



HAL
open science

Channel estimation: unified view of optimal performance and pilot sequences

Luc Le Magoarou, Stéphane Paquelet

► **To cite this version:**

Luc Le Magoarou, Stéphane Paquelet. Channel estimation: unified view of optimal performance and pilot sequences. 2020. hal-02472297v1

HAL Id: hal-02472297

<https://hal.science/hal-02472297v1>

Preprint submitted on 10 Feb 2020 (v1), last revised 1 Sep 2020 (v3)

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

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

Channel estimation: unified view of optimal performance and pilot sequences

Luc Le Magoarou, Stéphane Paquelet

Abstract—Channel estimation is of paramount importance in most communication systems in order to maximize data rate. In modern systems, the possibly large number of transmit/receive antennas and subcarriers makes this task difficult. Several parametric models, either linear or nonlinear, have thus been proposed to ease channel estimation by reducing the dimension of the estimation problem. In this paper, the problem of estimating an object of interest parameterized according to a model, based on noisy complex linear observations is studied. The Cramér-Rao bound (CRB) for this general problem is given, highlighting its key dependency on the introduced *variation space*. Then, identifiability conditions are given, and the minimal value of the CRB, as well as a general strategy to build minimal length power constrained observation matrices are determined. The results are finally illustrated in a general wideband MIMO context. They allow to retrieve well known previous results, but also to exhibit minimal length optimal pilot sequences in a simple nonlinear physical model setting. Note that the results of this paper are general and could perfectly be used in other applications as well.

Index Terms—Channel estimation, parametric model, Cramér-Rao bound.

I. INTRODUCTION

COMMUNICATION systems make use of a physical channel to convey information between a transmitter and a receiver [1]. Knowing the channel state at both ends of the link allows to maximize the data rate, hence the need to estimate the channel. This can be carried out by sending pilot signals known by both the transmitter and the receiver to gather noisy observations used to estimate the channel.

Recently, the ever-growing need for data rate in modern communication networks led to use channels of very high dimension, which makes channel estimation difficult. For example, it has been recently proposed to use massive multiple input multiple output (massive MIMO) wireless systems [2], [3], [4] with a large number of transmit and receive antennas in the millimeter-wave band [5], [6], where a large bandwidth can be exploited. In that case the channel comprises hundreds or even thousands of complex numbers, whose estimation is a very challenging signal processing problem [7].

In some cases, it is possible to express the channel as a function of a small number of parameters compared to its apparent dimension. This way, the dimension of the channel estimation problem is reduced, which allows to reduce both the estimation error and the quantity of pilots.

Contributions and organization. In this paper, we study channel estimation based on noisy linear observations and a *deterministic* channel model, be it linear or not, provided it is differentiable with respect to the parameters. The studied problem is formulated in section II. The notion of *variation space* is then introduced in section III, where a general result about \mathbb{R} -vector spaces is also stated in order to ease the remaining part of the paper. Then, the Cramér-Rao bound (CRB) [8], [9] of the considered problem is stated in section IV-A, where its dependency on the variation space is highlighted. Identifiability conditions on the observation matrices and the minimal number of observations for which they can be fulfilled are given in section IV-B. In section IV-C, we express the minimal variance of any unbiased estimator by optimizing the CRB under a power constraint on the observation matrix. Associated observation matrices of minimal length are also exhibited. Note that the strength of this study lies in its generality, since it encompasses all deterministic models, linear or not, depending on real or complex parameters, and is valid for any unbiased estimator. This renders the obtained results potentially useful well beyond the scope of channel estimation. These results are illustrated in a wideband MIMO context in section V. It is shown that the general framework allows to retrieve well-known results previously established for linear models, but also to derive optimal pilot sequences of minimal length in settings for which they had not been determined yet. The example of a very simple MIMO channel made of a single path is taken to illustrate this fact.

Note that this paper is based on some of our previous work [10], [11], in which the Cramér-Rao bound in the specific case of a physical channel model was stated. The novelty of this paper is that the Cramér-Rao bound is here optimized, and the derivation is more general since it is valid for any parametric model.

Related work. This paper generalizes results obtained in the case of a linear deterministic channel model, both in a MIMO context [12], [13] and for multicarrier systems [14], [15]. A notable difference is that the analysis of the present paper is based on the CRB (which depends only on the model and not on the estimation method) and not directly on the error incurred by a specific estimator, as is the case for the aforementioned papers.

There is also a vast body of literature regarding optimal pilot sequences in a bayesian channel estimation setting [16], [17], [18], for which the channel is assumed to follow a known distribution. In that case, the optimal pilot sequences have to span the space generated by the leading eigenvectors of the channel covariance matrix (which correspond in some

sense to the space most likely to contain the channel). These approaches are different in nature from the one of this paper, since (i) they consider a specific estimator (the linear minimum mean squared error (LMMSE)), and (ii) their objective is to minimize the estimation error in average over the channel distribution. On the other hand, the present paper is estimator independent and its objective can be seen as the minimization of the error for a given channel realization.

Recently, it has also been proposed to look for optimal pilot sequences in a multi-user bayesian setting. In [19], sequences are found by numerical optimization, minimizing a weighted sum of the channel estimation errors of each user. In [20] and [21], heuristics are proposed which amount to send pilot sequences that span the union of the spaces generated by the leading eigenvectors of the channel correlation matrices of all users.

The optimization part of this paper is in spirit similar to the one of [22]. Indeed, in [22] the objective is to find the optimal power constrained covariance matrix which maximizes the mutual information formula. Here, it is to find the optimal power constrained observation matrix which minimizes the CRB. Moreover, the proof of [22] uses a bound on the determinant of a matrix by the product of its diagonal entries, with an equality if and only if the matrix is diagonal (which is the case at the optimal). Here, we use a bound on the trace of the inverse of a matrix by the sum of the inverses of its diagonal entries, with an equality also if and only if the matrix is diagonal (which is also the case at the optimal).

II. PROBLEM FORMULATION

Notations. Matrices and vectors are denoted by bold upper-case and lower-case letters: \mathbf{A} and \mathbf{a} (except 3D “*spatial*” vectors that are denoted \vec{a}); the i th column of a matrix \mathbf{A} by \mathbf{a}_i ; its entry at the i th line and j th column by a_{ij} . $\mathbf{A}_{[i:j,:]}$ denotes the matrix built taking the rows i to j of \mathbf{A} (matlab style indexing). A matrix transpose, conjugate and transconjugate is denoted by \mathbf{A}^T , \mathbf{A}^* and \mathbf{A}^H respectively. The trace of a linear transformation represented by \mathbf{A} is denoted $\text{Tr}(\mathbf{A})$. The linear span of a set of vectors \mathcal{A} and its dimension (if it is a vector space) are denoted: $\text{span}_{\mathbb{R}}(\mathcal{A})$ and $\text{dim}_{\mathbb{R}}(\mathcal{A})$ when considering linear combinations with real coefficients, or $\text{span}_{\mathbb{C}}(\mathcal{A})$ and $\text{dim}_{\mathbb{C}}(\mathcal{A})$ when considering linear combinations with complex coefficients. The orthogonal complement of a subspace \mathcal{W} is denoted \mathcal{W}^\perp . The Kronecker product is denoted by \otimes . The identity matrix is denoted by \mathbf{Id} . $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the standard complex gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathbb{E}(\cdot)$ denotes the expectation and $\text{cov}(\cdot)$ the covariance of its argument.

We consider the general channel estimation setting where a channel $\mathbf{h} \in \mathbb{C}^{N_d}$ is to be estimated, where N_d is the number of complex dimensions of the channel. We assume it is deterministic and follows a parametric model depending on N_p real parameters. It can then be seen as a function $\mathbf{h} : \mathbb{R}^{N_p} \rightarrow \mathbb{C}^{N_d}$ that maps each parameters value to the corresponding channel (we will denote the channel indifferently \mathbf{h} or $\mathbf{h}(\boldsymbol{\theta})$ depending on the context). Note that this is the most generic setting since

complex parameters can always be decomposed into real and imaginary parts (or modulus and angle) and thus correspond to two real parameters each. The only assumption that we make about the channel model is that the function \mathbf{h} is differentiable with respect to the parameters.

Estimation is made based on N_m noisy linear measurements of the form

$$\mathbf{y} = \mathbf{M}^H \mathbf{h} + \mathbf{n}, \quad (1)$$

where $\mathbf{n} \in \mathbb{C}^{N_m}$ corresponds to the noise whose entries are assumed i.i.d. complex gaussians of variance σ^2 , so that $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{Id}_{N_m})$, and $\mathbf{M} \in \mathbb{C}^{N_d \times N_m}$ is the observation matrix, which is entirely determined by the pilot sequences sent by the emitter and the combining operations done at the receiver. this way of expressing the observations is very general (an example in a MIMO wideband context is given in section V).

The estimator of the parameters is a function mapping observations to estimates, denoted $\hat{\boldsymbol{\theta}} : \mathbb{C}^{N_m} \rightarrow \mathbb{R}^{N_p}$. The estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$ will be denoted $\hat{\boldsymbol{\theta}}$ (as the estimator) for shorter notations. The channel estimate is given by the model function, as $\mathbf{h}(\hat{\boldsymbol{\theta}})$.

The error is measured by the mean squared error (MSE):

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\theta}}) &\triangleq \mathbb{E} \left[\|\mathbf{h}(\boldsymbol{\theta}) - \mathbf{h}(\hat{\boldsymbol{\theta}})\|_2^2 \right] \\ &= \|\mathbf{h}(\boldsymbol{\theta}) - \mathbb{E}[\mathbf{h}(\hat{\boldsymbol{\theta}})]\|_2^2 + \mathbb{E} \left[\|\mathbf{h}(\hat{\boldsymbol{\theta}}) - \mathbb{E}[\mathbf{h}(\hat{\boldsymbol{\theta}})]\|_2^2 \right], \end{aligned}$$

where the expectation is taken over the noise distribution, and the second line corresponds to the well-known bias-variance decomposition [23]. We assume throughout the paper that the considered channel estimators are unbiased with respect to $\mathbf{h}(\boldsymbol{\theta})$, which reads $\mathbb{E}[\mathbf{h}(\hat{\boldsymbol{\theta}})] = \mathbf{h}(\boldsymbol{\theta})$. It follows

$$\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E} \left[\|\mathbf{h}(\hat{\boldsymbol{\theta}}) - \mathbb{E}[\mathbf{h}(\hat{\boldsymbol{\theta}})]\|_2^2 \right] = \text{Tr}[\text{cov}(\mathbf{h}(\hat{\boldsymbol{\theta}}))]. \quad (2)$$

This way, the bias is null and the MSE is entirely due to the variance of the estimator $\mathbf{h}(\hat{\boldsymbol{\theta}})$.

In this paper, we express this variance term as a function of the observation matrix and optimize it under a power constraint in order to exhibit optimal pilot sequences and the associated minimal variance, for any deterministic channel model.

III. VARIATION SPACE

In this section, the notion of variation space, which plays a central role in the analysis we propose, is first introduced. Then, a result allowing to decompose it in order to ease the following analysis is stated.

A. Preliminaries

We begin this section by defining notions needed for our analysis. First, let $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \triangleq \left(\frac{\partial \mathbf{h}}{\partial \theta_1}, \dots, \frac{\partial \mathbf{h}}{\partial \theta_{N_p}} \right) \in \mathbb{C}^{N_d \times N_p}$ be the complex gradient of the channel with respect to its real parameters. We will make extensive use of the set defined below.

Definition 1. (Variation space) *Let the set*

$$\mathcal{V}_{\boldsymbol{\theta}} \triangleq \left\{ \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathbf{x}, \mathbf{x} \in \mathbb{R}^{N_p} \right\}$$

which proves our claim for the first step of the descent. The exact same reasoning can be applied at each subsequent step of the descent. \square

It is interesting to notice that another (more algebraic) proof of this lemma is possible, which gives a practical way to obtain the basis vectors corresponding to the decomposition. Indeed, let \mathbf{U} be any matrix whose columns form an \mathbb{R} -orthonormal basis of \mathcal{E} . Then,

$$\mathbf{U}^H \mathbf{U} = \mathbf{Id} + j\mathbf{A},$$

where the matrix $\mathbf{A} = \Im\{\mathbf{U}^H \mathbf{U}\}$ is skew-symmetric, so that it admits the following normal form [24, Theorem 8.16] :

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \begin{pmatrix} 0 & -c_1 & & & & \\ c_1 & 0 & & & & \\ & & 0 & -c_2 & & \\ & & c_2 & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \triangleq \mathbf{\Gamma},$$

with $\mathbf{Q} \in \mathbb{R}^{d \times d}$ a real orthogonal matrix ($\mathbf{Q}^T \mathbf{Q} = \mathbf{Id}$) whose columns are eigenvectors of the symmetric positive semi-definite matrix $\mathbf{A}^T \mathbf{A} = -\mathbf{A}^2$ (whose nonzero eigenvalues are all of multiplicity two and correspond to c_1^2, c_2^2, \dots), and where $0 \leq c_k \leq 1, \forall k$. This yields

$$\mathbf{Q}^T \mathbf{U}^H \mathbf{U} \mathbf{Q} = \mathbf{Id} + j\mathbf{\Gamma} = \begin{pmatrix} 1 & -jc_1 & & & & \\ jc_1 & 1 & & & & \\ & & 1 & -jc_2 & & \\ & & jc_2 & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix},$$

so that the columns of the matrix $\mathbf{U} \mathbf{Q}$ form an \mathbb{R} -orthonormal basis of \mathcal{E} in which vectors can be grouped by two so that vectors of different groups are \mathbb{C} -orthogonal. This is exactly the main claim of lemma 1.

We preferred giving a geometric proof here in order to give more insight on the interaction between \mathbb{R} -vector spaces and \mathbb{C} -vector spaces. Indeed, our proof highlights the fact that the quantity c_i can be nicely interpreted as the squared cosine of the i -th principal angle [25] between \mathcal{E} and $j\mathcal{E}$.

2) *Application to \mathcal{V}_θ* : Applying lemma 1 to the variation space \mathcal{V}_θ (assuming it is of dimension N_p and N_p is even), it is possible to decompose it as

$$\mathcal{V}_\theta = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{v}_1, \mathbf{w}_1, \dots, \mathbf{v}_{\frac{N_p}{2}}, \mathbf{w}_{\frac{N_p}{2}} \right\} \right) \quad (8)$$

where $\mathbf{v}_m^H \mathbf{v}_n = \delta_{mn}$, $\mathbf{w}_m^H \mathbf{w}_n = \delta_{mn}$ and $\mathbf{v}_m^H \mathbf{w}_n = -\delta_{mn} j c_m$ (δ being the Kronecker symbol). Let us introduce the matrix

$$\mathbf{V} \triangleq \left(\mathbf{v}_1, \mathbf{w}_1, \dots, \mathbf{v}_{\frac{N_p}{2}}, \mathbf{w}_{\frac{N_p}{2}} \right) \quad (9)$$

whose columns form an \mathbb{R} -orthonormal basis of \mathcal{V}_θ . Similarly, if N_p is odd, the decomposition reads

$$\mathcal{V}_\theta = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{v}_1, \mathbf{w}_1, \dots, \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor}, \mathbf{w}_{\lfloor \frac{N_p}{2} \rfloor}, \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor + 1} \right\} \right), \quad (10)$$

where $\mathbf{v}_m^H \mathbf{v}_n = \delta_{mn}$, $\mathbf{w}_m^H \mathbf{w}_n = \delta_{mn}$ and $\mathbf{v}_m^H \mathbf{w}_n = -\delta_{mn} j c_m$, and the matrix \mathbf{V} can be built the same way. In practice, the matrix \mathbf{V} can be obtained by exhibiting any matrix \mathbf{U} whose columns form an \mathbb{R} -orthonormal basis of \mathcal{V}_θ , then finding a matrix \mathbf{Q} whose columns form a basis of eigenvectors for the matrix $-\Im\{\mathbf{U}^H \mathbf{U}\}^2$ (eigenvalues sorted in decreasing order). Then, we simply have $\mathbf{V} = \mathbf{U} \mathbf{Q}$ and c_k corresponds to the square root of the k -th greatest eigenvalue of $-\Im\{\mathbf{U}^H \mathbf{U}\}^2$.

IV. CRAMÉR-RAO BOUND AND OPTIMAL OBSERVATION MATRICES

In this section, the Cramér-Rao bound is analyzed in order to exhibit optimal observation matrices, from which can be deduced optimal pilot sequences. The expression of the CRB is first given. The identifiability of the parameters is then studied. Finally, optimal observations matrices under a power constraint are characterized.

A. Expression of the CRB

Let us give a general expression for the Cramér-Rao bound [8], [9], which is a lower bound on the variance of any unbiased estimator. The complex CRB [26] takes the form

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{M}) \triangleq \text{Tr} \left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathbf{I}(\boldsymbol{\theta}, \mathbf{M})^{-1} \frac{\partial \mathbf{h}^H}{\partial \boldsymbol{\theta}} \right]$$

where $\mathbf{I}(\boldsymbol{\theta}, \mathbf{M}) \in \mathbb{R}^{N_p \times N_p}$ is the Fisher information matrix (FIM) which quantifies the amount of information about the parameters $\boldsymbol{\theta}$ that the observation \mathbf{y} carries when using the observation matrix \mathbf{M} . This is a lower bound in the sense that

$$\text{Tr} \left[\text{cov}(\mathbf{h}(\hat{\boldsymbol{\theta}})) \right] \geq \text{CRB}(\boldsymbol{\theta}, \mathbf{M}).$$

In the general setting we consider, the CRB can be expressed in a very simple way, as shown by the following theorem.

Theorem 1. *Provided $\dim_{\mathbb{R}}(\mathcal{V}_\theta) = N_p$, the Cramér-Rao bound is expressed as*

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{M}) = \frac{\sigma^2}{2} \text{Tr} \left[\Re \left\{ \mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U} \right\}^{-1} \right],$$

where \mathbf{U} is any matrix whose columns form an \mathbb{R} -orthonormal basis of the variation space \mathcal{V}_θ .

Proof. The observation defined in (1) follows a gaussian distribution,

$$\mathbf{y} \sim \mathcal{CN}(\mathbf{M}^H \mathbf{h}, \sigma^2 \mathbf{Id}),$$

so that the FIM is given by the Slepian-Bangs formula [27], [28], [29]:

$$\mathbf{I}(\boldsymbol{\theta}, \mathbf{M}) = \frac{2}{\sigma^2} \Re \left\{ \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \mathbf{M} \mathbf{M}^H \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \right\}.$$

In this form, the FIM is difficult to invert, because it involves the real part of a complex matrix. In [10], we proposed to use real representations of complex matrices to get rid of this problem. Here, in order to gain a deeper geometric understanding of the bound, let us use the Gram-Schmidt

process on the gradient matrix $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$, with the real inner product $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ to decompose it as

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} = \mathbf{U}\mathbf{R},$$

where $\mathbf{U} \in \mathbb{C}^{N_d \times K}$ is a matrix whose columns are \mathbb{R} -orthonormal (meaning that $\Re\{\mathbf{U}^H \mathbf{U}\} = \mathbf{I}_{d_K}$), $\mathbf{R} \in \mathbb{R}^{K \times N_p}$ is a real upper-triangular matrix, and $K = \dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}})$. This decomposition of the gradient matrix allows to rewrite the FIM

$$\mathbf{I}(\boldsymbol{\theta}, \mathbf{M}) = \frac{2}{\sigma^2} \mathbf{R}^T \Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\} \mathbf{R},$$

since $\Re\{\mathbf{A}^H \mathbf{B} \mathbf{A}\} = \mathbf{A}^T \Re\{\mathbf{B}\} \mathbf{A}$ as soon as \mathbf{A} is a real matrix. Now, if and only if \mathbf{R} is invertible, which is equivalent to $K = N_p$, the CRB is expressed

$$\begin{aligned} \text{CRB}(\boldsymbol{\theta}, \mathbf{M}) &= \text{Tr} \left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathbf{I}(\boldsymbol{\theta}, \mathbf{M})^{-1} \frac{\partial \mathbf{h}^H}{\partial \boldsymbol{\theta}} \right] \\ &= \frac{\sigma^2}{2} \text{Tr} \left[\mathbf{U} \mathbf{R} \mathbf{R}^{-1} \Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}^{-1} \mathbf{R}^{-T} \mathbf{R}^T \mathbf{U}^H \right] \\ &= \frac{\sigma^2}{2} \text{Tr} \left[\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}^{-1} \right]. \end{aligned}$$

In order to conclude, one can remark that

$$\text{Tr} \left[\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}^{-1} \right] = \text{Tr} \left[\Re\{\mathbf{Q}^T \mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U} \mathbf{Q}\}^{-1} \right]$$

for any real orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{N_p \times N_p}$, so that the equation holds true for any matrix whose columns form an \mathbb{R} -orthogonal basis of $\mathcal{V}_{\boldsymbol{\theta}}$. \square

The result of this theorem shows an invariance property, it is true for any matrix \mathbf{U} whose columns are an \mathbb{R} -orthonormal basis of $\mathcal{V}_{\boldsymbol{\theta}}$. Moreover, the matrix $\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}$ can be given a nice interpretation. Indeed, the orthogonal projection $\mathbf{P}_{\mathcal{V}_{\boldsymbol{\theta}}}$ of any vector \mathbf{z} onto $\mathcal{V}_{\boldsymbol{\theta}}$ is expressed

$$\mathbf{P}_{\mathcal{V}_{\boldsymbol{\theta}}} \mathbf{z} = \sum_{i=1}^{N_p} \langle \mathbf{u}_i, \mathbf{z} \rangle_{\mathbb{R}} \mathbf{u}_i = \mathbf{U} \Re\{\mathbf{U}^H \mathbf{z}\},$$

so that $\Re\{\mathbf{U}^H \mathbf{z}\}$ corresponds to the coordinates of the projection in the basis given by \mathbf{U} . Now, if $\mathbf{z} = \mathbf{M} \mathbf{M}^H \mathbf{t}$ with $\mathbf{t} \in \mathcal{V}_{\boldsymbol{\theta}}$ then $\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{t}\} = \Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\} \mathbf{r}$ for some $\mathbf{r} \in \mathbb{R}^{N_p}$ corresponding to the coordinates of \mathbf{t} in the basis given by \mathbf{U} . It means that $\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}$ is the matrix that corresponds to the operator $\mathbf{P}_{\mathcal{V}_{\boldsymbol{\theta}}} \mathbf{M} \mathbf{M}^H$ restricted to $\mathcal{V}_{\boldsymbol{\theta}}$ when expressed in the basis given by \mathbf{U} . Such an operator corresponds to the notion of compression in functional analysis.

Definition 2. (Compression [30, p.120]) *Let \mathcal{H} be a subspace of a Hilbert space \mathcal{K} , let $\mathbf{P}_{\mathcal{H}}$ be the orthogonal projection from \mathcal{K} onto \mathcal{H} , and let $\mathbf{B} : \mathcal{K} \rightarrow \mathcal{K}$ be a linear operator on \mathcal{K} . The linear operator $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}$ is the compression of \mathbf{B} to \mathcal{H} , denoted $[\mathbf{B}]_{\mathcal{H}}$, if*

$$\mathbf{A} \mathbf{x} = \mathbf{P}_{\mathcal{H}} \mathbf{B} \mathbf{x}, \quad \forall \mathbf{x} \in \mathcal{H}.$$

In the following, and when no confusion is possible, we denote the same way a matrix \mathbf{A} and the operator associated

to the multiplication by \mathbf{A} . Moreover, for an operator $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}$ where \mathcal{H} is a \mathbb{K} -vector space ($\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$), we define its trace as

$$\text{Tr}[\mathbf{A}] \triangleq \sum_{i=1}^{N_p} \langle \mathbf{v}_i, \mathbf{A} \mathbf{v}_i \rangle_{\mathbb{K}},$$

where $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_p}\}$ is any \mathbb{K} -orthonormal basis of \mathcal{H} . It coincides with the sum of the diagonal elements of a matrix when the operator action is a matrix multiplication. These two notions allow to express the CRB in a simpler form, as in the following corollary (which is nothing more than a coordinate-free version of theorem 1).

Corollary 1. *Provided $\dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}}) = N_p$, the Cramér-Rao bound admits an intrinsic expression as*

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{M}) = \frac{\sigma^2}{2} \text{Tr} \left[\left([\mathbf{M} \mathbf{M}^H]_{\mathcal{V}_{\boldsymbol{\theta}}} \right)^{-1} \right],$$

where $[\mathbf{M} \mathbf{M}^H]_{\mathcal{V}_{\boldsymbol{\theta}}}$ is the compression of $\mathbf{M} \mathbf{M}^H$ to the variation space $\mathcal{V}_{\boldsymbol{\theta}}$.

This form of the CRB shows that the minimal variance of any unbiased estimator is determined by the interaction between the observation matrix \mathbf{M} and the potential directions of variations of the channel due to infinitesimal variations of the parameters around their value, represented by the set $\mathcal{V}_{\boldsymbol{\theta}}$. This fact, which is key in our analysis, is further exploited in the following subsections.

B. Identifiability

Parameters are said to be identifiable if and only if the CRB is finite,

$$\text{Identifiability} \Leftrightarrow \text{CRB}(\boldsymbol{\theta}, \mathbf{M}) < +\infty.$$

Identifiability imposes conditions on the variation space $\mathcal{V}_{\boldsymbol{\theta}}$ and on the observation matrix \mathbf{M} , as stated in the following theorem.

Theorem 2. *The parameters are identifiable if and only if*

$$\dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}}) = N_p$$

and

$$\mathcal{V}_{\boldsymbol{\theta}} \cap \text{im}_{\mathbb{C}}(\mathbf{M})^{\perp} = \{\mathbf{0}\}.$$

Proof. The first condition $\dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}}) = N_p$ is equivalent to the invertibility of \mathbf{R} that was shown to be a necessary condition for the CRB to be finite in section IV-A. When this condition is fulfilled, identifiability holds if and only if the matrix $\Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\}$ is invertible. This matrix being symmetric, it is invertible if and only if

$$\forall \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^{N_p}, \mathbf{x}^T \Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\} \mathbf{x} \neq 0.$$

Moreover, for any real vector \mathbf{x} , $\mathbf{x}^T \Re\{\mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U}\} \mathbf{x} = \mathbf{x}^T \mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U} \mathbf{x}$. Thus, recalling that $\mathcal{V}_{\boldsymbol{\theta}} = \text{im}_{\mathbb{R}}(\mathbf{U})$, identifiability holds if and only if

$$\forall \mathbf{z} \neq \mathbf{0} \in \mathcal{V}_{\boldsymbol{\theta}}, \mathbf{z}^H \mathbf{M} \mathbf{M}^H \mathbf{z} = \|\mathbf{M}^H \mathbf{z}\|_2^2 \neq 0,$$

which is equivalent (since $\ker(\mathbf{M}^H) = \text{im}_{\mathbb{C}}(\mathbf{M})^\perp$) to

$$\mathcal{V}_\theta \cap \text{im}_{\mathbb{C}}(\mathbf{M})^\perp = \{\mathbf{0}\}.$$

□

Interpretations. The first identifiability condition $\dim_{\mathbb{R}}(\mathcal{V}_\theta) = N_p$ means that the columns of $\frac{\partial \mathbf{h}}{\partial \theta}$ have to be linearly independent over \mathbb{R} for identifiability to be possible, whatever the observation matrix. Said differently, the number of degrees of freedom of the variation space has to be equal to the number of parameters to estimate, so that small variations of the channel \mathbf{h} due to an infinitesimal variation in the value of any parameter cannot be mistaken with small variations of the channel due to infinitesimal variations in the values of the other parameters. Note that since $\dim_{\mathbb{R}}(\mathcal{V}_\theta) \leq 2N_d$, this condition implies $N_p \leq 2N_d$, which means that it is impossible to identify a number of parameters that is more than twice the dimension of the observations space.

Then, if the first condition is fulfilled, the second condition $\mathcal{V}_\theta \cap \text{im}_{\mathbb{C}}(\mathbf{M})^\perp = \{\mathbf{0}\}$ means that no nonzero vector in the space of variations \mathcal{V}_θ can be orthogonal to the column space of the observation matrix \mathbf{M} for identifiability to hold. Said differently, the observation matrix has to preserve some energy for any element of the space of variations, every infinitesimal variation in the values of the parameters has to cause a change in the observation vector \mathbf{y} .

Number of observations. Identifiability directly imposes a minimal number of observations N_m , as stated in the following corollary.

Corollary 2. *Parameters can be identifiable only if*

$$N_m \geq \frac{N_p}{2}.$$

Proof. Identifiability can be stated:

$$\forall \mathbf{z} \neq \mathbf{0} \in \mathcal{V}_\theta, \mathbf{M}^H \mathbf{z} \neq \mathbf{0},$$

which is possible only if the \mathbb{R} -dimension of $\ker(\mathbf{M}^H)$ plus the \mathbb{R} -dimension of \mathcal{V}_θ is no greater than the \mathbb{R} -dimension of the ambient space \mathbb{C}^{N_d} (so that they can have a trivial intersection). This writes

$$\dim_{\mathbb{R}}(\ker(\mathbf{M}^H)) + N_p \leq 2N_d.$$

Moreover, $\dim_{\mathbb{R}}(\ker(\mathbf{M}^H)) = 2N_d - \dim_{\mathbb{R}}(\text{im}_{\mathbb{C}}(\mathbf{M}^H))$ (rank-nullity theorem), so that we end up with

$$\dim_{\mathbb{R}}(\text{im}_{\mathbb{C}}(\mathbf{M}^H)) \geq N_p.$$

The \mathbb{R} -dimension of a \mathbb{C} -vector space being twice its \mathbb{C} -dimension and the \mathbb{C} -dimension being upper-bounded by the number of columns, we finally get

$$N_m \geq \dim_{\mathbb{C}}(\text{im}_{\mathbb{C}}(\mathbf{M}^H)) \geq \frac{N_p}{2},$$

which proves the result. □

We just showed that the minimal number of observations N_m required for identifiability to be possible is $\lceil \frac{N_p}{2} \rceil$. In other

words, the matrix \mathbf{M} has to have at least $\lceil \frac{N_p}{2} \rceil$ columns for the CRB to be finite. As will be shown in the next subsection, there always exist an optimal observation matrix having $\lceil \frac{N_p}{2} \rceil$ columns.

C. Optimality

Let us now determine the minimal value of the CRB under a power constraint, and the observation matrices allowing to attain it. This corresponds to solve the optimization problem:

$$\begin{aligned} & \underset{\mathbf{M}}{\text{minimize}} \quad \text{CRB}(\boldsymbol{\theta}, \mathbf{M}), \\ & \text{subject to} \quad \|\mathbf{M}\|_F^2 = P. \end{aligned} \quad (11)$$

Note that the quantity $\|\mathbf{M}\|_F^2 = P = \text{Tr}(\mathbf{M}\mathbf{M}^H)$ corresponds to the observation power, which is proportional to the received power not directly equal to the transmitted power. The two quantities are linked in section V. The main result of this paper is given by the following theorem.

Theorem 3. *If N_p is even, the minimal value of the CRB is*

$$\text{CRB}_{\min}(\boldsymbol{\theta}) \triangleq \frac{2\sigma^2}{P} \left(\sum_{k=1}^{\frac{N_p}{2}} \frac{1}{\sqrt{1+c_k}} \right)^2,$$

where the scalars c_k are defined at (8). It is attained with

$$\mathbf{M} = \sqrt{\frac{P}{C}} \begin{pmatrix} \mathbf{v}_1 + \mathbf{j}\mathbf{w}_1 & \dots & \mathbf{v}_{\frac{N_p}{2}} + \mathbf{j}\mathbf{w}_{\frac{N_p}{2}} \\ \frac{1}{(1+c_1)^{\frac{3}{4}}} & \dots & \frac{1}{(1+c_{\frac{N_p}{2}})^{\frac{3}{4}}} \end{pmatrix},$$

where $C \triangleq 2 \sum_{l=1}^{\frac{N_p}{2}} \frac{1}{\sqrt{1+c_l}}$ and the vectors $\mathbf{v}_k, \mathbf{w}_k$ are defined at (8).

If N_p is odd, the minimal value of the CRB is

$$\text{CRB}_{\min}(\boldsymbol{\theta}) \triangleq \frac{2\sigma^2}{P} \left(\sum_{k=1}^{\lfloor \frac{N_p}{2} \rfloor} \frac{1}{\sqrt{1+c_k}} + \frac{1}{2} \right)^2,$$

where the scalars c_k are defined at (10). It is attained with

$$\mathbf{M} = \sqrt{\frac{P}{C}} \begin{pmatrix} \mathbf{v}_1 + \mathbf{j}\mathbf{w}_1 & \dots & \frac{\mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor} + \mathbf{j}\mathbf{w}_{\lfloor \frac{N_p}{2} \rfloor}}{(1+c_{\lfloor \frac{N_p}{2} \rfloor})^{\frac{3}{4}}} & \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor + 1} \\ \frac{1}{(1+c_1)^{\frac{3}{4}}} & \dots & \frac{1}{(1+c_{\lfloor \frac{N_p}{2} \rfloor})^{\frac{3}{4}}} & \frac{1}{2} \end{pmatrix},$$

where $C \triangleq 2 \sum_{l=1}^{\lfloor \frac{N_p}{2} \rfloor} \frac{1}{\sqrt{1+c_l}} + 1$ and the vectors $\mathbf{v}_k, \mathbf{w}_k$ are defined at (10).

Proof. Let us first consider the case where N_p is even. Starting from the result of theorem 1 and using the fact that it holds true for any matrix whose columns form an \mathbb{R} -orthonormal basis of \mathcal{V}_θ , we express the CRB as

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{M}) = \frac{\sigma^2}{2} \text{Tr} \left[\Re \left\{ \mathbf{V}^H \mathbf{M} \mathbf{M}^H \mathbf{V} \right\}^{-1} \right],$$

where \mathbf{V} is the matrix defined in (9) when applying lemma 1 to \mathcal{V}_θ .

Next, using the fact that for a symmetric positive semidefinite matrix \mathbf{A} , $(\mathbf{A}^{-1})_{ii} \geq \frac{1}{a_{ii}}, \forall i$ [31, Theorem 7.7.15], we get

$$\text{Tr} \left[\Re \left\{ \mathbf{V}^H \mathbf{M} \mathbf{M}^H \mathbf{V} \right\}^{-1} \right] \geq \sum_{k=1}^{\frac{N_p}{2}} \frac{1}{\|\mathbf{M}^H \mathbf{v}_k\|_2^2} + \frac{1}{\|\mathbf{M}^H \mathbf{w}_k\|_2^2},$$

with an equality if and only if the matrix $\Re\{\mathbf{V}^H \mathbf{M} \mathbf{M}^H \mathbf{V}\}$ is diagonal. In order to proceed, let us define

$$\tilde{\mathbf{u}}_k^+ = \frac{1}{\sqrt{2(1+c_k)}}(\mathbf{v}_k + j\mathbf{w}_k)$$

and

$$\tilde{\mathbf{u}}_k^- = \frac{1}{\sqrt{2(1-c_k)}}(\mathbf{v}_k - j\mathbf{w}_k),$$

which are unitary vectors such that $\tilde{\mathbf{u}}_k^+ \perp_{\mathbb{C}} \tilde{\mathbf{u}}_k^-$, $\forall k$. These vectors allow to express

$$\begin{aligned} \|\mathbf{M}^H \mathbf{v}_k\|_2^2 &= \frac{1}{2} \left[(1+c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^+\|_2^2 + (1-c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^-\|_2^2 \right. \\ &\quad \left. + \sqrt{1-c_k^2} \Re\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\} \right], \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{M}^H \mathbf{w}_k\|_2^2 &= \frac{1}{2} \left[(1+c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^+\|_2^2 + (1-c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^-\|_2^2 \right. \\ &\quad \left. - \sqrt{1-c_k^2} \Re\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\} \right]. \end{aligned}$$

Now, let us define $P_k^+ \triangleq \|\mathbf{M}^H \tilde{\mathbf{u}}_k^+\|_2^2$, $P_k^- \triangleq \|\mathbf{M}^H \tilde{\mathbf{u}}_k^-\|_2^2$ and $d_k \triangleq \sqrt{1-c_k^2} \Re\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\}$, so that we have

$$\begin{aligned} &\sum_{k=1}^{\frac{N_p}{2}} \frac{1}{\|\mathbf{M}^H \mathbf{v}_k\|_2^2} + \frac{1}{\|\mathbf{M}^H \mathbf{w}_k\|_2^2} \\ &= \sum_{k=1}^{\frac{N_p}{2}} \frac{2}{(1+c_k)P_k^+ + (1-c_k)P_k^- + d_k} \\ &\quad + \frac{2}{(1-c_k)P_k^- + (1+c_k)P_k^+ - d_k} \\ &\geq \sum_{k=1}^{\frac{N_p}{2}} \frac{4}{(1-c_k)P_k^- + (1+c_k)P_k^+}, \end{aligned}$$

the last inequality being a direct consequence of the fact that $\frac{1}{a+b} + \frac{1}{a-b} \geq \frac{2}{a}$ (because of the convexity of the inverse function on \mathbb{R}_+). It becomes an equality if and only if $d_k = \sqrt{1-c_k^2} \Re\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\} = 0$, $\forall k$.

In summary, we have

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{M}) = \frac{\sigma^2}{2} \sum_{k=1}^{\frac{N_p}{2}} \frac{4}{(1-c_k)P_k^- + (1+c_k)P_k^+}$$

if and only if $\Re\{\mathbf{V}^H \mathbf{M} \mathbf{M}^H \mathbf{V}\}$ is diagonal and $\Re\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\} = 0$, $\forall k$. Moreover, $\|\mathbf{M}\|_F^2 = \text{Tr}[\mathbf{M} \mathbf{M}^H] \geq \sum_{k=1}^{\frac{N_p}{2}} P_k^+ + P_k^-$, with an equality if and only if $\text{im}_{\mathbb{C}}(\mathbf{M}) \subset \text{span}_{\mathbb{C}}(\{\tilde{\mathbf{u}}_k^+, \tilde{\mathbf{u}}_k^-\}_{k=1}^{\frac{N_p}{2}})$. The optimization problem (11) is thus lower-bounded by the simpler problem

$$\begin{aligned} &\underset{P_k^+, P_k^-, k=1, \dots, \frac{N_p}{2}}{\text{minimize}} && \sum_{k=1}^{\frac{N_p}{2}} \frac{4}{(1-c_k)P_k^- + (1+c_k)P_k^+}, \\ &\text{subject to} && \sum_{k=1}^{\frac{N_p}{2}} P_k^+ + P_k^- = P. \end{aligned} \quad (12)$$

Let us solve this problem and then identify matrices \mathbf{M} for which the optimal values of (12) and (11) coincide. It is

obvious that at the optimum of (12), $P_k^- = 0$, $\forall k$, so that it is equivalent to solve the even simpler problem

$$\begin{aligned} &\underset{P_k^+, k=1, \dots, \frac{N_p}{2}}{\text{minimize}} && \sum_{k=1}^{\frac{N_p}{2}} \frac{4}{(1+c_k)P_k^+}, \\ &\text{subject to} && \sum_{k=1}^{\frac{N_p}{2}} P_k^+ = P. \end{aligned} \quad (13)$$

Using the Lagrange multipliers method, it is straightforward to obtain the optimal powers

$$(P_k^+)_{\text{opt}} = \frac{P}{\sqrt{1+c_k} \sum_{j=1}^{\frac{N_p}{2}} \frac{1}{\sqrt{1+c_j}}},$$

and the optimal value of the optimization problems (12) and (13) is

$$\sum_{k=1}^{\frac{N_p}{2}} \frac{4}{(1+c_k)(P_k^+)_{\text{opt}}} = \frac{4}{P} \left(\sum_{k=1}^{\frac{N_p}{2}} \frac{1}{\sqrt{1+c_k}} \right)^2.$$

It is also the optimal value of problem (11), since it is attained with the observation matrix

$$\mathbf{M}_{\text{opt}} = \left(\sqrt{(P_1^+)_{\text{opt}}} \tilde{\mathbf{u}}_1^+, \dots, \sqrt{\left(\frac{P_{\frac{N_p}{2}}^+}{2} \right)_{\text{opt}}} \tilde{\mathbf{u}}_{\frac{N_p}{2}}^+ \right),$$

which indeed guarantees that $P_k^+ = (P_k^+)_{\text{opt}}$ and $d_k = 0$, $\forall k$, that $\Re\{\mathbf{V}^H \mathbf{M}_{\text{opt}} \mathbf{M}_{\text{opt}}^H \mathbf{V}\}$ is diagonal, and that $\|\mathbf{M}_{\text{opt}}\|_F^2 = P$.

The proof is very similar in the case where N_p is odd, the only difference being that the decomposition of $\mathcal{V}_{\boldsymbol{\theta}}$ is the one given in (10) rather than the one given in (8). \square

This theorem exhibits the fact that the optimal CRB depends on the noise level σ^2 , the observation power P and the properties of the variation space $\mathcal{V}_{\boldsymbol{\theta}}$, namely its dimension N_p and the quantities c_k . Moreover it can be bounded above and below as

$$\frac{\sigma^2 N_p^2}{4P} \leq \text{CRB}_{\min}(\boldsymbol{\theta}) \leq \frac{\sigma^2 N_p^2}{2P},$$

with an equality on the left if and only if N_p is even and $c_k = 1$, $\forall k$ ($\mathcal{V}_{\boldsymbol{\theta}}$ is then a \mathbb{C} -vector space), and equality on the right if and only if $c_k = 0$, $\forall k$ ($\mathcal{V}_{\boldsymbol{\theta}}$ is then \mathbb{R} -orthogonal to $j\mathcal{V}_{\boldsymbol{\theta}}$).

Regarding the optimal observation matrices, a generic strategy to design them is the following:

- 1) Compute $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$,
- 2) Perform a QR factorization to get \mathbf{U} ,
- 3) Compute the eigendecomposition of $\Im\{\mathbf{U}^H \mathbf{U}\}$ to get $\mathbf{v}_k, \mathbf{w}_k, c_k, \forall k$,
- 4) Apply theorem 3.

Note that this strategy does not always lead to schemes that are realizable in practice. Indeed, $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$ may depend on the true parameters value $\boldsymbol{\theta}$ (which is unknown). The strategy can be implemented only if; either $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$ is a constant with respect to $\boldsymbol{\theta}$, or it can be estimated beforehand (as evoked in the next section in the case of MIMO channel estimation, where it can be determined based on the uplink channel estimates).

V. ILLUSTRATIONS OF THE RESULTS

Let us now introduce the most widespread parametric models used for channel estimation in a wideband MIMO context, which will be used to illustrate the applicability of the main results of the paper. However, note that the results are applicable to any parametric model and any application context, not just those presented here.

Setting. We consider the general case where the channel to estimate is between N_t transmit antennas and N_r receive antennas, on N_f subcarriers. This situation leads to a complex channel vector of dimension $N_d = N_r N_t N_f$ denoted $\mathbf{h} \in \mathbb{C}^{N_r N_t N_f}$, where h_{ijk} is the channel between the j -th transmit antenna and the i -th receive antenna on the k -th subcarriers. The observation matrix \mathbf{M} takes a particular form in this context. Indeed, if the transmitter sends a pilot sequence of length T corresponding to the matrix $\mathbf{X} \in \mathbb{C}^{N_t \times T}$ on N_{ps} pilot subcarriers, then the signal at the receive antennas can be written as in (1) with

$$\mathbf{M} = \mathbf{Id}_{N_r} \otimes \mathbf{X} \otimes \mathbf{F} \in \mathbb{C}^{N_r N_t N_f \times N_r T N_{ps}}, \quad (14)$$

where $\mathbf{F} \in \{0, 1\}^{N_f \times N_{ps}}$ is a column-sampled identity matrix, keeping only the columns corresponding to the selected pilot subcarriers. In such a setting, the number of complex observations is $N_m = N_r T N_{ps}$, and the transmitted power is $P_t \triangleq N_{ps} \|\mathbf{X}\|_2^2$. On the other hand, the observation power which is constrained in the optimization problem (11), is expressed $P = \|\mathbf{M}\|_2^2 = N_r N_{ps} \|\mathbf{X}\|_2^2$. We thus have $P = N_r P_t$, acknowledging the fact that adding receive antennas increases the received power P without changing the transmitted power P_t .

A. Least squares model

The most direct way to parameterize the channel is to take as $N_p = 2N_r N_t N_f = 2N_d$ parameters the real and imaginary parts of the channel entries,

$$\boldsymbol{\theta}_{LS} \triangleq (\Re(\mathbf{h}), \Im(\mathbf{h}))^T \in \mathbb{R}^{2N_r N_t N_f},$$

This leads to a linear channel model, expressed in function of the parameters as

$$\mathbf{h}_{LS}(\boldsymbol{\theta}_{LS}) = (\mathbf{Id}, \mathbf{jId}) \boldsymbol{\theta}_{LS}. \quad (15)$$

The observation defined in (1) then reads $\mathbf{y} = (\mathbf{M}^H, \mathbf{jM}^H) \boldsymbol{\theta}_{LS} + \mathbf{n}$ and the maximum likelihood estimation problem becomes a least squares problem, hence the name of the model. In this case, $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}_{LS}} = (\mathbf{Id}, \mathbf{jId})$, which, following the definition of the variation space gives

$$\mathcal{V}_{\boldsymbol{\theta}_{LS}} = \mathbb{C}^{N_r N_t N_f}. \quad (16)$$

This particular variation space can be decomposed according to lemma 1 as

$$\mathcal{V}_{\boldsymbol{\theta}_{LS}} = \text{span}_{\mathbb{R}} \left(\{ \mathbf{b}_1, -\mathbf{j}\mathbf{b}_1, \dots, \mathbf{b}_{N_r N_t N_f}, -\mathbf{j}\mathbf{b}_{N_r N_t N_f} \} \right),$$

where $\{ \mathbf{b}_1, \dots, \mathbf{b}_{N_r N_t N_f} \}$ is any \mathbb{C} -orthonormal basis of $\mathbb{C}^{N_r N_t N_f}$, and $c_1 = \dots = c_{N_r N_t N_f} = 1$.

Optimal CRB. Applying theorem 3, the optimal CRB of this model is then

$$\text{CRB}_{\min}(\boldsymbol{\theta}_{LS}) = \frac{\sigma^2 N_t^2 N_f^2 N_r^2}{P}. \quad (17)$$

It is attained for observation matrices of the form

$$\mathbf{M}_{LS} = \sqrt{\frac{P}{N_r N_t N_f}} (\mathbf{b}_1, \dots, \mathbf{b}_{N_r N_t N_f}). \quad (18)$$

Such matrices have $N_r N_t N_f = \frac{N_p}{2}$ columns, which, according to corollary 2 is minimal for identifiability to be possible.

Let us now particularize this result, considering observation matrix of the form given in (14). In the case of the least squares model, the variation space is structured as a tensor product of three simple vector spaces: $\mathcal{V}_{\boldsymbol{\theta}_{LS}} = \mathbb{C}^{N_r} \otimes \mathbb{C}^{N_t} \otimes \mathbb{C}^{N_f}$. This factorization allows to derive easily optimal sequence matrices \mathbf{X} . Indeed, in order to span $\mathbb{C}^{N_r} \otimes \mathbb{C}^{N_t} \otimes \mathbb{C}^{N_f}$, the columns of \mathbf{F} necessarily have to span \mathbb{C}^{N_f} and those of \mathbf{X} necessarily have to span \mathbb{C}^{N_t} . One simple way to satisfy these constraints is to take

$$\mathbf{F} = \mathbf{Id}_{N_f} \text{ and } \mathbf{X} = \sqrt{\frac{P_t}{N_t N_f}} \mathbf{S}, \quad (19)$$

where \mathbf{S} is unitary. This amounts to send N_t mutually orthogonal pilot sequences of duration $T = N_t$ on the different transmit antennas, on each subcarrier, and leads to

$$\text{CRB}_{\min}(\boldsymbol{\theta}_{LS}) = \frac{\sigma^2 N_t^2 N_f^2 N_r}{P_t}. \quad (20)$$

Equations (19) and (20) are nothing but a restatement of a well-known result [12], in a more general setting (several subcarriers). The method we propose here indeed allows to derive results for any linear channel model. However, it is more powerful since it generalizes also to nonlinear models, as shown in the next subsections.

B. Physical models

Another way to parameterize the channel is to assume that \mathbf{h} is the sum of L atomic channels corresponding to distinct physical paths, characterized by their direction of departure (DoD) \vec{u}_t^l , direction of arrival (DoA) \vec{u}_r^l , delay τ and complex gain β . Parameters of this kind of model are thus given by

$$\boldsymbol{\theta} = \left[(\Re(\beta_l), \Im(\beta_l), \vec{u}_r^l, \vec{u}_t^l, \tau)_{l=1}^L \right]^T. \quad (21)$$

The corresponding number of parameters is $N_p = mL$, where m is the number of real parameters to be estimated per physical path. The quantity m can take different values depending on the considered setting. The complex gain β_l always has to be estimated and accounts for two real parameters, so that $m \geq 2$. The DoD \vec{u}_t^l can either be assumed: (i) known (or irrelevant if $N_t = 1$) and thus account for zero real parameter, or (ii) unknown in which case it corresponds to a single real parameter if only the azimuth or elevation is estimated (for example in the case of a uniform linear array on the transmit side) or to two real parameters if both azimuth and elevation are estimated (for example in the case of a uniform planar array). The same goes for the DoA \vec{u}_r^l , replacing transmit

by receive in the previous sentence. The delay τ_l can either be assumed: (i) known (or irrelevant if $N_f = 1$) and thus account for zero real parameter, or (ii) unknown in which case it corresponds to a single real parameter. In summary, we have $2 \leq m \leq 7$ depending on the situation. In any case, the channel is expressed

$$\mathbf{h}(\boldsymbol{\theta}) = \sum_{l=1}^L \beta_l \mathbf{e}(\vec{u}_{r,l}, \vec{u}_{t,l}, \tau_l), \quad (22)$$

where $\mathbf{e}(\vec{u}_{r,l}, \vec{u}_{t,l}, \tau_l)$ is the atomic channel of unit norm corresponding to DoA $\vec{u}_{r,l}$, DoD $\vec{u}_{t,l}$ and delay τ_l , that we will denote \mathbf{e}_l below for compactness. This model is generally nonlinear (although it is linear with respect to the complex gains β_l). Such physical models are standard [32], [33]. A specific expression for \mathbf{e}_l can be found for example in [11].

1) *Linear physical model:* There is an interesting particular case of the physical model described above, in which all physical parameters except the ones corresponding to the gains are assumed known. This corresponds for example to the situation in which the directions and delays are estimated in the uplink, and only the gains have to be estimated in the downlink, thanks to the partial reciprocity of the channel. In [34], such a setting is studied, with $N_r = N_f = 1$ and an ULA at the base station, leading to a single parameter ϕ to estimate for the DoD. This leads to the physical model of the previous paragraph, with $m = 2$, and the parameters being

$$\boldsymbol{\theta}_{\text{LP}} \triangleq [(\Re(\beta_l)_{l=1}^L, \Im(\beta_l)_{l=1}^L)]^T. \quad (23)$$

We thus have $N_p = 2L$, and the channel is expressed

$$\mathbf{h}_{\text{LP}}(\boldsymbol{\theta}_{\text{LP}}) = (\mathbf{E}, \mathbf{jE}) \boldsymbol{\theta}_{\text{LP}}, \quad (24)$$

where $\mathbf{E} \in \mathbb{C}^{N_d \times L} \triangleq (\mathbf{e}(\phi_1), \dots, \mathbf{e}(\phi_L))$ is a matrix of atomic channels corresponding to DoDs estimated in the uplink, so that $\mathbf{e}(\phi_i)^H \mathbf{e}(\phi_j) = \delta_{ij}$. We call this model *linear physical model*, since knowing all the physical parameters except for the gains linearizes the estimation problem. Then, $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}_{\text{LP}}} = (\mathbf{E}, \mathbf{jE})$ and the variation space of this model can be decomposed according to lemma 1 as

$$\mathcal{V}_{\boldsymbol{\theta}_{\text{LP}}} = \text{span}_{\mathbb{R}}(\{\mathbf{e}_1, -\mathbf{j}\mathbf{e}_1, \dots, \mathbf{e}_L, -\mathbf{j}\mathbf{e}_L\}),$$

and $c_1 = \dots = c_L = 1$.

Optimal CRB. In the case where $N_r = N_f = 1$, (14) reduces to $\mathbf{M} = \mathbf{X}$ and consequently $P_t = P$. Thus, applying theorem 3 in that particular case, the optimal CRB of the linear physical model is

$$\text{CRB}_{\min}(\boldsymbol{\theta}_{\text{LP}}) = \frac{\sigma^2 L^2}{P_t}. \quad (25)$$

It is attained for pilot sequences of the form

$$\mathbf{X}_{\text{LP}} = \mathbf{M}_{\text{LP}} = \sqrt{\frac{P}{L}} (\mathbf{e}_1, \dots, \mathbf{e}_L). \quad (26)$$

Such matrices have $L = \frac{N_p}{2}$ columns, which, according to corollary 2 is minimal for identifiability to be possible, and is thus the minimal duration of the pilot sequences. Note that equations (25) and (26) allow to retrieve the variance of the channel estimators proposed in [34] (without proof of optimality).

2) *Nonlinear physical model:* In order to demonstrate the generality of the proposed results, let us give an example of their application to a simple nonlinear physical model. In [34], the DoDs are estimated in the uplink and considered unchanged at the time of the downlink channel estimation. This assumption could lead to a large bias in the model if the channel changes with time. In order to track channel paths more tightly, one could re-estimate these directions during the downlink channel estimation phase. To simplify the exposition, let us consider for this paragraph a channel made of a single path ($L = 1$). This would lead to a channel of the form

$$\mathbf{h}_{\text{PHY}}(\boldsymbol{\theta}_{\text{PHY}}) \triangleq \beta \mathbf{e}(\phi), \quad (27)$$

where the $N_p = 3$ real parameters to estimate are

$$\boldsymbol{\theta}_{\text{PHY}} \triangleq [(\Re(\beta), \Im(\beta), \phi)]^T. \quad (28)$$

$$\frac{\partial \mathbf{h}_{\text{PHY}}}{\partial \boldsymbol{\theta}_{\text{PHY}}} = \left(\mathbf{e}(\phi), \mathbf{j}\mathbf{e}(\phi), \beta \frac{\partial \mathbf{e}(\phi)}{\partial \phi} \right),$$

We thus have As stated in [10], when the antenna positions are expressed with respect to the centroid of the antenna array, we have $\mathbf{e}(\phi)^H \frac{\partial \mathbf{e}(\phi)}{\partial \phi} = 0$, so that the variation space of this model can be decomposed according to lemma 1 as

$$\mathcal{V}_{\boldsymbol{\theta}_{\text{PHY}}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{e}(\phi), -\mathbf{j}\mathbf{e}(\phi), \frac{\partial \mathbf{e}(\phi)}{\partial \phi} \right\} \right),$$

with $c_1 = 1$, and where $\frac{\partial \mathbf{e}(\phi)}{\partial \phi}$ is simply the normalized version of $\frac{\partial \mathbf{e}(\phi)}{\partial \phi}$.

Optimal CRB. Applying theorem 3 for this model (the odd version this time) yield

$$\text{CRB}_{\min}(\boldsymbol{\theta}_{\text{PHY}}) = \frac{\sigma^2}{P_t} \left(1 + \frac{1}{\sqrt{2}} \right)^2. \quad (29)$$

It is attained for pilot sequences of the form

$$\mathbf{X}_{\text{PHY}} = \mathbf{M}_{\text{PHY}} = \sqrt{\frac{P_t}{\sqrt{2} + 1}} \left(2^{\frac{1}{4}} \mathbf{e}(\phi), \frac{\partial \mathbf{e}(\phi)}{\partial \phi} \right). \quad (30)$$

Such matrices have $2 = \left\lceil \frac{N_p}{2} \right\rceil$ columns, which, according to corollary 2 is minimal for identifiability to be possible, and is thus the minimal duration of the pilot sequences. To the best of the authors' knowledge, this is the first time the minimal CRB and associated optimal pilot sequences are proposed for a nonlinear physical model, even in this very simple form.

One can notice that sending these optimal pilot sequences requires knowing ϕ . This is not realizable in practice since ϕ is a parameter to estimate. However, if a good estimate of ϕ is obtained in the uplink channel estimation phase, one can use it to build the downlink pilot sequences. Under the hypothesis that the DoD changes slowly, the built sequence is close to optimal and the variance of efficient estimators should be close to (29).

VI. CONCLUSION

In this paper, we studied the problem of estimating an object of interest parameterized according to a nonlinear model, based on noisy complex linear measurements, obeying (1).

The Cramér-Rao bound of such a general problem is established, showing its key dependency on an \mathbb{R} -vector space we called *variation space* (theorem 1). The CRB is shown to be proportional to the trace of the inverse compression of the observation matrix to the variation space (corollary 1).

A general result about \mathbb{R} -vector spaces is provided (lemma 1), which allows to decompose the variation space into \mathbb{C} -orthogonal subspaces. Such a decomposition proves useful in the study of observation matrices which is carried out next.

The identifiability conditions on the observation matrix are given (theorem 2), as well as a minimal number of measurements for identifiability to be possible (corollary 2).

The minimal CRB and associated observation matrices of minimal length are determined (theorem 3). They are shown to depend only on the observation power, the noise level and intrinsic properties of the variation space.

The results obtained for the general estimation problem are then particularized to MIMO channel estimation. It is shown that the general framework allows to retrieve well-known results, but also to derive optimal pilot sequences of minimal length in settings for which they had not been determined yet. The example of a very simple MIMO channel made of a single path is taken to illustrate this fact.

In the future, the theoretical results provided here could be applied to more practical MIMO systems, for example including hybrid precoding and combining [7], [35], [36]. They could allow to determine optimal pilot sequences in this context, as well as to quantify the suboptimality of existing or simpler schemes. They could also very well be applied outside the MIMO channel estimation scope, for any estimation problem whose observation model fits (1).

REFERENCES

- [1] S. Haykin, *Communication systems*. John Wiley & Sons, 2008.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
- [3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742–758, 2014.
- [5] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
- [6] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive mimo: the next wireless revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, 2014.
- [7] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.
- [8] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–89, 1945.
- [9] H. Cramér, *Mathematical Methods of Statistics*. Princeton university press, 1946, vol. 9.
- [10] L. Le Magoarou and S. Paquelet, "Parametric channel estimation for massive MIMO," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2018. [Online]. Available: <http://arxiv.org/abs/1710.08214>
- [11] L. Le Magoarou and S. Paquelet, "Performance of MIMO channel estimation with a physical model," *arXiv e-prints*, p. arXiv:1902.07031, Feb 2019.
- [12] M. Biguesh and A. B. Gershman, "Training-based mimo channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE transactions on signal processing*, vol. 54, no. 3, pp. 884–893, 2006.
- [13] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," in *Proceedings of the Annual Allerton Conference on Communication Control and Computing*, vol. 37. Citeseer, 1999, pp. 958–966.
- [14] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, 2003.
- [15] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for mimo ofdm systems in mobile wireless channels," *IEEE Transactions on signal processing*, vol. 51, no. 6, pp. 1615–1624, 2003.
- [16] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated mimo channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, 2004.
- [17] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, 2009.
- [18] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for fdd massive mimo systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.
- [19] S. Bazzi and W. Xu, "Downlink training sequence design for fdd multiuser massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp. 4732–4744, 2017.
- [20] —, "On the amount of downlink training in correlated massive mimo channels," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2286–2299, 2018.
- [21] S. Bazzi, S. Stefanatos, L. Le Magoarou, S. E. Hajri, M. Assaad, S. Paquelet, G. Wunder, and W. Xu, "Exploiting the massive mimo channel structural properties for minimization of channel estimation error and training overhead," *IEEE Access*, vol. 7, pp. 32 434–32 452, 2019.
- [22] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [24] W. H. Greub, *Linear algebra*. Springer Science & Business Media, 1975, vol. 23.
- [25] Åke Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of computation*, vol. 27, no. 123, pp. 579–594, 1973. [Online]. Available: <http://www.jstor.org/stable/2005662>
- [26] A. Van den Bos, "A cramér-rao lower bound for complex parameters," *IEEE Transactions on Signal Processing [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 42 (10), 1994.
- [27] D. Slepian, "Estimation of signal parameters in the presence of noise," *Transactions of the IRE Professional Group on Information Theory*, vol. 3, no. 3, pp. 68–89, 1954.
- [28] G. W. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale university, CT, USA, 1971.
- [29] O. Besson and Y. I. Abramovich, "On the fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, 2013.
- [30] P. R. Halmos, *A Hilbert space problem book*. Springer Science & Business Media, 1982, vol. 19.
- [31] R. A. Horn and C. R. Johnson, *Matrix Analysis: Second Edition*. Cambridge university press, 2012.
- [32] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.
- [33] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.
- [34] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for tdd/fdd massive mimo systems with spatial basis expansion model," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3170–3184, 2017.

- [35] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [36] A. M. Sayeed and J. H. Brady, *Millimeter-Wave MIMO Transceivers: Theory, Design and Implementation*. John Wiley & Sons, Ltd, 2016, pp. 231–253. [Online]. Available: <http://dx.doi.org/10.1002/9781119116493.ch10>