIMO system: receive antennas $N_r = 3$, transmit antennas $N_t = 2$, the true channel order $L_{tr} = 5$ and the number of data samples $N = 50$. Particularly, the MIMO channel H is generated at random whose entries are distributed i.i.d from $\mathcal{CN}(0, 1)$, while the transmitted symbols $\{x[t]\}_{t=0}^{N-1}$ are drawn from $\mathcal{CN}(0, I_{N_t})$.

Fig. 1 plots trace of lower bounds (w.r.t. the channel parameters) versus $\text{SNR} = 10\log(1/\sigma_n^2)$ when $L_{pt} = 7 > L_{tr} = 5$. We can see that the two proposed MCRBs are higher than the classical CRBs which indicates that estimators from the true model can be more efficient than the ones from the overspecified model. The stochastic MCRB is lower than the deterministic one. Both the deterministic and stochastic MCRBs are proportional to the noise variance in a similar manner to CRBs. Indeed, the pseudo-true channel is estimated as $h_{pt} = [\mathbf{0}, h^\top]^\top$ and hence the mean and covariance are correctly specified even if the number of parameters of interest is ill-determined. Intuitively, deterministic bounds are weaker than stochastic ones in regular problems. However, it is difficult to draw conclusions about the relation when the estimation problem is singular. The experimental result shown in Fig. 1 provides an example to indicate that the stochastic Gaussian bounds may be lower than the deterministic ones.

B. Underspecification ($L_{pt} < L_{tr}$)

Due to the underspecification of the channel order, we exclude some channel parameters that actually appear in the system model. Without loss of generality, the true model of (31) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{\mathcal{X}}\boldsymbol{h} + \boldsymbol{n} = \left[\boldsymbol{\mathcal{X}}_{\texttt{omit}}\ \tilde{\boldsymbol{\mathcal{X}}}\right]\begin{bmatrix}\boldsymbol{h}_{\texttt{omit}} \\ \tilde{\boldsymbol{h}}\end{bmatrix} + \boldsymbol{n}. \tag{75}$$

Instead of (75), the users estimate the assumed model

$$\boldsymbol{y} = \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}} + \boldsymbol{n}. \tag{76}$$

Unlike the order overspecification, the underspecification will introduce bias in estimators. Moreover, it leads to many more problems than the overspecification which may give rise to the inefficiency of channel estimation only.

Proposition 5. *Normally, the exclusion of $\boldsymbol{h}_{omit}$ leads to biased estimates.*

Proof. The estimator $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ of $\tilde{\boldsymbol{h}}$ from the assumed model (76) is given by

$$\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y}) = \left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{y}. \tag{77}$$

Taking the expectation of $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ results in

$$\begin{aligned}\mathbb{E}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} &= \mathbb{E}_f\left\{\left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\left(\tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{h}} + \boldsymbol{\mathcal{X}}_{\texttt{omit}}\boldsymbol{h}_{\texttt{omit}} + \boldsymbol{n}\right)\right\} \\ &= \tilde{\boldsymbol{h}} + \mathbb{E}_f\left\{\tilde{\boldsymbol{\mathcal{X}}}^{\#}\boldsymbol{\mathcal{X}}_{\texttt{omit}}\right\}\boldsymbol{h}_{\texttt{omit}} \triangleq \tilde{\boldsymbol{h}} + \boldsymbol{b}. \end{aligned} \tag{78}$$

Accordingly, $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ is biased. The bias $\boldsymbol{b}$ will disappear if $\boldsymbol{h}_{\texttt{omit}} = \boldsymbol{0}$ or $P\left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{\mathcal{X}}_{\texttt{omit}} = \boldsymbol{0}\right) = 1$ (i.e., $\tilde{\boldsymbol{\mathcal{X}}}$ and $\boldsymbol{\mathcal{X}}_{\texttt{omit}}$ are orthogonal) which is rarely common practice. □

Since $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ is biased, the usual CRBs do not apply directly. In this case, we consider the mean square error (MSE) of $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ which is given by

$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} = \mathbb{E}_f\left\{\left(\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y}) - \tilde{\boldsymbol{h}}\right)\left(\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y}) - \tilde{\boldsymbol{h}}\right)^H\right\}, \tag{79}$$

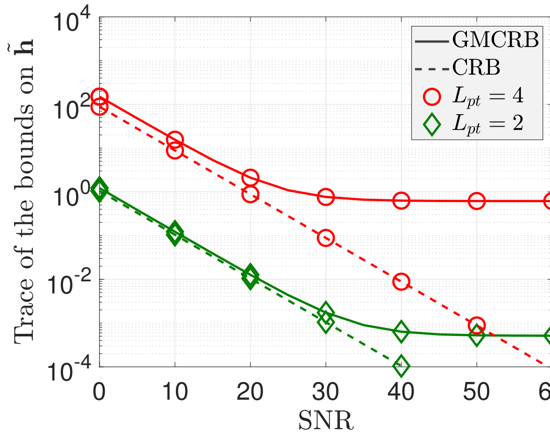
$$\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y}) - \tilde{\boldsymbol{h}} = \boldsymbol{b} + \tilde{\boldsymbol{\mathcal{X}}}^{\#}\boldsymbol{n}. \tag{80}$$

<u>*Case 1: $\boldsymbol{x}$ is deterministic*</u>*:* When the data are deterministic, the MSE of $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ is given by

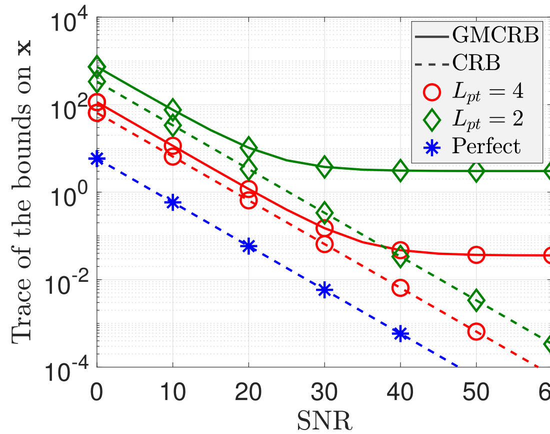
$$\begin{aligned}\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} &= \left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\mathbb{E}_f\left\{\boldsymbol{n}\boldsymbol{n}^H\right\}\tilde{\boldsymbol{\mathcal{X}}}\left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1} + \boldsymbol{b}\boldsymbol{b}^H \\ &= \sigma_n^2\left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1} + \boldsymbol{b}\boldsymbol{b}^H. \end{aligned} \tag{81}$$

Applying the same arguments in (68), the best linear unbiased estimator of $\tilde{\boldsymbol{h}}$ from the true model of (75) is given by

$$\begin{aligned}\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y}) &= \left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\texttt{omit}}\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\texttt{omit}}\boldsymbol{y} \\ &= \tilde{\boldsymbol{h}} + \left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\texttt{omit}}\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\texttt{omit}}\boldsymbol{n}, \end{aligned} \tag{82}$$



(a) Trace of the bounds on $\tilde{\boldsymbol{h}}$.



(b) Trace of the bounds on $\boldsymbol{x}$.

Fig. 2: Underspecification: Deterministic model with the "flat" channel $\boldsymbol{H}$ of $L_{tr} = 5$. The blue dashed line denotes the trace of CRB w.r.t. $\boldsymbol{x}$ when $L_{pt} = L_{tr} = 5$ (i.e. perfect specification).

where $\boldsymbol{Z}_{\texttt{omit}} = \boldsymbol{I} - \boldsymbol{\mathcal{X}}_{\texttt{omit}}\left(\boldsymbol{\mathcal{X}}_{\texttt{omit}}^H\boldsymbol{\mathcal{X}}_{\texttt{omit}}\right)^{-1}\boldsymbol{\mathcal{X}}_{\texttt{omit}}^H$. The MSE of $\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})$ is then determined as

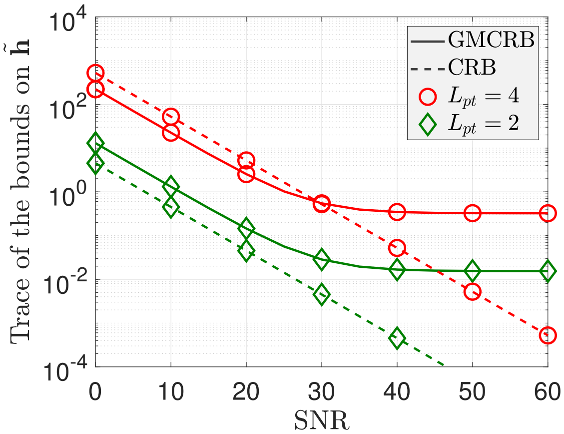
$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\} = \sigma_n^2\left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\texttt{omit}}\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}. \tag{83}$$

Next, we may want to compare $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\}$ in (81) with $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\}$ in (83). Thanks to (72), (83) is always greater than the first term of (81). Therefore, if the bias is small enough, it will lead to $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} \prec \mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\}$. On the other hand, if the bias $\boldsymbol{b}$ is large, the resulting $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\}$ may be greater than $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\}$.

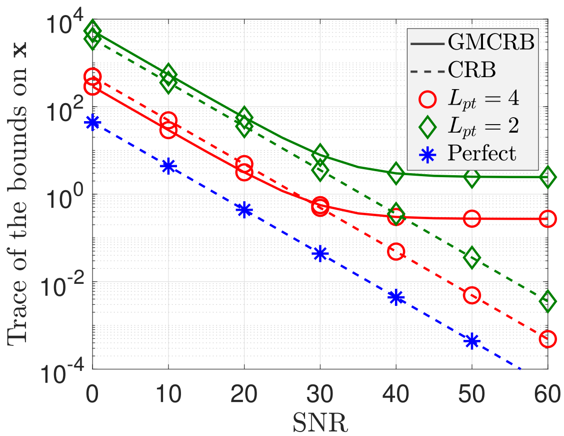
In order to illustrate this point, we here reuse the system model in Section VI-A, but consider two types of channels: (i) the "flat" channel, and (ii) the "exponential decay" channel. They are respectively defined as

$$\begin{cases}\text{flat}: & \boldsymbol{H}[\ell] = \boldsymbol{C}[\ell], \\ \text{decay}: & \boldsymbol{H}[\ell] = e^{-\ell/10}\boldsymbol{C}[\ell],\end{cases} \ell = 0, 1, \ldots, L_{tr} - 1,$$

where $\boldsymbol{C}[\ell] \in \mathbb{C}^{N_r \times N_r}$ is a "fixed" or deterministic matrix whose entries are generated randomly from $\mathcal{CN}(0, 1)$.



(a) Trace of the bounds on $\tilde{\boldsymbol{h}}$.



(b) Trace of the bounds on $\boldsymbol{x}$.

Fig. 3: Underspecification: Deterministic model with the "exponential decay" channel $\boldsymbol{H}$ of $L_{tr} = 5$. The blue dashed line denotes the trace of CRB w.r.t. $\boldsymbol{x}$ when $L_{pt} = L_{tr} = 5$ (i.e. perfect specification).

The experimental results are shown in Fig. 2 and Fig. 3. When the channel is flat, i.e., all components $\{\boldsymbol{H}[\ell]\}_{\ell=0}^{L_{tr}-1}$ share the same relative importance to $\boldsymbol{y}$, the exclusion of some $\boldsymbol{H}[\ell]$ may result in a large bias $\boldsymbol{b}$. Then this may lead to that $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} > \mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\}$, and the classical CRB is lower than the proposed GMCRB, as seen in Fig. 2(a). In this case, estimators derived from the underspecified model are less efficient than the ones from the true model. Note that, the trace of GMCRBs is plotted w.r.t. the channel parameter $\tilde{\boldsymbol{h}}$; the lower bound simply means that we are dealing with the model with the less number of parameters of interest. We know that our underlying problem considers the joint estimation of channel parameters and data symbols. To have a fair comparison, we therefore compare these bounds w.r.t. the input parameter $\boldsymbol{x}$ instead of $\tilde{\boldsymbol{h}}$ because $\boldsymbol{x}$ is correctly specified. As can be seen from Fig. 2(b) that the lower the misspecified channel order is, the less accuracy estimators can attain.

Fig. 3 shows that when we deal with the exponential decay channel, our GMCRB is, however, lower than the classical CRB at low SNRs. Probably because the output $\boldsymbol{y}$ is strongly affected by the first terms of $\boldsymbol{H}$, so the the

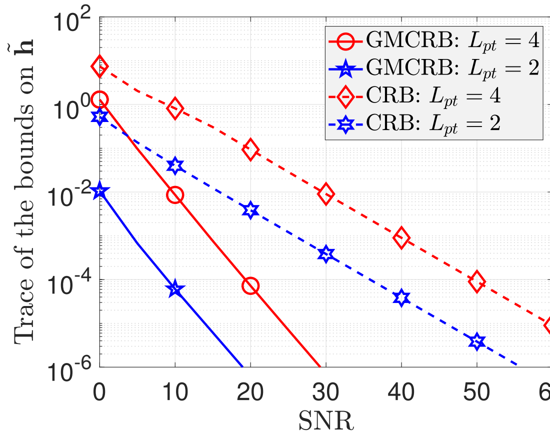


Fig. 4: Underspecification: Stochastic model with $L_{tr} = 5$.

resulting bias is small. At high SNRs, the classical CRB is much lower than GMCRB. The proposed GMCRBs tend to converge towards an error level as SNR increases. Since the error mean $\boldsymbol{e} = \boldsymbol{\mu} - \tilde{\boldsymbol{\mathcal{X}}}_{pt}\tilde{\boldsymbol{h}}_{pt}$ is independent of the noise and hence $\sigma_{pt}^2 \approx \|\boldsymbol{e}\|_2^2/(N_r N) \gg \sigma_n^2$ at high SNRs, while the deterministic GMCRB is proportional to σ_{pt}^2. Similar to the system model with the "flat" channel, the data estimation accuracy degrades gracefully when the underspecified channel order decreases, as seen in Fig. 3(b).

Case 2: $\boldsymbol{x}$ is stochastic: In this case, the MSE of $\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})$ and $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ are given by

$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\} = \mathbb{E}_f\left\{\boldsymbol{G}\boldsymbol{n}\boldsymbol{n}^H\boldsymbol{G}^H\right\}, \tag{84}$$

$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} = \mathbb{E}_f\left\{\boldsymbol{b}\boldsymbol{b}^H\right\} + \mathbb{E}_f\left\{\tilde{\boldsymbol{\mathcal{X}}}^{\#}\boldsymbol{n}\boldsymbol{n}^H(\tilde{\boldsymbol{\mathcal{X}}}^{\#})^H\right\}, \tag{85}$$

where $\boldsymbol{G} = \left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\mathtt{omit}}\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\mathtt{omit}}$.

We observe that

$$\begin{aligned}\mathbb{E}_f\left\{\boldsymbol{V}\boldsymbol{n}\boldsymbol{n}^H\boldsymbol{V}^H\right\} &= \mathbb{E}_f\left[\mathbb{E}_f\left\{\boldsymbol{V}\boldsymbol{n}\boldsymbol{n}^H\boldsymbol{V}^H\middle|\boldsymbol{V}\right\}\right]\\ &= \mathbb{E}_f\left[\boldsymbol{V}\mathbb{E}_f\{\boldsymbol{n}\boldsymbol{n}^H\}\boldsymbol{V}^H\middle|\boldsymbol{V}\right]\\ &= \mathbb{E}_f\left[\sigma_n^2\boldsymbol{V}\boldsymbol{V}^H\right] = \sigma_n^2\mathbb{E}\left\{\boldsymbol{V}\boldsymbol{V}^H\right\}.\end{aligned} \tag{86}$$

Accordingly, we can simplify (84) and (85) as follows

$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\} = \sigma_n^2\mathbb{E}_f\left\{\left(\tilde{\boldsymbol{\mathcal{X}}}^H\boldsymbol{Z}_{\mathtt{omit}}\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\right\}, \tag{87}$$

$$\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\} = \mathbb{E}_f\left\{\boldsymbol{b}\boldsymbol{b}^H\right\} + \sigma_n^2\mathbb{E}_f\left\{\left(\tilde{\boldsymbol{\mathcal{X}}}^H\tilde{\boldsymbol{\mathcal{X}}}\right)^{-1}\right\}. \tag{88}$$

Intuitively, we expect that the estimator $\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})$ from the true model of (75) is more efficient than $\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})$ from the misspecified model of (76). However, making an unambiguous comparison between $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}_{tr}(\boldsymbol{y})\right\}$ and $\mathrm{MSE}_f\left\{\widehat{\tilde{\boldsymbol{h}}}(\boldsymbol{y})\right\}$ is difficult, because the result depends on the bias $\boldsymbol{b}$ and the stochastic $\boldsymbol{x}$. For instance, we can see from Fig. 4 that the stochastic GMCRB is lower than the stochastic CRB in both cases of channel order misspecification. Note that this is just an example to demonstrate that the channel order underspecification may introduce bias and give rise to misleading results.

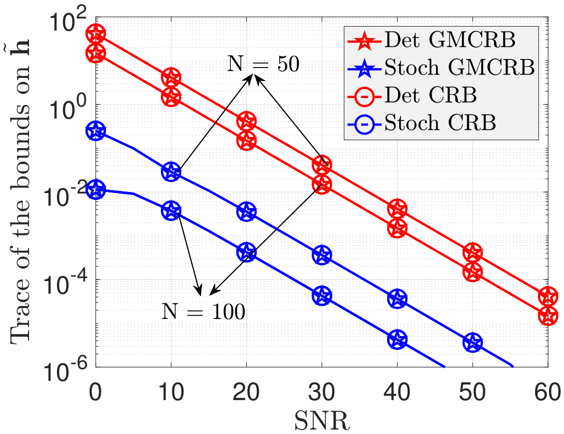


Fig. 5: Complete Specification: $L_{pt} = L_{tr} = 5$

C. Correct specification ($L_{pt} = L_{tr}$)

When the channel order is correctly specified, the proposed GMCRBs are identical to the classical CRBs, as seen in Fig. 5. In this case, the system model is correctly specified, i.e., $g(\boldsymbol{y}|\boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\phi})$, so the misspecified mean $\tilde{\boldsymbol{\mu}}$ and the misspecified covariance matrix $\boldsymbol{R}$ become the true ones. The pseudo-true $\boldsymbol{\theta}_{pt}$ is equal to the vector of true parameters $\boldsymbol{\phi}$. As a result, both two matrices $\boldsymbol{J}_{\boldsymbol{\theta}_{pt}}$ and $-\boldsymbol{A}_{\boldsymbol{\theta}_{pt}}$ reduce to the classical Fisher information matrix (FIM). Particularly, the deterministic GMCRB can be derived directly from the following FIM:

$$\mathrm{FIM}_{\boldsymbol{\theta}} = \frac{1}{\sigma_n^2}\begin{bmatrix} \boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{H}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{X}}^\top\boldsymbol{\mathcal{X}}^* & \boldsymbol{\mathcal{X}}^\top\boldsymbol{\mathcal{H}}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{X}}^* & \boldsymbol{\mathcal{H}}^\top\boldsymbol{\mathcal{H}}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{N_r N}{\sigma_n^2} \end{bmatrix}, \tag{89}$$

while the stochastic GMCRB turns out to be the stochastic CRB whose FIM is defined as

$$\mathrm{FIM}_{\boldsymbol{\theta}}(i,j) = \mathrm{tr}\left\{\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_i^*}\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_j}\right\}. \tag{90}$$

Remark 5. Although the analytic derivations presented in this section are simple, our results are of interest to reconfirm previous studies and draw a final conclusion: Channel order underspecification leads to more problems than the overspecification and we should always prefer overspecification to underspecification in practice.

VII. Conclusions

In this paper, we have addressed the problem of analyzing the theoretical performance limit of blind system identification techniques under the misspecification of the channel order through the lens of the MCRB. We have proposed the GMCRB – a new generalized interpretation of the MCRB – in order to deal with the nonexistence of the usual MCRB caused by the inherent ambiguities in blind estimation. Two closed-form expressions of the GMCRB were presented for the class of unbiased blind estimators when unknown symbols are (i) deterministic signals and (ii) stochastic signals. When the channel order is overspecified, the inclusion of some irrelevant parameters does not introduce bias but decreases the efficiency of estimation. By contrast, the order underspecification leads to bias and hence several misleading conclusions. Future works will derive the MCRBs for semi-blind system identification.

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] Y. Zeng and T. S. Ng, "Blind estimation of MIMO channels with an upper bound for channel orders," *Signal Process.*, vol. 86, no. 2, pp. 212–222, 2006.

[3] A. Tomasoni, D. Gatti, S. Bellini, M. Ferrari, and M. Siti, "Efficient OFDM channel estimation via an information criterion," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 3, pp. 1352–1362, 2013.

[4] T.-T. Lin and F.-H. Hwang, "On the CFO/channel estimation technique for MIMO-OFDM systems without using a prior knowledge of channel length," *EURASIP J. Wirel. Commun. Netw.*, vol. 2015, no. 1, pp. 1–13, 2015.

[5] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 694–705, 1997.

[6] Zhi Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3053–3061, 1997.

[7] H. Gazzah, P. A. Regalia, J. . Delmas, and K. Abed-Meraim, "A blind multichannel identification algorithm robust to order overestimation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1449–1458, 2002.

[8] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 499–511, 1997.

[9] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.

[12] 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.

[13] S. Kay, "Conditional model order estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1910–1917, 2001.

[14] J. Via, I. Santamaria, and J. Perez, "Effective channel order estimation based on combined identification/equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3518–3526, 2006.

[15] 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.

[16] 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.

[17] Y. Wang and L. Jin, "Blind order estimation based on subspace identification and weighted least squares equalization," in *IEEE Veh. Techn. Conf.*, 2015, pp. 1–5.

[18] 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.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[20] 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.

[21] E. de Carvalho, J. Cioffi, and D. T. M. Slock, "Cramer-Rao bounds for blind multichannel estimation," in *IEEE Global Telecommun. Conf.*, 2000, pp. 1036–1040.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] Q. H. Vuong, "Cramer-Rao bounds for misspecified models," *Caltech*, 1986.

[28] 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.

[29] 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.

[30] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, pp. 1–25, 1982.

[31] S. Fortunati, F. Gini, and M. S. Greco, "The misspecified Cramer-Rao bound and its application to scatter matrix estimation in complex elliptically symmetric distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2387–2399, 2016.

[32] ——, "The constrained misspecified Cramér–Rao bound," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 718–721, 2016.

[33] S. Fortunati, "Misspecified Cramér-Rao bounds for complex unconstrained and constrained parameters," in *European Signal Process. Conf.*, 2017, pp. 1644–1648.

[34] M. Pajovic, "Misspecified Bayesian Cramér-Rao bound for sparse Bayesian," in *IEEE Stat. Signal Process. Works.*, 2018, pp. 263–267.

[35] F. Roemer, "Misspecified Cramer-rao bound for delay estimation with a mismatched waveform: A case study," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5994–5998.

[36] F. Elvander, J. Ding, and A. Jakobsson, "On harmonic approximations of inharmonic signals," in *IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5360–5364.

[37] M. P. Little, W. F. Heidenreich, and G. Li, "Parameter identifiability and redundancy: theoretical considerations," *PLOS One*, vol. 5, no. 1, p. e8915, 2010.

[38] H. Wei and S.-i. Amari, "Dynamics of learning near singularities in radial basis function networks," *Neural Netw.*, vol. 21, no. 7, pp. 989–1005, 2008.

[39] W. Guo, Y. Ong, Y. Zhou, J. R. Hervas, A. Song, and H. Wei, "Fisher information matrix of unipolar activation function-based multilayer perceptrons," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 3088–3098, 2019.

[40] H. Sedghi, V. Gupta, and P. M. Long, "The singular values of convolutional layers," in *Int. Conf. Learn. Repres.*, 2019, pp. 1–12.

[41] S. Audoly, G. Bellu, L. D'Angio, M. P. Saccomani, and C. Cobelli, "Global identifiability of nonlinear models of biological systems," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 55–65, 2001.

[42] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angiò, "DAISY: A new software tool to test global identifiability of biological and physiological systems," *Comput. Methods Programs Biomed.*, vol. 88, no. 1, pp. 52–61, 2007.

[43] M. P. Little, W. F. Heidenreich, and G. Li, "Parameter identifiability and redundancy in a general class of stochastic carcinogenesis models," *PLOS One*, vol. 4, no. 12, p. e8520, 2009.

[44] C. Shin, R. W. Heath, and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, 2007.

[45] V. Choqueuse, A. Mansour, G. Burel, L. Collin, and K. Yao, "Blind channel estimation for STBC systems using higher-order statistics," *IEEE Trans. Wirel. Commun.*, vol. 10, no. 2, pp. 495–505, 2010.

[46] J. Zhang, X. Yuan, and Y. A. Zhang, "Blind signal detection in massive MIMO: Exploiting the channel sparsity," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 700–712, 2018.

[47] M. Drton, M. Plummer *et al.*, "A Bayesian information criterion for singular models," *J. R. Statist. Soc. B*, vol. 79, no. 2, pp. 323–380, 2017.

[48] S. Audoly, L. D'Angio, M. P. Saccomani, and C. Cobelli, "Global identifiability of linear compartmental models - a computer algebra algorithm," *IEEE. Trans. Biomed. Eng.*, vol. 45, no. 1, pp. 36–47, 1998.

[49] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 133–136, 2008.

[50] Y. C. Eldar *et al.*, "Rethinking biased estimation: Improving maximum likelihood and the Cramér–rao bound," *Found. Trends Signal Process.*, vol. 1, no. 4, pp. 305–449, 2008.

[51] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, 2001.

[52] A. Somekh-Baruch, A. Leshem, and V. Saligrama, "On the non-existence of unbiased estimators in constrained estimation problems," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5549–5554, 2018.

[53] E. Nitzan, T. Routtenberg, and J. Tabrikian, "Cramér–Rao bound for constrained parameter estimation using Lehmann-unbiasedness," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 753–768, 2019.

[54] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2026–2041, 1996.

[55] Y. Li and P. Yeh, "An interpretation of the Moore-Penrose generalized inverse of a singular Fisher information matrix," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5532–5536, 2012.

[56] P. Stoica and Boon Chong Ng, "On the Cramer-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[57] A. K. Jagannatham and B. D. Rao, "Cramer-Rao lower bound for constrained complex parameters," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 875–878, 2004.

[58] L. T. Thanh, K. Abed-Meraim, and N. L. Trung, "Performance lower bounds of blind system identification techniques in the presence of channel order estimation error," in *Eur. Signal Process. Conf.*, 2021, pp. 1646–1650.

[59] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Rev.*, vol. 57, no. 2, pp. 225–251, 2015.

[60] U. Baysal and R. L. Moses, "On the geometry of isotropic arrays," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1469–1478, 2003.

[61] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, 2012.

[62] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2019.

[63] V. Simoncini, "Computational methods for linear matrix equations," *SIAM Rev.*, vol. 58, no. 3, pp. 377–441, 2016.

[64] T. Stykel, "Analysis and numerical solution of generalized lyapunov equations," Ph.D. dissertation, Institut für Mathematik, Technische Universität, Berlin, 2002.

[65] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[66] A. M. Farah, "Generalized and quadratic eigenvalue problems with Hermitian matrices." Ph.D. dissertation, University of Birmingham, 2012.

[67] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[68] 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.

[69] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, 1996.

[70] F. Benjamin and A. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1557–1567, 1995.

[71] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, 2016.

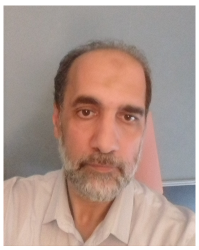
[72] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, and A. Renaux, "Slepian-Bangs-type formulas and the related misspecified Cramér-Rao bounds for complex elliptically symmetric distributions," *Signal Process.*, vol. 142, pp. 320–329, 2018.

[73] A. C. Rencher and G. B. Schaalje, *Linear models in statistics*. John Wiley & Sons, 2008.

[74] R. Davidson and J. G. MacKinnon, *Econometric theory and methods*. Oxford University Press, 2004.

Le Trung Thanh received B.Sc. and M.Sc. degrees in Electronics and Telecommunications from VNU University of Engineering and Technology, Vietnam National University, Hanoi (VNU) in 2016 and 2018 respectively. He is now pursuing his Ph.D. study at the University of Orleans, France.

His research interests include signal processing, subspace tracking, tensor analysis, and system identification.



Karim Abed-Meraim (Fellow, IEEE) was born in 1967. He received the State Engineering Degree from the École Polytechnique, Palaiseau, France, in 1990, the State Engineering Degree from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1992, the M.Sc. degree from Paris XI University, Orsay, France, in 1992, and the Ph.D. degree in the field of signal processing and communications from ENST, in 1995.

From 1995 to 1998, he was a Research Staff with the Electrical Engineering Department, The University of Melbourne, where he worked on several research project related to “Blind System Identification for Wireless Communications”, “Blind Source Separation,” and “Array Processing for Communications.” From 1998 to 2012, he has been an Assistant Professor, then an Associate Professor with the Signal and Image Processing Department, Télécom ParisTech. In September 2012, he joined the University of Orléans, France (PRISME Laboratory), as a Full Professor. His research interests include signal processing for communications, adaptive filtering and tracking, array processing, and statistical performance analysis. He is the author of about 500 scientific publications, including book chapters, international journal and conference papers, and patents.

Dr. Abed-Meraim is currently a member of the IEEE SAM-TC and a Senior Area Editor of the IEEE Transactions on Signal Processing.

Nguyen Linh Trung (Senior Member, IEEE) received the B.Eng. and Ph.D. degrees in electrical engineering from the Queensland University of Technology, Brisbane, QLD, Australia, in 1998 and 2005, respectively.

Since 2006, he has been with the Faculty of VNU University of Engineering and Technology, Vietnam National University, Hanoi (VNU), where he is currently an Associate Professor of electronic engineering with the Faculty of Electronics and Telecommunications and the Director of the Advanced Institute of Engineering and Technology. His research focuses on signal processing methods, including time-frequency signal analysis, blind processing, adaptive filtering, compressive sampling, tensor-based signal analysis, graph signal processing, and apply them to wireless communication, networking, and biomedical engineering.