An Interpretation of the Moore-Penrose Generalized Inverse of a Singular Fisher Information Matrix

Yen-Huan Li
Graduate Institute of Communication Engineering
National Taiwan University
Taipei, Taiwan
Email: r97942042@ntu.edu.tw

Ping-Cheng Yeh
Graduate Institute of Communication Engineering and
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan
Email: pcyeh@cc.ee.ntu.edu.tw

***Abstract*—An interpretation of the Moore-Penrose generalized inverse of a singular Fisher information matrix (FIM) is presented in this paper, from the perspective of Cramér-Rao bound (CRB). CRB is a lower bound on the variance of unbiased estimators, and can be calculated as the Moore-Penrose generalized inverse of the FIM [1]. There are some existing facts, however, that render the application of CRB questionable when the FIM is singular. For example, Soica *et al.* shown that unbiased estimators with finite variances do not exist if the FIM is singular [2]. The application of CRB with singular FIMs, therefore, seems meaningless. Noting that unbiased estimators with finite variances may exist if some deterministic constraints are put on the unknown parameter, we solve this confusing situation by showing that the Moore-Penrose generalized inverse of a singular FIM is *the CRB corresponding to the minimum variance among all choices of minimum constraint functions*. Our result not only provides a way to interpret the CRB obtained as the Moore-Penrose generalized inverse of a singular FIM, but also enables future research on the joint design of constraint functions and unbiased estimators.**

***Index Terms*—Cramér-Rao bound (CRB), constrained parameters, singular Fisher information matrix (FIM).**

I. Introduction

An interpretation of the Moore-Penrose generalized inverse of a singular information matrix is presented in this paper, from the perspective of Cramér-Rao bound (CRB). CRB is a lower bound on the covariance matrices of estimators in a parametric estimation problem. The most general form of CRB says that the covariance matrix of any unbiased estimator is larger than the generalized inverse of the Fisher information matrix (FIM) under Löwner partial order [1]. This general form of CRB holds for parametric estimation problems with both singular and non-singular FIMs.

There are, however, facts in literature which renders the application of CRB questionable when the FIM is singular. Rothenberg proves in [3] that under some regularity conditions, the non-singularity of the FIM is equivalent to the local identifiability of the parameter to be estimated[1]; Stoica *et al.* prove in [2] that unbiased estimators with finite variances do not exist when the FIM is singular, except for some "unusual" conditions[2]. If the parameter to be estimated is locally non-identifiable, or all of the unbiased estimators will have infinite variances, it is meaningless to discuss the performance of unbiased estimators.

As mentioned in [2], one may change the nature of an estimation problem to allow the existence of reasonable estimators. The first approach is to introduce *a priori* information about the probability distribution of the parameter to be estimated; in this way the estimation problem becomes a Bayesian one. There are abundant literature on Bayesian approach [4], but we should keep in mind that *a priori* information is not always already known. The second approach is to consider *biased* estimators instead of *unbiased* estimators. This approach is discussed in [2], where the necessary condition for the bias function to ensure the existence of reasonable estimators is derived. There are a number of situations, however, where biased estimators are not allowed. For example, almost all estimation problems encountered in the design of a communication system, including the estimation of carrier phase and symbol timing for synchronization, the estimation of channel response for equalization, etc., require unbiased estimators. The third approach is to put some deterministic constraints on the parameter to be estimated. The deterministic constraints result in a parametric estimation problem with reduced dimension, where reasonable unbiased estimators may exist. This paper focuses on the third approach.

We believe that the third approach has practical significance. Take blind channel estimation problems for example. The goal of blind channel estimation is to estimate the channel response $\boldsymbol{h}$ from $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h} + \boldsymbol{n}$, the convolution of the channel response $\boldsymbol{h}$ and the input data sequence $\boldsymbol{s}$ corrupted by the additive noise $\boldsymbol{n}$. The unknown parameter $\boldsymbol{\theta} \triangleq (\boldsymbol{s}, \boldsymbol{h})$ is not identifiable since $(\alpha \boldsymbol{s}, \frac{1}{\alpha}\boldsymbol{h})$ and $(\boldsymbol{s}, \boldsymbol{h})$ are observationally equivalent for any constant $\alpha \neq 0$, so unbiased estimators do not exist. Practically this so-called scalar ambiguity problem is resolved by assigning a pre-determined value to one of the element of $\boldsymbol{s}$ [5]. That is, a constraint function $f(\boldsymbol{\theta}) \triangleq s_n - c = 0$ is put on the parameter $\boldsymbol{\theta}$, where s_n denotes the nth element of $\boldsymbol{s}$ and c is some pre-determined constant. This is exactly the

[1] A parameter $\boldsymbol{\theta}$ is locally identifiable if there exists an open neighbourhood $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}$ such that no other $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$ is observationally equivalent to $\boldsymbol{\theta}$.

[2] More accurately, the "unusual" conditions suggest that if the FIM is singular, only unbiased estimators for some functions of the unknown parameter, instead of the unknown parameter itself, may exist with finite variances.

third approach mentioned above.

CRB with a parameter constraint, or constrained CRB, is already derived in [6], [7], [8]. The value of the constrained CRB depends on the choice of the constraint function; different constraint functions lead to different values of the CRB. This bound is useful when the constraint function is exogenously given, but there are situations where we are able to modify the constraint function. Take blind channel estimation problems for example again. Suppose an engineer chooses the constraint function as $f(\boldsymbol{\theta}) \triangleq s_1 - c = 0$ and designs an unbiased estimator corresponding to this constraint function, and finds the resulted mean squared error (MSE), although almost achieving the constrained CRB, is still unsatisfactory compared with the target value. How can the engineer tell the unsatisfactory result is caused by the inappropriate choice of the constraint function, or simply because the target value is not attainable for *any* choice of the constraint function? We need a bound for the joint design of the unbiased estimator and the constraint function.

The main contribution of this paper is the following theorem: The Moore-Penrose generalized inverse of a singular FIM is *the CRB corresponding to the minimum variance among all choices of minimum constraint functions*. According to the theorem, a meaningful interpretation of the CRB obtained as the Moore-Penrose generalized inverse of a singular FIM is presented, and a CRB for the joint design of the unbiased estimator and the constraint function is obtained. In additional to a performance bound, we also derive the sufficient condition for a constraint function to achieve the bound, and show that this bound can always be achieved by a linear constraint function, which facilitates the optimal design of minimum constraint functions.

A mathematical definition of minimum constraint functions will be given in Section IV-A, but the meaning is conceptually easy to understand. In blind channel estimation problems, only a one-dimensional constraint on $\boldsymbol{\theta}$ is needed to resolve the scalar ambiguity, such as $f(\boldsymbol{\theta}) = s_n - 1$, any constraint function $\boldsymbol{f}$ that is essentially a one-dimensional constraint is a minimum constraint function as long as the constrained CRB exists.

The rest of the paper is organized as follows. The necessary background knowledge is given in II. Then we show that the Moore-Penrose generalized inverse [9] of the FIM can be viewed as a CRB for constrained parameters with some constraint function in Section III. Finally we prove the main result of this paper, that the Moore-Penrose generalized inverse of the FIM is the CRB corresponding to the minimum variance among all choices of constraint functions in Section IV. Conclusions are presented in Section V.

Notation

Bold-faced lower case letters represent column vectors, and bold-faced upper case letters are matrices. Superscripts such as $\boldsymbol{v}^*$, $\boldsymbol{v}^T$, $\boldsymbol{M}^{-1}$, and $\boldsymbol{M}^\dagger$ denote the conjugate, transpose, inverse, and the Moore-Penrose generalized inverse of the corresponding vector or matrix. The matrix $\mathsf{diag}(\boldsymbol{\theta})$ means the matrix whose diagonal elements are elements of the vector $\boldsymbol{\theta}$. The vector $\mathsf{E}\left[\boldsymbol{v}\right]$ denotes the expectation of $\boldsymbol{v}$. The matrix $\mathsf{cov}(\boldsymbol{u}, \boldsymbol{v})$ is defined as $\mathsf{cov}(\boldsymbol{u}, \boldsymbol{v}) \triangleq \mathsf{E}\left[(\boldsymbol{u} - \mathsf{E}(\boldsymbol{u}))(\boldsymbol{v} - \mathsf{E}(\boldsymbol{v}))^T\right]$, which is the cross-covariance matrix of random vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. We use the notation $\boldsymbol{A} \geq \boldsymbol{B}$ to mean that $\boldsymbol{A} - \boldsymbol{B}$ is a nonnegative-definite matrix.

II. Preliminaries

In this section, some background knowledge required to begin the discussions in the following sections are presented. We restrict our attention to the case of *unbiased estimators for the unknown parameter*, so the theorems presented in this section may be a simplified version of the original one.

When we refer to the *CRB for unconstrained parameters*, we mean the following theorem.

Theorem II.1 (CRB for unconstrained parameters)**.** *Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^n$ based on observation $\boldsymbol{y}$, which is characterized by its probability density function (pdf) $p(\boldsymbol{y}; \boldsymbol{\theta})$. Then for any such $\hat{\boldsymbol{\theta}}$, we have*

$$\mathsf{cov}\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}\right) \geq \boldsymbol{J}^\dagger, \tag{1}$$

where $\boldsymbol{J}$ is the FIM defined as

$$\boldsymbol{J} \triangleq \mathsf{E}\left[\frac{\partial p}{\partial \boldsymbol{\theta}} \frac{\partial p}{\partial \boldsymbol{\theta}^T}\right]. \tag{2}$$

Proof: See [1]. ■

The above theorem is always correct given that unbiased estimators exist. Stoica *et al.*, however, proved the following theorem in [2].

Theorem II.2. *If the information matrix $\boldsymbol{J}$ is singular, then there does not exist an unbiased estimator with finite variance.*

Proof: See [2][3]. ■

That is, there does not exist any reasonable estimator $\hat{\boldsymbol{\theta}}$ if the FIM is singular, so the CRB fails to provide any useful information.

When we refer to the *CRB for constrained parameters*, we mean the following theorem.

Theorem II.3 (CRB for constrained parameters)**.** *Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^n$ based on observation $\boldsymbol{y}$, which is characterized by its pdf $p(\boldsymbol{y}; \boldsymbol{\theta})$. Furthermore, we require the parameter $\boldsymbol{\theta}$ to satisfy a possibly nonlinear constraint function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$,*

$$\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{0}. \tag{3}$$

Assume that $\partial \boldsymbol{f} / \partial \boldsymbol{\theta}^T$ is full rank. Choose a matrix $\boldsymbol{U}$ with $(n - m)$ orthonormal columns such that

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T} \boldsymbol{U} = \boldsymbol{0}. \tag{4}$$

[3]When we restrict our attention to unbiased estimators for the unknown parameter only, the condition for the existence of an unbiased estimator with finite variance in [2] becomes $\boldsymbol{J}\boldsymbol{J}^\dagger = \boldsymbol{I}$, which is impossible for singular FIMs.

If $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is nonsingular, then

$$\mathsf{cov}\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\theta}}\right) \geq \boldsymbol{U}\left(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^T, \tag{5}$$

where $\boldsymbol{J}$ is the FIM defined as in (2).

Proof: See [6], [7], [8]. ∎

The following theorem gives a necessary sufficient condition to check the existence of a finite constrained CRB.

Theorem II.4. *A necessary and sufficient condition for the existence of a finite constrained CRB is an invertible $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$.*

Proof: See [8]. ∎

Now we are able to discuss the relationship between the Moore-Penrose generalized inverse of an FIM and constrained CRB.

III. $\boldsymbol{J}^\dagger$ as a CRB for Constrained Parameters

The main result of this section is the following theorem.

Theorem III.1. *If the FIM $\boldsymbol{J}$ is singular, and let the singular value decomposition (SVD) of $\boldsymbol{J}$ be*

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{U}_s & \boldsymbol{U}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_s^T \\ \boldsymbol{U}_n^T \end{bmatrix}, \tag{6}$$

the diagonal elements of $\boldsymbol{\Sigma}$ being nonzero, then $\boldsymbol{J}^\dagger$ is a CRB for constrained parameters with constraint function

$$\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{U}_n^T \mathsf{diag}(\boldsymbol{\theta}) + \boldsymbol{C} = \boldsymbol{0} \tag{7}$$

for some constant matrix $\boldsymbol{C}$.

To prove the theorem, we first prove the following lemma.

Lemma III.1. *Let the SVD of a hermitian matrix $\boldsymbol{J}$ be the same as in (6). Then*

$$\boldsymbol{J}^\dagger = \boldsymbol{U}_s\left(\boldsymbol{U}_s^H\boldsymbol{J}\boldsymbol{U}_s\right)^{-1}\boldsymbol{U}_{\boldsymbol{s}}^T. \tag{8}$$

Proof: Substitute $\boldsymbol{J}$ as $\boldsymbol{J} = \boldsymbol{U}_s\boldsymbol{\Sigma}\boldsymbol{U}_s^T$ into (8). ∎

Now we are able to prove Theorem III.1.

Proof for Theorem III.1: By examining the lemma and Theorem II.3, we can think of $\boldsymbol{J}^\dagger$ as a constrained CRB with some constraint function $\boldsymbol{f}(\boldsymbol{\theta})$ such that

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T}\boldsymbol{U}_s = \boldsymbol{0}. \tag{9}$$

Since $\boldsymbol{U}_n^T\boldsymbol{U}_s = \boldsymbol{0}$ by the definition of SVD, a constraint function $\boldsymbol{f}$ that satisfies (9) can be chosen such that

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T} = \boldsymbol{U}_n^T. \tag{10}$$

The above equation can be satisfied by a linear constraint function,

$$\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{U}_n^T \mathsf{diag}(\boldsymbol{\theta}) + \boldsymbol{C} = \boldsymbol{0}, \tag{11}$$

and the theorem is proved. ∎

IV. Interpretation of $\boldsymbol{J}^\dagger$ as a CRB for Constrained Parameters

In this section we prove that $\boldsymbol{J}^\dagger$ is not only a CRB for constrained parameters, but *the CRB corresponding to the minimum variance among all choices of minimum constraint functions*. We first give a definition minimum constraint functions, and then prove the claim.

A. Definition of Minimum Constraint Functions

Minimum constraint functions are defined as follows.

Definition IV.1. *A differentiable constraint function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$ for an estimation problem with singular FIM $\boldsymbol{J}$ is a minimum constraint if*

1) *$\partial\boldsymbol{f}/\partial\boldsymbol{\theta}^T$ is full rank,*
2) *$\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is nonsingular, and*
3) *$\mathsf{rank}\ \partial\boldsymbol{f}/\partial\boldsymbol{\theta}^T + \mathsf{rank}\ \boldsymbol{J} = n$,*

where $\boldsymbol{U}$ is chosen as in Theorem II.3.

The first requirement is to ensure that $\boldsymbol{f}$ does not contain any redundant constraints [6], [7]. The second requirement is to ensure the existence of CRB according to Theorem II.4. The third requirement means that $\boldsymbol{f}$ contains the minimum number of independent constraints. Take blind channel estimation problems as example. From discussions in Section I we know that once we choose one symbol as a pilot symbol with some predefined value, we eliminate the scalar ambiguity and thus an unbiased estimator exists. Note that the nullity of the FIM is also one by [10], [11]. We can see the third requirement holds.

Now we give a formal proof that if the first two requirements are satisfied, then the third requirement ensures that $\boldsymbol{f}$ contains the minimum number of independent constraints.

Theorem IV.1. *For any constraint function $\boldsymbol{f}$ in Definition IV.1 that satisfies the first and the second requirements,*

$$\min_{\boldsymbol{f}} \mathsf{rank}\ \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T} = n - \mathsf{rank}\ \boldsymbol{J}. \tag{12}$$

Proof: First we show that in order to satisfy the first and the second requirements,

$$\mathsf{rank}\ \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T} \geq n - \mathsf{rank}\ \boldsymbol{J}, \tag{13}$$

and then we show that the equality is achievable.

If

$$\mathsf{rank}\ \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}^T} < n - \mathsf{rank}\ \boldsymbol{J}, \tag{14}$$

by the definition of $\boldsymbol{U}$ (see Theorem II.3), $\boldsymbol{U}$ is a n-by-$(\mathsf{rank}\ \boldsymbol{U})$ matrix with

$$n \geq \mathsf{rank}\ \boldsymbol{U} > \mathsf{rank}\ \boldsymbol{J}. \tag{15}$$

By the fact that

$$\mathsf{rank}\ \boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{J}\boldsymbol{U} \leq \min\{\mathsf{rank}\ \boldsymbol{U}, \mathsf{rank}\ \boldsymbol{J}\} \leq \mathsf{rank}\ \boldsymbol{J} < \mathsf{rank}\ \boldsymbol{U}, \tag{16}$$

where the last inequality follows by (15), and noting that $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ is a $(\mathsf{rank}\ \boldsymbol{U})$-by-$(\mathsf{rank}\ \boldsymbol{U})$ square matrix, $\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}$ cannot be full-rank. Thus (13) is proved.

The achievability of equality in (13) is easy to prove. Choose the constraint function $\boldsymbol{f}$ as in (7), and we can see such a constraint function satisfies all of the requirements of a minimum constraint function. ■

By the above theorem we can see the third requirement is in fact requiring $\partial \boldsymbol{f}/\partial \boldsymbol{\theta}^T$ to have the minimum rank. The reason why such a constraint function $\boldsymbol{f}$ can be considered as the constraint function with *minimum constraints* can be found by the following theorem.

Theorem IV.2. *Let $A \subset \mathbb{R}^n$ be open and let $\boldsymbol{f} : A \to \mathbb{R}^m$ be a differentiable function such that $\partial \boldsymbol{f}/\partial \boldsymbol{\theta}^T$ has rank m whenever $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$. Then $\boldsymbol{f}^{-1}(\boldsymbol{0})$ defines an $(n-m)$-dimensional manifold in $\mathbb{R}^n$.*

Proof: See [12]. ■

Constraint functions $\boldsymbol{f}$ with the minimum $\mathsf{rank}\ \partial \boldsymbol{f}/\partial \boldsymbol{\theta}^T$ ensures that the resulting manifolds have the maximal degree of freedom, so we call them as constraint functions with the minimum constraints.

B. $\boldsymbol{J}^\dagger$ is the CRB corresponding to the minimum variance among all choices of minimum constraint functions

This subsection is to prove the claim that $\boldsymbol{J}^\dagger$ is the CRB corresponding to the minimum variance among all choices of minimum constraint functions. For convenience, the ith largest eigenvalue of a matrix $\boldsymbol{M}$ is denoted by $\lambda_i(\boldsymbol{M})$ in the following discussions.

The main result of this subsection is the following theorem.

Theorem IV.3. *In Theorem II.3, if $\boldsymbol{f}$ is a minimum constraint function, then*

$$\mathsf{tr}\left(\mathsf{cov}\left[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}\right]\right) \geq \mathsf{tr}\left(\boldsymbol{J}^\dagger\right). \tag{17}$$

Furthermore, equality can be achieved by choosing the constraint function $\boldsymbol{f}$ as in Theorem III.1.

Note that the trace of a covariance matrix is the sum of the variances of the elements of $\hat{\boldsymbol{\theta}}$. In this way, we have proved that the Moore-Penrose generalized inverse of the FIM is the CRB corresponding to the minimum variance among all choices of minimum constraint functions.

Theorem IV.3 is in fact a corollary of the following theorem.

Theorem IV.4. *Let the SVD of a m-by-m nonnegative definite matrix $\boldsymbol{J}$ with rank n be*

$$\boldsymbol{J} = \left[\begin{array}{cc} \boldsymbol{U}_s & \boldsymbol{U}_n \end{array}\right] \left[\begin{array}{cc} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{array}\right] \left[\begin{array}{c} \boldsymbol{U}_s^T \\ \boldsymbol{U}_n^T \end{array}\right], \tag{18}$$

where $\boldsymbol{\Sigma}$ is a n-by-n diagonal matrix. Then

$$\begin{aligned} \lambda_i(\boldsymbol{V}\left(\boldsymbol{V}^T\boldsymbol{J}\boldsymbol{V}\right)^{-1}\boldsymbol{V}^T) &\geq \lambda_i(\boldsymbol{U}_s\left(\boldsymbol{U}_s^T\boldsymbol{J}\boldsymbol{U}_s\right)^{-1}\boldsymbol{U}_s^T) \\ &= \lambda_i(\boldsymbol{J}^\dagger) \quad \forall i \end{aligned} \tag{19}$$

for any matrix $\boldsymbol{V}$ with the same size as $\boldsymbol{U}_s$ and $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$.

If the above theorem holds, then Theorem IV.3 can be proved as follows.

Proof for Theorem IV.3: Note that $\boldsymbol{J}$ is a nonnegative definite matrix, and the resulting $\boldsymbol{U}$ for every minimum constraint $\boldsymbol{f}$ should have the same size as $\boldsymbol{U_s}$ in Theorem IV.4, so the above theorem applies. Noting that $\boldsymbol{U}\left(\boldsymbol{U}^T\boldsymbol{J}\boldsymbol{U}\right)^{-1}\boldsymbol{U}^T = \boldsymbol{J}^\dagger$ according to Lemma III.1, the corollary follows because trace equals to the sum of eigenvalues. ■

See Appendix for the proof of Theorem IV.4.

V. Conclusions

We have proved the main theorem of this paper: The Moore-Penrose generalized inverse of a singular FIM is the CRB corresponding to the minimum variance among all choices of minimum constraint functions. According to the theorem, a meaningful interpretation of the Moore-Penrose generalized inverse of a singular FIM is presented, and a CRB for the joint design of the unbiased estimator and the constraint function is obtained. In additional to a performance bound, we also derive the sufficient condition for a constraint function to achieve the bound, and show that this bound can always be achieved by a linear constraint function, which facilitates the optimal design of minimum constraint functions.

Appendix

The proof is mainly based on Poincaré separation theorem and a lemma. We first show Poicaré seperation theorem below.

Theorem A.1 (Poincaré separation theorem)**.** *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a Hermitian matrix, and let $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ satisfy $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$. Define $\boldsymbol{B}_r \triangleq \boldsymbol{U}^T\boldsymbol{A}\boldsymbol{U}$. Then*

$$\lambda_i(\boldsymbol{B}_r) \leq \lambda_i(\boldsymbol{A}) \tag{20}$$

for all $k \in \{1, \ldots, r\}$.

Proof: See [9]. ■

Then we prove the following lemma.

Lemma A.1. *For any nonnegative definite matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, and any matrix $\boldsymbol{V} \in \mathbb{R}^{m\times n}$, $m \geq n$, with $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$,*

$$\lambda_i(\boldsymbol{V}\boldsymbol{M}\boldsymbol{V}^T) = \lambda_i(\boldsymbol{M}) \tag{21}$$

for all $i \in \{1, \ldots, n\}$, and

$$\lambda_i(\boldsymbol{V}\boldsymbol{M}\boldsymbol{V}^T) = 0 \tag{22}$$

for all $i \in \{n+1, \ldots, m\}$.

Proof: Define $\overline{\boldsymbol{V}} \triangleq \boldsymbol{V}$ for notational convenience. By the definition of $\overline{\boldsymbol{V}}$, there exist a matrix $\widetilde{\boldsymbol{V}}$ such that $\left[\begin{array}{cc} \overline{\boldsymbol{V}} & \widetilde{\boldsymbol{V}} \end{array}\right]$ is a unitary matrix.

Let the SVD of $\boldsymbol{M}$ be $\overline{\boldsymbol{U}}\boldsymbol{\Sigma}\overline{\boldsymbol{U}}^T$. We can construct a m-by-m unitary matrix as

$$\boldsymbol{U} \triangleq \left[\begin{array}{cc} \overline{\boldsymbol{U}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{array}\right], \tag{23}$$

and we have

$$\begin{aligned}
&\overline{\boldsymbol{V}}\boldsymbol{M}\overline{\boldsymbol{V}}^T \\
&= \begin{bmatrix} \overline{\boldsymbol{V}} & \widetilde{\boldsymbol{V}} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{U}}\boldsymbol{\Sigma}\overline{\boldsymbol{U}}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{V}}^T \\ \widetilde{\boldsymbol{V}}^T \end{bmatrix} \\
&= \begin{bmatrix} \overline{\boldsymbol{V}} & \widetilde{\boldsymbol{V}} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{U}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \\
&\quad \begin{bmatrix} \overline{\boldsymbol{U}}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}^T \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{V}}^T \\ \widetilde{\boldsymbol{V}}^T \end{bmatrix}. \qquad (24)
\end{aligned}$$

Note that the product of two unitary matrices are also a unitary matrix. Therefore (24) is the SVD of the matrix $\overline{\boldsymbol{V}}\boldsymbol{M}\overline{\boldsymbol{V}}^T$. Since $\overline{\boldsymbol{V}}\boldsymbol{M}\overline{\boldsymbol{V}}^T$ is a nonnegative definite matrix, we can infer from its SVD that its eigenvalues are $\lambda_1(\boldsymbol{M}), \ldots, \lambda_n(\boldsymbol{M})$ and $(m-n)$ zeroes, and the theorem follows. ■

Now we are able to prove Theorem IV.4.

Proof for Theorem IV.4: By Lemma A.1, we know that

$$\lambda_i \left(\boldsymbol{V} \left(\boldsymbol{V}^T \boldsymbol{J} \boldsymbol{V} \right)^{-1} \boldsymbol{V}^T \right) = \lambda_i \left(\boldsymbol{U}_s \left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right)^{-1} \boldsymbol{U}_s^T \right) = 0 \qquad (25)$$

for $i \in \{n+1, n+2, \ldots, m\}$, and

$$\lambda_i \left(\boldsymbol{V} \left(\boldsymbol{V}^T \boldsymbol{J} \boldsymbol{V} \right)^{-1} \boldsymbol{V}^T \right) = \lambda_i \left(\left(\boldsymbol{V}^T \boldsymbol{J} \boldsymbol{V} \right)^{-1} \right), \qquad (26)$$

$$\lambda_i \left(\boldsymbol{U}_s \left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right)^{-1} \boldsymbol{U}_s^T \right) = \lambda_i \left(\left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right)^{-1} \right), \qquad (27)$$

for $i \in \{1, 2, \ldots, n\}$, so it suffices to prove

$$\lambda_i \left(\left(\boldsymbol{V}^T \boldsymbol{J} \boldsymbol{V} \right)^{-1} \right) \geq \lambda_i \left(\left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right)^{-1} \right), \qquad (28)$$

for $i \in \{n+1, n+2, \ldots, m\}$, or equivalently,

$$\lambda_i \left(\boldsymbol{V}^T \boldsymbol{J} \boldsymbol{V} \right) \leq \lambda_i \left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right). \qquad (29)$$

Noting that $\lambda_i \left(\boldsymbol{U}_s^T \boldsymbol{J} \boldsymbol{U}_s \right) = \lambda_i \left(\boldsymbol{J} \right)$ because they have the same first n eigenvalues, and by the fact that a nonnegative definite matrix is always Hermitian, we can see (29) is just a result of Poicaré separation theorem. Therefore the theorem follows. ■

REFERENCES

[1] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York, NY: John Wiley & Sons, 1973.

[2] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, pp. 87–90, Jan. 2001.

[3] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–591, May 1971.

[4] H. L. Van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Piscataway, NJ: IEEE Press, 2007.

[5] B. Su and P. P. Vaidyanathan, "Subspace-based blind channel identification for cyclic prefix systems using few received blocks," *IEEE Trans. Signal Process.*, vol. 55, pp. 4979–4993, Oct. 2007.

[6] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 26, pp. 1285–1301, Nov. 1990.

[7] T. L. Marzetta, "A simple derivation of the constrained multiple parameter Cramer-Rao bound," *IEEE Trans. Signal Process.*, vol. 41, pp. 2247–2249, Jun. 1993.

[8] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, pp. 177–179, Jul. 1998.

[9] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[10] E. de Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind, and training sequence based channel estimation," in *1st IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communications*, 1997, pp. 129–132.

[11] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Performance analysis of a deterministic channel estimator for block transmission systems with null guard intervals," *IEEE Trans. Signal Process.*, vol. 50, pp. 684–695, Mar. 2002.

[12] M. Spivak, *Calculus on Manifolds*. Cambridge, MA: Perseus Books, 1965.