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 optimize the data rate/energy consumption tradeoff. In modern systems, the possibly large number of transmit/receive antennas and subcarriers makes this task difficult. Designing pilot sequences of reasonable size yielding good performance is thus critical. Classically, the number of pilots is reduced by viewing the channel as a random vector and assuming knowledge of its distribution. In practice, this requires estimating the channel covariance matrix, which can be computationally costly and not adapted to scenarios with high mobility. In this paper, an alternative view is considered, in which the channel is a function of unknown deterministic parameters. In this setting, the problem of designing optimal pilot sequences of smallest possible size is studied for *any* parametric channel model. To do so, the Cramér-Rao bound (CRB) for this general channel estimation problem is given, highlighting its key dependency on the introduced *variation space*. Then, the minimal size of pilot sequences and minimal value of the CRB are determined. Moreover, a general strategy to build optimal minimal length power constrained pilots sequences is given, based on an estimation of the variation space. The theoretical results are finally illustrated in a massive MIMO system context. They conveniently allow to retrieve well known previous results, but also to exhibit minimal length optimal pilot sequences for a new strategy based on a nonlinear physical model.

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].

Designing pilot sequences that lead to low estimation error and are of reduced size (compared to the channel dimension) is

Luc Le Magoarou and Stéphane Paquelet are both with bcom, Rennes, France. Contact addresses: `luc.lemagoarou@b-com.com`, `stephane.paquelet@b-com.com`.

thus a critical issue in massive MIMO systems. Classically, it has been done by considering the channel as a random vector whose distribution is known a priori, which naturally leads to the use of bayesian methods and estimators such as the linear minimum mean squared error (LMMSE). The minimal size of the pilot sequence is then determined by the effective rank of the channel covariance matrix [8], and efficient strategies such as the joint spatial division and multiplexing (JSDM) [9], [10] can be implemented. However, the main drawback of such methods is that it requires to estimate the channel covariance matrix, which can be computationally costly and unfit for high mobility scenarios (since the channel covariance then changes fast).

Another solution to envision channel estimation, which does not require covariance estimation, is to consider the channel as a function of parameters being deterministic unknown quantities, such as the channel coefficients or the directions and complex gains of the most significant propagation paths. This naturally leads to classical estimators based on the maximum likelihood (ML) principle. Following this line of thought, modern approaches have emerged [11] [12] that exploit some prior knowledge regarding the parameters to estimate in order to design efficient transmission strategies. In this setting, which quantity does determine the minimal size of pilot sequences? What is the best attainable performance? How to design optimal pilot sequences? Based on which a priori information?

Contributions. In this paper, we tackle these questions in a general unified way, for *any* parametric channel model (linear or not). Based on the Cramér-Rao bound (CRB) [13], [14] of the considered problem, we show that the crucial object for pilot sequences determination is the *variation space* of the channel, which is a notion we introduce. Identifiability conditions, the minimal size of pilot sequence, the minimal attainable variance and a strategy to build optimal pilot sequences of minimal size are given, all based on the variation space. We argue that the variation space is an object whose estimation may be simpler than that of the covariance. The theoretical part of the paper (which constitute the main contribution) is then illustrated on several MIMO channel models, and it allows to determine optimal pilot sequences and optimal performance for a new promising channel estimation strategy we propose for MIMO systems that operate in frequency division duplex (FDD) mode.

Related work. On the theoretical side, this paper is a generalization and an unification of many results obtained in the case of linear deterministic channel models (in which the model parameters are simply the channel coefficients), both in a

MIMO context [8], [15] and for multicarrier systems [16], [17], [18]. Indeed, the present analysis based on the variation space allows to treat simple linear models and more elaborate nonlinear physical channel models the same way. Another significant difference with prior work 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.

There is also a vast body of literature regarding optimal pilot sequences in a bayesian channel estimation setting, for which the channel is assumed to follow a known Gaussian [19], [20], [21], or more elaborate Gaussian mixture [22] distribution. 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 estimated distribution. On the other hand, the analysis of 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 [23], sequences are found by numerical optimization, minimizing a weighted sum of the channel estimation errors of each user. In [24] and [25], 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.

On the practical side, the analysis performed in this paper allows us to suggest a new transmission strategy for massive MIMO systems operating in FDD mode. It relies on the physical assumptions that the angles of arrival for channel propagation paths vary slowly and are reciprocal between the uplink and the downlink. This assumption is also at the origin of recent proposals [11] [12]. The strategy we propose is similar to this prior work in that is uses previous angle estimates (indifferently acquired in the uplink or downlink) to design pilot sequences. However, it is different since it allows to reestimate the angles at each step (with a small additional overhead), which leads to better performance lower bounds, as shown in section V.

Organization of the paper. The studied problem is formulated in section II. The notion of variation space is introduced, and an expression of the Cramér-Rao bound (CRB) based on it is given in section III. Identifiability conditions on the observation matrices and the minimal number of observations for which they can be fulfilled are given in section IV-A. In section IV-B, 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 in section IV-C, as well as an algorithm to build it based on an estimation of the variation space. These results are illustrated in section V, where it is shown that the proposed theoretical framework allows to retrieve well-known results previously established for linear models, but also to propose a new efficient transmission strategy for massive MIMO systems operating in FDD mode. For convenience and in order to keep the flow of the paper, most technical proofs are given in appendix.

Note that this paper is partially based on some of our previous work [26], [27], 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.

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 $\overrightarrow{a}$); the ith column of a matrix $\mathbf{A}$ by $\mathbf{a}_i$; its entry at the ith line and jth 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 $\mathrm{Tr}(\mathbf{A})$. The linear span of a set of vectors $\mathcal{A}$ and its dimension (if it is a vector space) are denoted: $\mathrm{span}_{\mathbb{R}}(\mathcal{A})$ and $\dim_{\mathbb{R}}(\mathcal{A})$ when considering linear combinations with real coefficients, or $\mathrm{span}_{\mathbb{C}}(\mathcal{A})$ and $\dim_{\mathbb{C}}(\mathcal{A})$ 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 $\mathbf{Id}$. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the standard complex gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. $\mathbb{E}(\cdot)$ denotes the expectation and $\mathrm{cov}(\cdot)$ the covariance of its argument.

A. Observations

We consider the general channel estimation setting where a channel $\mathbf{h} \in \mathbb{C}^{N_d}$ is to be estimated, N_d being the total number of complex dimensions of the channel. For example, in the case of a channel between N_t transmit antennas and N_r receive antennas on N_f subcarriers, we have $N_d = N_r N_t N_f$. We assume it is deterministic and follows a parametric model depending on N_p real parameters. It can then be seen as a function $\mathbf{h} : \mathbb{R}^{N_p} \to \mathbb{C}^{N_d}$ 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. We denote $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \triangleq \left(\frac{\partial \mathbf{h}}{\partial \theta_1}, \ldots, \frac{\partial \mathbf{h}}{\partial \theta_{N_p}} \right) \in \mathbb{C}^{N_d \times N_p}$ the complex gradient of the channel with respect to its real parameters.

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

$$\mathbf{y} = \mathbf{M}^H \mathbf{h} + \mathbf{n}, \tag{1}$$

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 $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{Id}_{N_m}\right)$, and $\mathbf{M} \in \mathbb{C}^{N_d \times N_m}$ is the matrix representing the measurement process, that we hereafter denote the *observation matrix*. It is entirely determined by the pilot sequences sent by the transmitter and the combining operations

done at the receiver. this way of expressing the observations is very general (an example in a generic MIMO wideband context is given in section V).

B. Estimation

The estimator of the parameters is a function mapping obervations to estimates, denoted $\hat{\boldsymbol{\theta}} : \mathbb{C}^{N_m} \to \mathbb{R}^{N_p}$. 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):

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

where the expectation is taken over the noise distribution, and the second line corresponds to the well-known bias-variance decomposition [28]. 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

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

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

C. Cramér-Rao bound

The variance of any unbiased estimator is bounded below by the Cramér-Rao bound [13], [14], so that

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

where the complex CRB [29] takes the form

$$\mathrm{CRB}(\boldsymbol{\theta}, \mathbf{M}) \triangleq \mathrm{Tr}\left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathbf{I}(\boldsymbol{\theta}, \mathbf{M})^{-1} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \right], \quad (3)$$

$\mathbf{I}(\boldsymbol{\theta}, \mathbf{M}) \in \mathbb{R}^{N_p \times N_p}$ being 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 $\mathbf{M}$. The observation defined in (1) follows a gaussian distribution,

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

so that the FIM is given by the Slepian-Bangs formula [30], [31], [32]:

$$\mathbf{I}(\boldsymbol{\theta}, \mathbf{M}) = \frac{2}{\sigma^2} \mathfrak{Re}\left\{ \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \mathbf{M}\mathbf{M}^H \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \right\}. \quad (4)$$

This finally yields

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) \geq \frac{\sigma^2}{2} \mathrm{Tr}\left[\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathfrak{Re}\left\{ \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \mathbf{M}\mathbf{M}^H \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \right\}^{-1} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \right].$$

In this paper, we analyze the right-hand side of this inequality. It is expressed in a compact way with help of the introduced variation space in section III. Then, viewed as a function of the observation matrix $\mathbf{M}$, it is optimized under a power constraint in section IV in order to exhibit optimal pilot sequences and

the associated minimal error, for *any* deterministic channel model. Note that the approach we propose can be generalized to deal with improper measurements [33], [34], [35], using a more general form of the Slepian-Bangs formula [36].

III. CRB BASED ON THE VARIATION SPACE

In this section, the notion of variation space, which plays a central role in the analysis we propose, is first introduced and discussed. Then, the CRB is expressed as a function of the variation space.

A. Variation space and related notions

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

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

be the variation space around the parameters value $\boldsymbol{\theta}$. This is the set corresponding to the potential directions of variation of the channel due to infinitesimal variations in the parameters value.

It is interesting to note that the variation space has the structure of an $\mathbb{R}$-vector space, since it contains all linear combinations of the columns of $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$ with *real* coefficients. However, $\mathcal{V}_{\boldsymbol{\theta}}$ is not necessarily a $\mathbb{C}$-vector space, since it does not contain all linear combinations of the columns of $\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}$ with *complex* coefficients (because we consider real parameters). This subtle distinction will play a major role in the subsequent analysis, as evidenced in section IV. To makes things clearer, we define also the real inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}} \triangleq \mathfrak{Re}\{\mathbf{x}^H \mathbf{y}\}.$$

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be real-orthogonal (or $\mathbb{R}$-orthogonal) if $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}} = 0$. Let $\mathcal{E}$ be a $\mathbb{R}$-vector space, we denote $\dim_{\mathbb{R}}(\mathcal{E})$ its dimension with the scalar field $\mathbb{R}$. Similarly, we denote the classical complex inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{C}} \triangleq \mathbf{x}^H \mathbf{y},$$

and two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be complex-orthogonal ((or $\mathbb{C}$-orthogonal)) if $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{C}} = 0$. Let $\mathcal{F}$ be a $\mathbb{C}$-vector space, we denote $\dim_{\mathbb{C}}(\mathcal{F})$ its dimension with the scalar field $\mathbb{C}$. Note that any $\mathbb{C}$-vector space is also a $\mathbb{R}$-vector space of doubled dimension, so that $\dim_{\mathbb{R}}(\mathcal{F}) = 2\dim_{\mathbb{C}}(\mathcal{F})$, but the converse is *not* true (a $\mathbb{R}$-vector space is in general not a $\mathbb{C}$-vector space).

B. Expression of the CRB

In the general setting we consider, the CRB can be expressed in a very simple way, depending only on the variation space, the observation matrix and the noise level. The obtained form of the CRB will prove very useful in section IV in order to optimize the observation matrix, and thus the sent pilot sequences. It is given by the following theorem.

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

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

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

Proof. Let us start from (3) and (4). In this basic form, the FIM is difficult to invert, because it involves the real part of a complex matrix. In [26], 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 $\langle ., .\rangle_{\mathbb{R}}$ to decompose it as

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} = \mathbf{U}\mathbf{R}, \tag{5}$$

where $\mathbf{U} \in \mathbb{C}^{N_d \times K}$ is a matrix whose columns are $\mathbb{R}$-orthonormal (meaning that $\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{U}\right\} = \mathbf{Id}_K$), $\mathbf{R} \in \mathbb{R}^{K \times N_p}$ 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

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

since $\mathfrak{Re}\{\mathbf{A}^H\mathbf{B}\mathbf{A}\} = \mathbf{A}^T\mathfrak{Re}\{\mathbf{B}\}\mathbf{A}$ 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

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

$$= \frac{\sigma^2}{2}\mathrm{Tr}\left[\mathbf{U}\mathbf{R}\mathbf{R}^{-1}\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\right\}^{-1}\mathbf{R}^{-T}\mathbf{R}^T\mathbf{U}^H\right]$$

$$= \frac{\sigma^2}{2}\mathrm{Tr}\left[\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\right\}^{-1}\right].$$

In order to conclude, one can remark that

$$\mathrm{Tr}\left[\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\right\}^{-1}\right] = \mathrm{Tr}\left[\mathfrak{Re}\left\{\mathbf{B}^T\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\mathbf{B}\right\}^{-1}\right]$$

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

This theorem allows to express the CRB in a way that is particularly suited to the determination of optimal observation matrices that is carried out in section IV. Moreover, the CRB can be given an even simpler form. Indeed, it 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 $\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\right\}$ can be given a nice interpretation. Indeed, the orthogonal projection $\mathbf{P}_{\mathcal{V}_{\boldsymbol{\theta}}}\mathbf{z}$ of any vector $\mathbf{z}$ onto $\mathcal{V}_{\boldsymbol{\theta}}$ is expressed

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

so that $\mathfrak{Re}\{\mathbf{U}^H\mathbf{z}\}$ 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 $\mathfrak{Re}\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{t}\} = \mathfrak{Re}\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\}\mathbf{r}$ 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 $\mathfrak{Re}\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\}$ 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 [37, 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 $\mathbf{B} : \mathcal{K} \to \mathcal{K}$ be a linear operator on $\mathcal{K}$. The linear operator $\mathbf{A} : \mathcal{H} \to \mathcal{H}$ 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 $\mathbf{A} : \mathcal{H} \to \mathcal{H}$ where $\mathcal{H}$ is a $\mathbb{K}$-vector space ($\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$), we define its trace as

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

where $\{\mathbf{v}_1, \ldots, \mathbf{v}_{N_p}\}$ 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 and more intrinsic 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*

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

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 section.

IV. OPTIMIZED OBSERVATION MATRICES

In this section, the objective is to optimize the observation matrix $\mathbf{M}$ with respect to the particular form of the CRB given in theorem 1. We first give identifiability conditions, which allow to determine a minimal number of observations. Then, the optimal CRB and associated observation matrices of minimal size are given. Finally, we give a practical algorithm to design observation matrices based on an estimation of the variation space.

A. Identifiability

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

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

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

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

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 III-B. When this condition is fulfilled, identifiability holds if and only if the matrix $\mathfrak{Re}\left\{\mathbf{U}^H\mathbf{M}\mathbf{M}^H\mathbf{U}\right\}$ is invertible. This matrix being symmetric, it is invertible if and only if

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

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

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

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

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

$\square$

Interpretations. The first identifiability condition $\dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}}) = N_p$ means that the columns of $\frac{\partial\mathbf{h}}{\partial\boldsymbol{\theta}}$ 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 $\dim_{\mathbb{R}}(\mathcal{V}_{\boldsymbol{\theta}}) \leq 2N_d$, 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 channel.

Then, if the first condition is fulfilled, the second condition $\mathcal{V}_{\boldsymbol{\theta}} \cap im_{\mathbb{C}}(\mathbf{M})^{\perp} = \{\mathbf{0}\}$ means that no nonzero vector in the space of variations $\mathcal{V}_{\boldsymbol{\theta}}$ 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:

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

which is possible only if the $\mathbb{R}$-dimension of $\ker(\mathbf{M}^H)$ plus the $\mathbb{R}$-dimension of $\mathcal{V}_{\boldsymbol{\theta}}$ 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, $\dim_{\mathbb{R}}(\ker(\mathbf{M}^H)) = 2N_d - \dim_{\mathbb{R}}(im_{\mathbb{C}}(\mathbf{M}^H))$ (rank-nullity theorem), so that we end up with

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

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

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

which proves the result. $\square$

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.

B. 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:

$$\begin{aligned} \underset{\mathbf{M}}{\text{minimize}} & \quad \text{CRB}(\boldsymbol{\theta}, \mathbf{M}), \\ \text{subject to} & \quad \|\mathbf{M}\|_F^2 = P. \end{aligned} \quad (6)$$

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 and not directly equal to the transmitted power. The two quantities are linked in section V.

1) Decomposition of the variation space: The expression of the CRB given in theorem 1 is valid for any $\mathbb{R}$-orthogonal basis of $\mathcal{V}_{\boldsymbol{\theta}}$. In order to ease optimization, we exhibit here a specific basis with useful properties. To do so, let us state the following lemma that allows to decompose $\mathcal{V}_{\boldsymbol{\theta}}$ into a direct sum of $\mathbb{C}$-orthogonal subspaces.

Lemma 1. *(i) Any $\mathbb{R}$-vector space $\mathcal{E}$ of dimension d that belongs to a $\mathbb{C}$-vector space $\mathcal{F}$ (containing $j\mathcal{E}$) can be decomposed into the direct sum of subspaces of dimension 2 (and possibly a subspace of dimension one if d is odd) that are mutually $\mathbb{C}$-orthogonal. (ii) The subspaces of the aforementioned decomposition belong to eigenspaces of $\mathbf{P}_{\mathcal{E}} \circ \mathbf{P}_{j\mathcal{E}}$, where $\mathbf{P}_{\mathcal{E}}$ (resp. $\mathbf{P}_{j\mathcal{E}}$) is the orthogonal projection onto $\mathcal{E}$ (resp. $j\mathcal{E}$).*

Proof. This lemma is proven in appendix A. $\square$

Applying lemma 1 to the variation space $\mathcal{V}_{\boldsymbol{\theta}}$ (assuming it is of dimension N_p and N_p is even), it is possible to decompose it as

$$\mathcal{V}_{\boldsymbol{\theta}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{v}_1, \mathbf{w}_1, \ldots, \mathbf{v}_{\frac{N_p}{2}}, \mathbf{w}_{\frac{N_p}{2}} \right\} \right) \qquad (7)$$

where $\mathbf{v}_m^H \mathbf{v}_n = \delta_{mn}$, $\mathbf{w}_m^H \mathbf{w}_n = \delta_{mn}$ and $\mathbf{v}_m^H \mathbf{w}_n = -\delta_{mn}jc_m$ (with $0 \leq c_m \leq 1$, and δ being the Kronecker symbol). The quantities c_m can be seen as the lack of $\mathbb{C}$-orthogonality of the $\mathbb{R}$-orthogonal basis $\left\{ \mathbf{v}_1, \mathbf{w}_1, \ldots, \mathbf{v}_{\frac{N_p}{2}}, \mathbf{w}_{\frac{N_p}{2}} \right\}$. Let us introduce the matrix

$$\mathbf{V} \triangleq \left(\mathbf{v}_1, \mathbf{w}_1, \ldots, \mathbf{v}_{\frac{N_p}{2}}, \mathbf{w}_{\frac{N_p}{2}} \right) \qquad (8)$$

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

$$\mathcal{V}_{\boldsymbol{\theta}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{v}_1, \mathbf{w}_1, \ldots, \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor}, \mathbf{w}_{\lfloor \frac{N_p}{2} \rfloor}, \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor + 1} \right\} \right), \qquad (9)$$

where $\mathbf{v}_m^H \mathbf{v}_n = \delta_{mn}$, $\mathbf{w}_m^H \mathbf{w}_n = \delta_{mn}$ and $\mathbf{v}_m^H \mathbf{w}_n = -\delta_{mn}jc_m$, and the matrix $\mathbf{V}$ can be built the same way. Said differently, this result means that for any matrix $\mathbf{U}$ whose columns form an $\mathbb{R}$-orthogonal basis of $\mathcal{V}_{\boldsymbol{\theta}}$, there exists a real orthogonal matrix $\mathbf{B}$ such that

$$\mathbf{B}^T \mathfrak{Im}\{\mathbf{U}^H \mathbf{U}\} \mathbf{B} = \begin{pmatrix} 0 & -c_1 & & & \\ c_1 & 0 & & & \\ & & 0 & -c_2 & \\ & & c_2 & 0 & \\ & & & & \ddots \end{pmatrix} \triangleq \boldsymbol{\Gamma},$$

(10)

and then $\mathbf{V} = \mathbf{UB}$. In practice, the matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$ can be obtained by computing the real Schur decomposition of the matrix $\mathfrak{Im}\{\mathbf{U}^H \mathbf{U}\}$ and reordering the diagonal blocks.

2) Optimal CRB and observation matrices: Using this decomposition allows us to state the main result of this paper in the following theorem.

Theorem 3. *The minimal value of the CRB is*

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

where the scalars c_k are defined at (7) (resp. (9)) and $\epsilon = 0$ (resp. $\epsilon = 1$) if N_p is even (resp. odd).

It is attained with the observation matrix of minimal size

$$\mathbf{M} = \sqrt{\frac{P}{C}} \left(\frac{\mathbf{v}_1 + j\mathbf{w}_1}{(1+c_1)^{\frac{3}{4}}}, \ldots, \frac{\mathbf{v}_{\frac{N_p}{2}} + j\mathbf{w}_{\frac{N_p}{2}}}{(1+c_{\frac{N_p}{2}})^{\frac{3}{4}}} \right),$$

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 (7) if N_p is even, and with

$$\mathbf{M} = \sqrt{\frac{P}{C}} \left(\frac{\mathbf{v}_1 + j\mathbf{w}_1}{(1+c_1)^{\frac{3}{4}}}, \ldots, \frac{\mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor} + j\mathbf{w}_{\lfloor \frac{N_p}{2} \rfloor}}{(1+c_{\lfloor \frac{N_p}{2} \rfloor})^{\frac{3}{4}}}, \mathbf{v}_{\lfloor \frac{N_p}{2} \rfloor + 1} \right),$$

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 (9) if N_p is odd.

Proof. This theorem is proven in appendix B. $\square$

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

$$\frac{\sigma^2 N_p^2}{4P} \leq \text{CRB}_{\min}(\boldsymbol{\theta}) \leq \frac{\sigma^2 N_p^2}{2P}, \qquad (11)$$

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 $j\mathcal{V}_{\boldsymbol{\theta}}$).

C. Observation matrix design

Based on theorem 3, it is possible to build optimal observation matrices of minimal size $\lceil \frac{N_p}{2} \rceil$, provided the variation space is known. Since the variation space depends itself on the parameters to estimate, this result may seem of little use. However, in some cases, an estimation $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$ of the variation space can be obtained. For example, this is the case for MIMO systems operating in FDD mode using a physical model, where the variation space can be determined based on the previous uplink or downlink channel estimates since it depends only on the directions of arrival of the channel paths, which vary in general slowly. This interesting application is studied in details in the next section.

Algorithm 1 Observation matrix determination (N_p even)

Input: An estimate $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$ of the variation space (i.e. a matrix $\mathbf{G}$ whose columns are a generating family of $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$ with real scalars), the observation power P.

1: Find a matrix $\mathbf{U}$ whose columns form an $\mathbb{R}$-orthonormal basis of $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$:

  Real QR decomposition (5): $\begin{pmatrix} \mathfrak{Re}\{\mathbf{G}\} \\ \mathfrak{Im}\{\mathbf{G}\} \end{pmatrix} = \mathbf{QR}$,

  $\mathbf{U} \leftarrow \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \mathbf{R}^{-1}$

2: Decompose $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$ as in lemma 1:

  Real Schur decomposition (10): $\mathfrak{Im}\{\mathbf{U}^H \mathbf{U}\} = \mathbf{B}\boldsymbol{\Gamma}\mathbf{B}^T$,

  $c_1 \leftarrow \gamma_{21}$, $c_2 \leftarrow \gamma_{43}, \ldots, c_{\frac{N_p}{2}} \leftarrow \gamma_{N_p, N_p-1}$,

  $\mathbf{V} \leftarrow \mathbf{UB}$

3: Build the observation matrix according to theorem 3:

$$\mathbf{S} \leftarrow \begin{pmatrix} 1 & 0 & \cdots \\ j & 0 & \cdots \\ 0 & 1 & \cdots \\ 0 & j & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \in \mathbb{C}^{N_p \times \frac{N_p}{2}}, \; C \leftarrow 2 \sum_{l=1}^{\frac{N_p}{2}} \frac{1}{\sqrt{1+c_l}},$$

$$\mathbf{D} \leftarrow \sqrt{\frac{P}{C}} \begin{pmatrix} \frac{1}{(1+c_1)^{\frac{3}{4}}} & 0 & \\ 0 & \frac{1}{(1+c_2)^{\frac{3}{4}}} & \\ & & \ddots \end{pmatrix} \in \mathbb{C}^{\frac{N_p}{2} \times \frac{N_p}{2}},$$

  $\mathbf{M} \leftarrow \mathbf{VSD}$

Output: The observation matrix $\mathbf{M} \in \mathbb{C}^{N_d \times \frac{N_p}{2}}$ that is optimal with respect to $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$.

The strategy we propose to build observation matrices based on an estimate of the variation space is given in algorithm 1 for an even number of parameters (the algorithm is almost the same for an odd number, except for the third step being slightly modified according to theorem 3). It comprises three steps. The first one amounts to find an $\mathbb{R}$-orthogonal basis of $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$. The second one corresponds to apply the decomposition of lemma 1 to $\hat{\mathcal{V}}_{\boldsymbol{\theta}}$. Finally the third one uses the result of theorem 3 based on the aforementioned decomposition to build the observation matrix. Overall, the computational complexity of this algorithm is $\mathcal{O}(N_d N_p^2)$.

V. Illustrations of the results

Let us now illustrate the applicability of the presented theoretical results, whose ultimate goal is to facilitate the design of optimal and short pilot sequences for any deterministic channel model. To do so, we consider a massive MIMO system and compare various models.

Scenarios. Three scenarios are chosen to apply our results:

- First, in section V-A, the classical least squares model is studied in the proposed framework, showing that our approach is trivial in that case and allows to retrieve previous results. This model makes no use of any a priori information, so that pilot sequences have to be as long as the channel dimension and the optimal CRB is proportional to the square of the channel dimension.
- Then, in section V-B, physical models are investigated, which allow to design shorter pilot sequences and theoretically lead to better performance, thanks to the lower number of parameters to estimate. In that case, our framework allows to theoretically justify previous approaches based on estimates of the channel directions of arrival (DoA), assuming their reciprocity [12] or time persistence [11]. These approaches lead to the length of optimal pilot sequences being proportional to the number of dominant paths and the optimal CRB being proportional to the square of this quantity. However, such methods induce a biased model, due to the fact that DoA estimates are considered perfect and kept fixed, which lacks robustness in practical scenarios.
- Finally, in section V-C, still in the physical models context, we show that our result allows to suggest a new strategy that is more robust to the DoA estimation error. It is based on an update of the DoA estimates, and comes with better theoretical guarantees than previous approaches. Indeed, it corrects their bias and causes only a small increase of the optimal CRB and pilot sequence length (due to the DoA update). This approach is compared numerically to the one of section V-B, taking into account the DoA estimation error, showing empirically its advantage in terms of robustness.

Setting. In the general case where the channel to estimate is between N_t transmit antennas and N_r receive antennas, on N_f subcarriers, the channel is a complex 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, and can be linked directly to the sent pilot sequence. 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

$$\mathbf{M} = \mathbf{Id}_{N_r} \otimes \mathbf{X} \otimes \mathbf{F} \in \mathbb{C}^{N_r N_t N_f \times N_r T N_{\mathrm{ps}}}, \qquad (12)$$

where $\mathbf{F} \in \{0,1\}^{N_f \times N_{\mathrm{ps}}}$ 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 $N_m = N_r T N_{\mathrm{ps}}$, and the transmitted power is $P_t \triangleq N_{\mathrm{ps}} \|\mathbf{X}\|_2^2$. On the other hand, the observation power which is constrained in the optimization problem (6), is expressed $P = \|\mathbf{M}\|_2^2 = N_r N_{\mathrm{ps}} \|\mathbf{X}\|_2^2$. 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.

In this section, let us perform the analysis considering a massive MIMO setting where the base station is equipped with an uniform linear array (ULA) with half-wavelength separated antennas aligned with the y-axis, and user terminals are equipped with a single antenna ($N_r = 1$). Let us also consider a single subcarrier ($N_f = 1$) for ease of exposition, but note that the study straightforwardly extends to the multi-carrier case. These assumptions directly imply the direct equality of the observation matrix and the pilot sequences matrix:

$$\mathbf{M} = \mathbf{X}, \; P = P_t, \qquad (13)$$

which greatly simplifies the following subsections.

A. Application to the least squares model

The most direct and simple way to parameterize the channel with no a priori is to take as $N_p = 2N_t = 2N_d$ parameters the real and imaginary parts of the channel entries,

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

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

$$\mathbf{h}_{\mathrm{LS}}(\boldsymbol{\theta}_{\mathrm{LS}}) = (\mathbf{Id}, \mathrm{j}\mathbf{Id}) \, \boldsymbol{\theta}_{\mathrm{LS}}. \qquad (14)$$

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

$$\mathcal{V}_{\boldsymbol{\theta}_{\mathrm{LS}}} = \mathbb{C}^{N_t}. \qquad (15)$$

Regarding the framework proposed in this paper, this is a trivial case since the variation space is independent of the parameters value, due to the linearity of the model. Consequently, no estimation of the variation space is needed to design optimal pilot sequences. Indeed, this particular variation space can decomposed according to lemma 1 as

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

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

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

$$\text{CRB}_{\min}(\boldsymbol{\theta}_{\text{LS}}) = \frac{\sigma^2 N_t^2}{P_t}. \tag{16}$$

It is attained for observation matrices (pilot sequences) of the form

$$\mathbf{X}_{\text{LS}} = \mathbf{M}_{\text{LS}} = \sqrt{\frac{P}{N_t}} \left(\mathbf{b}_1, \ldots, \mathbf{b}_{N_t} \right). \tag{17}$$

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

Equations (16) and (17) are nothing but a restatement of a well-known result [8]. 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. Application to 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) $\overrightarrow{u_t}$, direction of arrival (DoA) $\overrightarrow{u_r}$, delay τ and complex gain β. Parameters of this kind of model are thus given by

$$\boldsymbol{\theta} = \left[\left(\mathfrak{Re}(\beta_l), \mathfrak{Im}(\beta_l), \overrightarrow{u_{r,l}}, \overrightarrow{u_{t,l}}, \tau_l \right)_{l=1}^L \right]^T. \tag{18}$$

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.

In the massive MIMO setting studied here, (18) reduces to

$$\boldsymbol{\theta}_{\text{PHY}} = \left[\left(\mathfrak{Re}(\beta_l), \mathfrak{Im}(\beta_l), \phi_l \right)_{l=1}^L \right]^T, \tag{19}$$

where ϕ_l is the azimuth angle for the l-th physical path. The downlink channel is then expressed

$$\mathbf{h}_{\text{PHY}}(\boldsymbol{\theta}_{\text{PHY}}) = \sum_{l=1}^L \beta_l \mathbf{e}(\phi_l), \tag{20}$$

where $\mathbf{e}(\phi) = \frac{1}{\sqrt{N_t}} (e^{-j\frac{2\pi}{\lambda}\left(\frac{N_t-1}{4}\right)\sin\phi}, \ldots, e^{j\frac{2\pi}{\lambda}\left(\frac{N_t-1}{4}\right)\sin\phi})^T$ is the steering vector associated with azimuth ϕ (for an even number of antennas). This model is nonlinear (although it is linear with respect to the complex gains β_l). Such physical models are quite standard [38], [39] and successful due to the possibility to take L small (typically less than ten) with good channel modeling accuracy (the channel is then called sparse). The variation space for such physical model is expressed

$$\mathcal{V}_{\boldsymbol{\theta}_{\text{PHY}}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{e}(\phi_l), -j\mathbf{e}(\phi_l), \frac{\partial \mathbf{e}(\phi_l)}{\partial \phi_l} \right\}_{l=1}^L \right). \tag{21}$$

As opposed to the least squares case, this variation space depends on the parameter value, so that it is not possible to build optimal short-length pilot sequences without relying on some prior information about the parameters. However, it is interesting to notice that the variation space depends only on the azimuth angles $\phi_1, \ldots, \phi_L$ and not on the path gains. Fortunately, the azimuth angles vary much more slowly than

the path gains and are the same for the uplink and downlink channels, so that one can hope to acquire good estimates of them in order to design pilot sequences.

Angle-constrained estimation strategy. Under these assumptions, it has already been proposed to design short length downlink pilot sequences considering azimuth angles are already known by the base station, either thanks to uplink channel estimates [12] (using the channel angle reciprocity), or thanks to previous downlink channel estimates [11] (using angle time persistence). In both cases, the communication workflow can be decomposed in two separate phases as follows:

1) **Paths estimation:** estimate all dominant path angles, either sending long downlink pilot sequences [11] (duration at least equal to the number of transmit antennas) or using the uplink pilots [12].
2) **Channel tracking:** track the downlink channel, by estimating path gains while keeping path angles fixed, sending short downlink pilot sequences (duration proportional to the number of dominant paths).

Note that the paths estimation phase (step 1) has to be done periodically, in order to take into account changes in path angles, the disappearance of old paths and the appearance of new paths. Otherwise, the link quality may degrade much due to the aging of the path angles information.

Let us analyze these methods that we call angle-constrained within our framework. We focus here on the channel tracking phase. Denoting $\hat{\phi}_1, \ldots, \hat{\phi}_L$ the estimated azimuths and $\hat{\mathbf{E}} \in \mathbb{C}^{N_d \times L} \triangleq (\mathbf{e}(\hat{\phi}_1), \ldots, \mathbf{e}(\hat{\phi}_L))$ the matrix of corresponding steering vectors, they amount to simplify the estimation problem considering only the path gains remain to estimate, yielding parameters

$$\boldsymbol{\theta}_{\text{AC}} \triangleq \left[\left(\mathfrak{Re}(\beta_l)_{l=1}^L, \mathfrak{Im}(\beta_l)_{l=1}^L \right) \right]^T. \tag{22}$$

It thus corresponds to have $N_p = 2L$, and the channel is expressed

$$\mathbf{h}_{\text{AC}}(\boldsymbol{\theta}_{\text{AC}}) = \left(\hat{\mathbf{E}}, j\hat{\mathbf{E}} \right) \boldsymbol{\theta}_{\text{AC}}. \tag{23}$$

Considering this model, the variation space simplifies to

$$\mathcal{V}_{\boldsymbol{\theta}_{\text{AC}}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{e}(\hat{\phi}_1), -j\mathbf{e}(\hat{\phi}_1), \ldots, \mathbf{e}(\hat{\phi}_L), -j\mathbf{e}(\hat{\phi}_L) \right\} \right).$$

Optimal CRB. Thus, applying theorem 3 in that particular case, the optimal CRB of the angle-constrained physical model is bounded as

$$\frac{\sigma^2 L^2}{P_t} \leq \text{CRB}_{\min}(\boldsymbol{\theta}_{\text{PHY}}) \leq 2 \times \frac{\sigma^2 L^2}{P_t}, \tag{24}$$

where the lower bound is attained if the columns of $\hat{\mathbf{E}}$ are mutually orthogonal (which is assumed in [11], [12]), in which case the decomposition of lemma 1 is trivial with $c_1 = \cdots = c_L = 1$. In that case, it is interesting to compare this CRB to the CRB of the least squares model (16): using the physical model allows to divide by $\left(\frac{N_t}{L} \right)^2$ the minimal attainable variance. This potentially huge gain is attained provided the azimuth estimates are perfect, and for downlink pilot sequences of the form

$$\mathbf{X}_{\text{AC}} = \mathbf{M}_{\text{AC}} = \sqrt{\frac{P_t}{L}} \left(\mathbf{e}(\hat{\phi}_1), \ldots, \mathbf{e}(\hat{\phi}_L) \right). \tag{25}$$

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, that corresponds to the number of estimated channel paths. In practice, the number of estimated paths is small (rarely more than ten), so that such sequences are short when compared to the ones required when using the least squares model (of length N_t). Note that equations (24) and (25) allow to retrieve the variance and pilot sequences proposed in [11] and [12] (without proof of optimality).

The bias problem. It is important to mention that the model simplification implied by the angle-constrained estimation strategy induces bias. Indeed, the obtained channel estimate $\mathbf{h}_{\text{AC}}(\hat{\boldsymbol{\theta}}_{\text{AC}})$ is constrained to belong to the range of the matrix $\hat{\mathbf{E}}$, which is not necessarily the case for the true channel $\mathbf{h}$ since the azimuth estimation is not error free and the azimuth angles may have changed since their estimation. We indeed have

$$\left\| \mathbf{h} - \mathbf{h}_{\text{AC}}(\hat{\boldsymbol{\theta}}_{\text{AC}}) \right\|_2^2 \geq \left\| \mathbf{h} - \hat{\mathbf{E}}(\hat{\mathbf{E}}^H \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^H \mathbf{h} \right\|_2^2,$$

$\hat{\mathbf{E}}(\hat{\mathbf{E}}^H \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^H$ being the matrix representing orthogonal projection onto the range of $\hat{\mathbf{E}}$. Combined with the CRB, we get the following bound on the MSE for the angle-constrained channel estimation strategy:

$$\text{MSE}(\hat{\boldsymbol{\theta}}_{\text{AC}}) \geq \max\left(\left\| \mathbf{h} - \hat{\mathbf{E}}(\hat{\mathbf{E}}^H \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^H \mathbf{h} \right\|_2^2, \frac{\sigma^2 L^2}{P_t} \right).$$
(26)

A high level summary of the channel tracking phase of the angle-constrained channel estimation strategy for a single user in given in algorithm 2. It fits into a multi-user framework and a complete transmission workflow nicely, as explained in [11], [12], but we do not give too much details on this here since the present illustration focuses only on the downlink channel tracking phase.

Algorithm 2 High level summary of channel tracking phase of the angle-constrained estimation strategy

---

Input: Estimates $\hat{\phi}_1, \ldots, \hat{\phi}_L$ of the azimuth angles (obtained through previous downlink or uplink channel estimates).
 1: Build $\hat{\mathbf{E}} = (\mathbf{e}(\hat{\phi}_1), \ldots, \mathbf{e}(\hat{\phi}_L))$.
 2: Send pilot sequences $\mathbf{X}_{\text{AC}} = \sqrt{\frac{P_t}{L}} \left(\mathbf{e}(\hat{\phi}_1), \ldots, \mathbf{e}(\hat{\phi}_L) \right)$ of duration L and receive user feedback to build observations following (1).
 3: Estimate path gains $\hat{\beta}_1, \ldots, \hat{\beta}_L$.
 4: Estimate channel $\mathbf{h}_{\text{AC}}(\hat{\boldsymbol{\theta}}_{\text{AC}}) = \sum_{l=1}^L \hat{\beta}_l \mathbf{e}(\hat{\phi}_l)$.
Output: $\text{MSE}(\hat{\boldsymbol{\theta}}_{\text{AC}}) \geq \max\left(\left\| \mathbf{h} - \hat{\mathbf{E}}(\hat{\mathbf{E}}^H \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^H \mathbf{h} \right\|_2^2, \frac{\sigma^2 L^2}{P_t} \right)$

C. A new channel estimation strategy for physical models

Let us now propose another strategy for the channel tracking phase that is more robust to the angle estimation error because it does not suffer from the same bias problem as the angle-constrained strategy. Moreover, the proposed strategy also leads to short pilot sequences thanks to the use of theorem 3

and algorithm 1 for pilot design, and incurs only a small variance increase compared to the angle-constrained strategy. We indeed argue that is is possible to make better use of previously acquired azimuth angles estimates $\hat{\phi}_1, \ldots, \hat{\phi}_L$ than to bias the model by considering them perfectly estimated and to constrain the channel estimate to belong to the range of $\hat{\mathbf{E}}$. Indeed, one could instead use the azimuth estimates to design pilot sequences according to algorithm 1, so as to update the angle estimation while estimating the channel. This amounts to consider the physical model of (19) and (20) as is, without fixing the angles. This way, the variation space is expressed as in (21).

Optimal CRB. Applying theorem 3, the optimal CRB of this full physical model is thus bounded as

$$\frac{9}{4} \times \frac{\sigma^2 L^2}{P_t} \leq \text{CRB}_{\min}(\boldsymbol{\theta}_{\text{PHY}}) \leq \frac{9}{2} \times \frac{\sigma^2 L^2}{P_t}.$$
(27)

These bounds are obtained by injecting $N_p = 3L$ into (11). The exact value of the CRB depends on the values of the scalars $c_1, \ldots, c_{\lfloor \frac{N_p}{2} \rfloor}$. These bounds are $\frac{9}{4}$ times larger than the CRB bounds of the angle-constrained strategy (24), because of the larger number of real parameters to estimate ($3L$ instead of $2L$). It is attained for perfect estimates of the azimuth angles. A natural estimate of the variation space based on azimuth estimates reads

$$\hat{\mathcal{V}}_{\boldsymbol{\theta}_{\text{PHY}}} = \text{span}_{\mathbb{R}} \left(\left\{ \mathbf{e}(\hat{\phi}_l), -j\mathbf{e}(\hat{\phi}_l), \frac{\partial \mathbf{e}(\hat{\phi}_l)}{\partial \hat{\phi}_l} \right\}_{l=1}^L \right).$$
(28)

This estimate can be used directly to design pilot sequences according to algorithm 1. The proposed channel tracking strategy is summarized in algorithm 3. It fits into a multi-user framework and a complete transmission strategy exactly as its angle-constrained counterpart, to which it is pretty similar. The two main differences with algorithm 2 are the slightly longer pilot sequences ($\lceil \frac{3L}{2} \rceil$ instead of L) and the fact that azimuth angles are not kept fixed but updated. The update of the angles estimates can be done for example using the maximum likelihood principle with exhaustive testing of angles sampling uniformly a small region around the previous angles estimates, or with a few steps of gradient descent initialized at the previous angles estimates. This update has the effect of suppressing the bias and leads to a lower bound on the MSE comprising only variance, which vanishes at high SNR. The two evoked strategies (angle-constrained and proposed) actually differ in the way prior information is used. For the angle-constrained strategy, channel estimates themselves are based on fixed and previously acquired angle estimates, while for the proposed strategy, only the downlink pilot sequences are based on previous angle estimates, which is more robust as shown empirically in the following paragraph, because pilot sequences stay quasi-optimal if the true angles stay in a neighborhood of the estimates. Moreover, the proposed strategy allows an update of the angle estimates, which may reduce the frequency at which the path estimation phase (which is computationally heavy) has to be carried out.

Comparison of strategies with azimuth error. Let us now compare numerically the proposed strategy with the angle-constrained estimation strategy, taking into account the az-

---

Algorithm 3 High level summary of channel tracking phase of the proposed estimation strategy (L even)

Input: Estimates $\hat{\phi}_1, \ldots, \hat{\phi}_L$ of the azimuth angles (obtained through previous downlink or uplink channel estimates).

1: Build $\hat{\mathcal{V}}_{\boldsymbol{\theta}_{\text{PHY}}} = \text{span}_{\mathbb{R}}\left(\left\{ \mathbf{e}(\hat{\phi}_l), -\mathrm{j}\mathbf{e}(\hat{\phi}_l), \frac{\partial \mathbf{e}(\hat{\phi}_l)}{\partial \hat{\phi}_l} \right\}_{l=1}^{L} \right)$.

2: Send pilot sequences $\mathbf{X}_{\text{PHY}}$ of duration $\lceil \frac{3L}{2} \rceil$ built using algorithm 1 and receive user feedback to build observations following (1).

3: Estimate path gains $\hat{\beta}_1, \ldots, \hat{\beta}_L$ and update $\hat{\phi}_1, \ldots, \hat{\phi}_L$.

4: Estimate channel $\mathbf{h}_{\text{PHY}}(\hat{\boldsymbol{\theta}}_{\text{PHY}}) = \sum_{l=1}^{L} \hat{\beta}_l \mathbf{e}(\hat{\phi}_l)$.

Output: $\text{MSE}(\hat{\boldsymbol{\theta}}_{\text{PHY}}) \geq \frac{9}{4} \times \frac{\sigma^2 L^2}{P_t}$

---

imuth estimation error, in order to assess the robustness of the strategies with respect to the angle estimation error. To do so, we consider $N_t = 64$ antennas at the base station (half-wavelength separated ULA). First, to illustrate the fundamental difference on a simple example, we consider a single path channel $\mathbf{h} = \beta \mathbf{e}(\phi)$ with estimated azimuth $\hat{\phi} = 0$ and true azimuth $\phi = \hat{\phi} + \Delta$, Δ being the azimuth estimation error. Then, a lower bound on the relative MSE (MSE divided by the squared norm of the channel) is computed for both strategies. Regarding the new proposed strategy, the lower bound is simply the relative CRB, which is the bound of theorem 1 divided by the squared norm of the channel:

$$\frac{\sigma^2}{2 \|\mathbf{h}\|_2^2} \text{Tr}\left[\mathfrak{Re}\left\{ \mathbf{U}^H \mathbf{M} \mathbf{M}^H \mathbf{U} \right\}^{-1} \right]. \tag{29}$$

with

$$\mathbf{U} = \left(\mathbf{e}(\phi), \overline{\frac{\partial \mathbf{e}(\phi)}{\partial \phi}} \right), \mathbf{M} = \sqrt{\frac{P_t}{\sqrt{2}+1}} \left(2^{\frac{1}{4}} \mathbf{e}(\hat{\phi}), \overline{\frac{\partial \mathbf{e}(\hat{\phi})}{\partial \hat{\phi}}} \right),$$

where $\overline{\frac{\partial \mathbf{e}(\phi)}{\partial \phi}}$ is simply the normalized version of $\frac{\partial \mathbf{e}(\phi)}{\partial \phi}$ (this observation matrix is the result of applying algorithm 1). In this case, the physical model comprises $N_p = 3$ parameters, leading to sequences of length $\lceil \frac{3L}{2} \rceil = 2$. Regarding the angle-constrained estimation strategy, the lower bound on the relative MSE is computed as the maximum of (29) and the relative bias

$$\frac{\left\| \mathbf{h} - \hat{\mathbf{E}}(\hat{\mathbf{E}}^H \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^H \mathbf{h} \right\|_2^2}{\|\mathbf{h}\|_2^2}, \tag{30}$$

with

$$\mathbf{U} = \mathbf{e}(\phi), \hat{\mathbf{E}} = \mathbf{e}(\hat{\phi}), \mathbf{M} = \sqrt{P_t} \mathbf{e}(\hat{\phi})$$

In this case, the simplified physical model comprises $N_p = 2$ parameters, leading to sequences of length $L = 1$. These error lower bounds are shown for $\Delta \in \{0°, 1.0°, 5.0°\}$ (0° meaning perfect angle estimates, leading to the optimal CRB) on figure 1 as a function of the potential signal to noise ratio (pSNR)

$$\text{pSNR} \triangleq \frac{P_t \|\mathbf{h}\|_2^2}{\sigma^2},$$

which is an upper bound on the classical SNR, attained only if the precoder is perfectly collinear to the channel. From the figure, one can notice that the optimal CRB (with perfect

angle estimate, $\Delta = 0$) is lower for the angle-constrained strategy than for the proposed strategy, as stated in the previous paragraph. However, it is interesting to notice that, as soon as there is some angle estimation error ($\Delta > 0$), the proposed strategy is always theoretically better than the angle constrained strategy at high pSNR. This is because it is not biased toward the previously estimated azimuth angles, as is the angle-constrained strategy, since the bias is independent of the pSNR. Moreover, at low pSNR, the proposed strategy and the angle-constrained one are within a few decibels, because the number of parameter to estimate is pretty close in both cases (proportional to the number of dominant paths). For a relatively small angle estimation error ($\Delta = 1$), the proposed strategy exhibits a robust behavior, being very close to the optimal CRB ($\Delta = 0$) for all pSNR values, which is not the case of the angle-constrained strategy. Finally, note that for a pretty large angle estimation error ($\Delta = 5$) the proposed strategy leads to better results than the angle-constrained one, and that even for a low pSNR. In summary, the azimuth estimation error seems to have much less effect on the performance of the proposed strategy than on that of the angle-constrained strategy. Once again, this is explained by the fact that updating the azimuths estimates allows to reduce the impact of the initial estimation error, yielding a more robust strategy. In the two phases communication workflow we consider here, the robustness is interesting since it allows to perform the channel tracking phase longer without degrading the link. In turn, this allows to avoid as long as possible the costly path estimation phase, in order to spare resources for communication purposes.
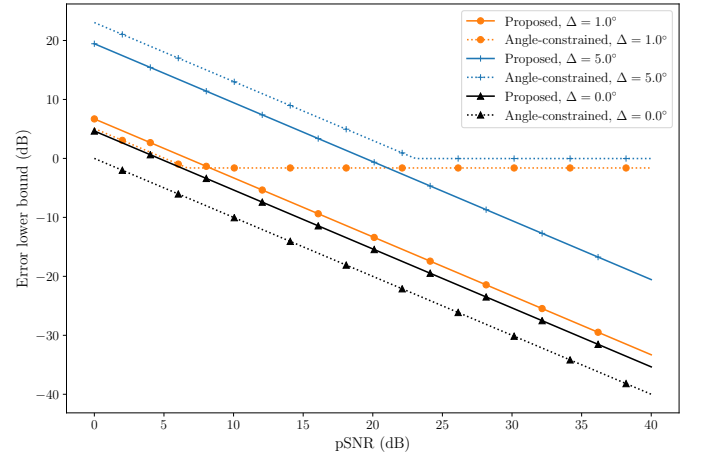

Fig. 1. Comparison of estimation strategies for single path channels ($L = 1$).

In order to further validate the approach, let us now study a more practical scenario, with multipath channels. To do so, we consider a clustered channel model at a frequency of 28 GHz, with L being equal to the number of clusters. The number of clusters and their powers are drawn according to the NYUSIM channel model [40] in the NLOS scenario, which yields $L \in [1, 7]$. Azimuth angles corresponding to the main azimuth of each cluster $\phi_1, \ldots, \phi_L$ are uniformly distributed between 0 and 2π. In order to simulate the azimuth estimation error,

the estimated azimuths are generated as $\hat{\phi}_l = \phi_l + \delta_l$, δ_l being uniformly distributed between $-\Delta$ and Δ. Pilot sequences are built using algorithm 1. Results for $\Delta \in \{0°, 1.0°, 5.0°\}$ are shown on figure 2. The curves exhibit qualitatively the same behavior as for the single path case studied before, definitely showing that the theoretical study performed in this paper allows to design efficient downlink pilot sequences yielding more robustness to the channel tracking phase of methods based on previous angle estimates [11], [12].
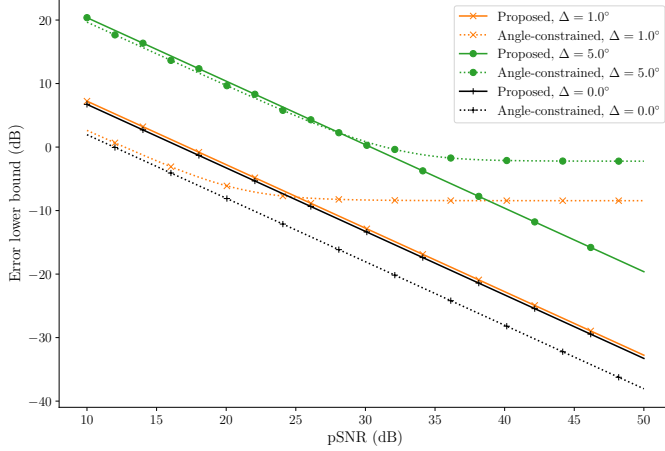

Fig. 2. Comparison of estimation strategies for multipath channels generated according to NYUSIM channel model [40]. Averages over 1000 channel realizations are shown.

VI. Conclusion

In this paper, we studied the problem of estimating a channel 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).

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).

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 optimal observation matrices which is carried out next.

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 a setting for which they had not been determined yet. In particular, it allows to design efficient downlink pilot

sequences yielding more robustness to the channel tracking phase of methods based on previous angle estimates [11], [12].

In the future, the theoretical results provided here could be applied to more practical MIMO systems, for example including hybrid precoding and combining [7], [41], [42]. 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). 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.

Appendix A
Proof of lemma 1

(i) The general strategy of the proof is to exhibit 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. The first step of the proof amounts to link the real and complex inner products as

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{C}} = \mathfrak{Re}\{\mathbf{a}^H \mathbf{b}\} + \mathrm{j}\mathfrak{Im}\{\mathbf{a}^H \mathbf{b}\} = \langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}} + \mathrm{j}\langle \mathrm{j}\mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}}. \tag{31}$$

Now, the idea is to maximize the second term of this sum (which will automatically cancel the first one) in order to build recursively an $\mathbb{R}$-orthonormal basis of $\mathcal{E}$ with the sought properties. To do so, let us choose

$$(\mathbf{v}_1, \mathbf{w}_1) \in \underset{\substack{(\mathbf{v}, \mathbf{w}) \in \mathcal{E}^2 \\ \|\mathbf{v}\|_2 = \|\mathbf{w}\|_2 = 1}}{\operatorname{argmax}} \langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}}, \tag{32}$$

which necessarily exist since the function $\langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}}$ to maximize is continuous and the constraint set is compact. Moreover, let

$$c_1 \triangleq \max_{\substack{(\mathbf{v}, \mathbf{w}) \in \mathcal{E}^2 \\ \|\mathbf{v}\|_2 = \|\mathbf{w}\|_2 = 1}} \langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}} = \langle \mathbf{v}_1, \mathrm{j}\mathbf{w}_1 \rangle_{\mathbb{R}}. \tag{33}$$

Note that by the Cauchy-Schwarz inequality, $c_1 \leq 1$. Moreover, $c_1 \geq 0$ because if $\langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}} \leq 0$, then $\langle -\mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}} \geq \langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}}$. The case $c_1 = 0$ is easily handled since in that case, any $\mathbb{R}$-orthogonal basis is automatically also $\mathbb{C}$-orthogonal and (i) is proven. For the case $0 < c_1 \leq 1$, let us write the Lagrangian of the constrained maximization problem (32):

$$\mathcal{L}(\mathbf{v}, \mathbf{w}, \alpha, \beta) \triangleq \langle \mathbf{v}, \mathrm{j}\mathbf{w} \rangle_{\mathbb{R}} + \alpha(\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbb{R}} - 1) + \beta(\langle \mathbf{w}, \mathbf{w} \rangle_{\mathbb{R}} - 1),$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are the Lagrange multipliers (voluntarily ignoring the constraint $(\mathbf{v}, \mathbf{w}) \in \mathcal{E}^2$ for now). Differentiating it with respect to $\mathbf{v}$ and $\mathbf{w}$ and writing the optimality conditions (introducing the constraint $(\mathbf{v}, \mathbf{w}) \in \mathcal{E}^2$) yields $\forall \mathbf{z} \in \mathcal{E}$,

$$\langle \mathrm{j}\mathbf{w}_1 + 2\alpha\mathbf{v}_1, \mathbf{z} \rangle_{\mathbb{R}} = 0 \tag{34}$$

and

$$\langle \mathrm{j}\mathbf{v}_1 + 2\beta\mathbf{w}_1, \mathbf{z} \rangle_{\mathbb{R}} = 0 \tag{35}$$

From there, injecting $\mathbf{z} = \mathbf{v}_1$ in (34) (resp. $\mathbf{z} = \mathbf{w}_1$ in (35)) yields $-2\alpha = c_1$ (resp. $2\beta = c_1$). Moreover, injecting $\mathbf{z} = \mathbf{w}_1$ in (34) yields $\mathbf{v}_1 \perp_{\mathbb{R}} \mathbf{w}_1$, so that $\operatorname{span}_{\mathbb{R}}(\{\mathbf{v}_1, \mathbf{w}_1\})$ is of dimension two. Moreover, if $\mathbf{z} \in \mathcal{E}$ is $\mathbb{R}$-orthogonal to both $\mathbf{v}_1$ and $\mathbf{w}_1$, then (34) implies that $\mathbf{z} \perp_{\mathbb{C}} \mathbf{w}_1$ and (35) implies that $\mathbf{z} \perp_{\mathbb{C}} \mathbf{v}_1$. This means that $\mathcal{E}$ can be decomposed into the direct sum of a subspace of dimension 2 ($\operatorname{span}_{\mathbb{R}}(\{\mathbf{v}_1, \mathbf{w}_1\})$) and a subspace of dimension $d-2$ (containing all the $\mathbf{z} \in \mathcal{E}$ that are $\mathbb{R}$-orthogonal to both $\mathbf{v}_1$ and $\mathbf{w}_1$) that are $\mathbb{C}$-orthogonal. The exact same reasoning can then be re-applied to the subspace of dimension $d-2$ to prove the lemma by descent, introducing the vectors $\mathbf{v}_2$ and $\mathbf{w}_2$ as the solution of (32) on this subspace and the quantity c_2 as the inner product $\langle \mathbf{v}_2, \mathrm{j}\mathbf{w}_2 \rangle$. The descent stops when the dimension of the remaining subspace is strictly smaller than two, so that if d is odd, the last subspace of the decomposition is of dimension one.

(ii) Now, let us prove that the subspace of dimension two identified at each step necessarily belongs to an eigenspace of the operator $\mathbf{P}_{\mathcal{E}} \circ \mathbf{P}_{\mathrm{j}\mathcal{E}}$. First of all, by the Hilbert projection theorem, for any $\mathbf{x} \in \mathcal{F}$ we can define $\mathbf{P}_{\mathcal{E}}\mathbf{x} \triangleq \operatorname{argmin}_{\mathbf{s} \in \mathcal{E}} \|\mathbf{x} - \mathbf{s}\|_2$ and $\mathbf{P}_{\mathrm{j}\mathcal{E}}\mathbf{x} \triangleq \operatorname{argmin}_{\mathbf{s} \in \mathrm{j}\mathcal{E}} \|\mathbf{x} - \mathbf{s}\|_2$, which are the orthogonal projections of $\mathbf{x}$ onto $\mathcal{E}$ and $\mathrm{j}\mathcal{E}$. One can notice that the two projections are linked since $\mathbf{P}_{\mathcal{E}}(\mathrm{j}\mathbf{x}) = \mathrm{j}\mathbf{P}_{\mathrm{j}\mathcal{E}}\mathbf{x}$. Then, combining the definition of the projection operators with (32) and (33) yields

$$c_1 \mathbf{v}_1 = \mathbf{P}_{\mathcal{E}}(\mathrm{j}\mathbf{w}_1) = \mathrm{j}\mathbf{P}_{\mathrm{j}\mathcal{E}}\mathbf{w}_1$$

and

$$c_1 \mathbf{w}_1 = \mathbf{P}_{\mathcal{E}}(\mathrm{j}\mathbf{v}_1) = \mathrm{j}\mathbf{P}_{\mathrm{j}\mathcal{E}}\mathbf{v}_1.$$

Combining these two equations, we get

$$\mathbf{P}_{\mathcal{E}} \circ \mathbf{P}_{\mathrm{j}\mathcal{E}}(\mathbf{v}_1) = -c_1^2 \mathbf{v}_1$$

and

$$\mathbf{P}_{\mathcal{E}} \circ \mathbf{P}_{\mathrm{j}\mathcal{E}}(\mathbf{w}_1) = -c_1^2 \mathbf{w}_1,$$

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.

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,

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

where the matrix $\mathbf{A} = \mathfrak{Im}\{\mathbf{U}^H \mathbf{U}\}$ is skew-symmetric, so that it admits the following real normal form [43, Theorem 8.16] also known as the Youla decomposition [44]:

$$\mathbf{B}^T \mathbf{A} \mathbf{B} = \begin{pmatrix} 0 & -c_1 & & & \\ c_1 & 0 & & & \\ & & 0 & -c_2 & \\ & & c_2 & 0 & \\ & & & & \ddots \end{pmatrix} \triangleq \mathbf{\Gamma}, \qquad (36)$$

with $\mathbf{B} \in \mathbb{R}^{d \times d}$ a real orthogonal matrix ($\mathbf{B}^T \mathbf{B} = \mathbf{Id}$) whose columns are eigenvectors of the symmetric positive semidefinite 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 $0 \le c_k \le 1, \forall k$. This yields

$$\mathbf{B}^T \mathbf{U}^H \mathbf{U} \mathbf{B} = \mathbf{Id} + \mathrm{j}\mathbf{\Gamma} = \begin{pmatrix} 1 & -\mathrm{j}c_1 & & & \\ \mathrm{j}c_1 & 1 & & & \\ & & 1 & -\mathrm{j}c_2 & \\ & & \mathrm{j}c_2 & 1 & \\ & & & & \ddots \end{pmatrix},$$

so that the columns of the matrix $\mathbf{U}\mathbf{B}$ 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. In practice, the matrix $\mathbf{B}$ and the values $c_1, c_2, \dots$ can be obtained by computing the real Schur decomposition of the matrix $\mathfrak{Im}\{\mathbf{U}^H \mathbf{U}\}$ and reordering the blocks.

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 [45] between $\mathcal{E}$ and $\mathrm{j}\mathcal{E}$.

APPENDIX B
PROOF OF THEOREM 3

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 $\mathcal{V}_{\boldsymbol{\theta}}$, we express the CRB as

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

where $\mathbf{V}$ is the matrix defined in (8) when applying lemma 1 to $\mathcal{V}_{\boldsymbol{\theta}}$.

Next, using the fact that for a symmetric positive semidefinite matrix $\mathbf{A}$, $(\mathbf{A}^{-1})_{ii} \ge \frac{1}{a_{ii}}, \forall i$ [46, Theorem 7.7.15], we get

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

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

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

and

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

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

$$\|\mathbf{M}^H \mathbf{v}_k\|_2^2 = \frac{1}{2} \Big[(1 + c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^+\|_2^2 + (1 - c_k) \|\mathbf{M}^H \tilde{\mathbf{u}}_k^-\|_2^2 + \sqrt{1 - c_k^2} \mathfrak{Re}\{(\tilde{\mathbf{u}}_k^+)^H \mathbf{M} \mathbf{M}^H \tilde{\mathbf{u}}_k^-\} \Big],$$

and

$$\left\|\mathbf{M}^H\mathbf{w}_k\right\|_2^2 = \frac{1}{2}\Big[(1+c_k)\left\|\mathbf{M}^H\tilde{\mathbf{u}}_k^+\right\|_2^2 + (1-c_k)\left\|\mathbf{M}^H\tilde{\mathbf{u}}_k^-\right\|_2^2$$
$$- \sqrt{1-c_k^2}\,\mathfrak{Re}\{(\tilde{\mathbf{u}}_k^+)^H\mathbf{M}\mathbf{M}^H\tilde{\mathbf{u}}_k^-\}\Big].$$

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

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

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 $d_k = \sqrt{1-c_k^2}\,\mathfrak{Re}\{(\tilde{\mathbf{u}}_k^+)^H\mathbf{M}\mathbf{M}^H\tilde{\mathbf{u}}_k^-\} = 0$, $\forall k$.

In summary, we have

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

if and only if $\mathfrak{Re}\{\mathbf{V}^H\mathbf{M}\mathbf{M}^H\mathbf{V}\}$ is diagonal and $\mathfrak{Re}\{(\tilde{\mathbf{u}}_k^+)^H\mathbf{M}\mathbf{M}^H\tilde{\mathbf{u}}_k^-\} = 0$, $\forall k$. Moreover, $\left\|\mathbf{M}\right\|_F^2 = \mathrm{Tr}[\mathbf{M}\mathbf{M}^H] \geq \sum_{k=1}^{\frac{N_p}{2}}P_k^+ + P_k^-$, with an equality if and only if $\mathrm{im}_{\mathbb{C}}(\mathbf{M}) \subset \mathrm{span}_{\mathbb{C}}(\{\tilde{\mathbf{u}}_k^+,\tilde{\mathbf{u}}_k^-\}_{k=1}^{\frac{N_p}{2}})$. The optimization problem (6) is thus lower-bounded by the simpler problem

$$\underset{P_k^+,P_k^-,k=1,\ldots,\frac{N_p}{2}}{\text{minimize}}\quad \sum_{k=1}^{\frac{N_p}{2}}\frac{4}{(1-c_k)P_k^- + (1+c_k)P_k^+},$$
$$\text{subject to}\quad \sum_{k=1}^{\frac{N_p}{2}}P_k^+ + P_k^- = P. \qquad (37)$$

Let us solve this problem and then identify matrices $\mathbf{M}$ for which the optimal values of (37) and (6) coincide. It is obvious that at the optimum of (37), $P_k^- = 0$, $\forall k$, so that it is equivalent to solve the even simpler problem

$$\underset{P_k^+,k=1,\ldots,\frac{N_p}{2}}{\text{minimize}}\quad \sum_{k=1}^{\frac{N_p}{2}}\frac{4}{(1+c_k)P_k^+},$$
$$\text{subject to}\quad \sum_{k=1}^{\frac{N_p}{2}}P_k^+ = P. \qquad (38)$$

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

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

and the optimal value of the optimization problems (37) and (38) is

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

It is also the optimal value of problem (6), since it is attained with the observation matrix

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

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

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 (9) rather than the one given in (7).

REFERENCES

[1] Simon Haykin. *Communication systems*. John Wiley & Sons, 2008.
[2] Fredrik Rusek, Daniel Persson, Buon Kiong Lau, Erik G Larsson, Thomas L Marzetta, Ove Edfors, and Fredrik Tufvesson. Scaling up mimo: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, 30(1):40–60, 2013.
[3] Erik G Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L Marzetta. Massive mimo for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, 2014.
[4] Lu Lu, Geoffrey Ye Li, A Lee Swindlehurst, Alexei Ashikhmin, and Rui Zhang. An overview of massive mimo: Benefits and challenges. *IEEE journal of selected topics in signal processing*, 8(5):742–758, 2014.
[5] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. Millimeter wave mobile communications for 5g cellular: It will work! *IEEE access*, 1:335–349, 2013.
[6] A Lee Swindlehurst, Ender Ayanoglu, Payam Heydari, and Filippo Capolino. Millimeter-wave massive mimo: the next wireless revolution? *IEEE Communications Magazine*, 52(9):56–62, 2014.
[7] Robert W Heath, Nuria Gonzalez-Prelcic, Sundeep Rangan, Wonil Roh, and Akbar M Sayeed. An overview of signal processing techniques for millimeter wave mimo systems. *IEEE journal of selected topics in signal processing*, 10(3):436–453, 2016.
[8] Mehrzad Biguesh and Alex B Gershman. Training-based mimo channel estimation: a study of estimator tradeoffs and optimal training signals. *IEEE transactions on signal processing*, 54(3):884–893, 2006.
[9] A. Adhikary, J. Nam, J. Ahn, and G. Caire. Joint spatial division and multiplexing—the large-scale array regime. *IEEE Transactions on Information Theory*, 59(10):6441–6463, 2013.
[10] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch. Joint spatial division and multiplexing for mm-wave channels. *IEEE Journal on Selected Areas in Communications*, 32(6):1239–1255, 2014.
[11] Z. Gao, L. Dai, Z. Wang, and S. Chen. Spatially common sparsity based adaptive channel estimation and feedback for fdd massive MIMO. *IEEE Transactions on Signal Processing*, 63(23):6169–6183, 2015.
[12] Hongxiang Xie, Feifei Gao, Shun Zhang, and Shi Jin. A unified transmission strategy for tdd/fdd massive mimo systems with spatial basis expansion model. *IEEE Transactions on Vehicular Technology*, 66(4):3170–3184, 2017.
[13] Calyampudi Radakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–89, 1945.
[14] Harald Cramér. *Mathematical Methods of Statistics*, volume 9. Princeton university press, 1946.
[15] Thomas 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*, volume 37, pages 958–966. Citeseer, 1999.
[16] Xiaoli Ma, Georgios B Giannakis, and Shuichi Ohno. Optimal training for block transmissions over doubly selective wireless fading channels. *IEEE Transactions on Signal Processing*, 51(5):1351–1366, 2003.
[17] Imad Barhumi, Geert Leus, and Marc Moonen. Optimal training design for mimo ofdm systems in mobile wireless channels. *IEEE Transactions on signal processing*, 51(6):1615–1624, 2003.
[18] Hlaing Minn and N. Al-Dhahir. Optimal training signals for MIMO OFDM channel estimation. *IEEE Transactions on Wireless Communications*, 5(5):1158–1168, 2006.
[19] Jayesh H Kotecha and Akbar M Sayeed. Transmit signal design for optimal estimation of correlated mimo channels. *IEEE Transactions on Signal Processing*, 52(2):546–557, 2004.

[20] Emil Bjornson and Björn Ottersten. A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance. *IEEE Transactions on Signal Processing*, 58(3):1807–1820, 2009.

[21] Junil Choi, David J Love, and Patrick 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*, 8(5):802–814, 2014.

[22] Y. Gu and Y. D. Zhang. Information-theoretic pilot design for downlink channel estimation in fdd massive MIMO systems. *IEEE Transactions on Signal Processing*, 67(9):2334–2346, 2019.

[23] Samer Bazzi and Wen Xu. Downlink training sequence design for fdd multiuser massive mimo systems. *IEEE Transactions on Signal Processing*, 65(18):4732–4744, 2017.

[24] Samer Bazzi and Wen Xu. On the amount of downlink training in correlated massive mimo channels. *IEEE Transactions on Signal Processing*, 66(9):2286–2299, 2018.

[25] Samer Bazzi, Stelios Stefanatos, Luc Le Magoarou, Salah Eddine Hajri, Mohamad Assaad, Stéphane Paquelet, Gerhard Wunder, and Wen Xu. Exploiting the massive mimo channel structural properties for minimization of channel estimation error and training overhead. *IEEE Access*, 7:32434–32452, 2019.

[26] Luc Le Magoarou and Stéphane Paquelet. Parametric channel estimation for massive MIMO. In *IEEE Statistical Signal Processing Workshop (SSP)*, 2018.

[27] Luc Le Magoarou and Stéphane Paquelet. Performance of MIMO channel estimation with a physical model. *arXiv e-prints*, page arXiv:1902.07031, Feb 2019.

[28] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[29] Adriaan 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.

[30] David Slepian. Estimation of signal parameters in the presence of noise. *Transactions of the IRE Professional Group on Information Theory*, 3(3):68–89, 1954.

[31] G. W. Bangs. *Array Processing With Generalized Beamformers*. PhD thesis, Yale university, CT, USA, 1971.

[32] Olivier Besson and Yuri I Abramovich. On the fisher information matrix for multivariate elliptically contoured distributions. *IEEE Signal Processing Letters*, 20(11):1130–1133, 2013.

[33] Donatella Darsena, Giacinto Gelli, Luigi Paura, and Francesco Verde. Subspace-based blind channel identification of siso-fir systems with improper random inputs. *Signal Processing*, 84(11):2021–2039, 2004.

[34] Donatella Darsena, Giacinto Gelli, Luigi Paura, and Francesco Verde. Widely linear equalization and blind channel identification for interference-contaminated multicarrier systems. *IEEE Transactions on Signal Processing*, 53(3):1163–1177, 2005.

[35] Saeed Abdallah and Ioannis N Psaromiligkos. Widely linear versus conventional subspace-based estimation of simo flat-fading channels: Mean squared error analysis. *IEEE Transactions on Signal Processing*, 60(3):1307–1318, 2011.

[36] J-P Delmas and Habti Abeida. Stochastic crame/spl acute/r-rao bound for noncircular signals with application to doa estimation. *IEEE Transactions on Signal Processing*, 52(11):3192–3199, 2004.

[37] Paul Richard Halmos. *A Hilbert space problem book*, volume 19. Springer Science & Business Media, 1982.

[38] Akbar M Sayeed. Deconstructing multiantenna fading channels. *IEEE Transactions on Signal Processing*, 50(10):2563–2579, 2002.

[39] 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*, 98(6):1058–1076, June 2010.

[40] Mathew K Samimi and Theodore S Rappaport. 3-d millimeter-wave statistical channel model for 5g wireless system design. *IEEE Transactions on Microwave Theory and Techniques*, 64(7):2207–2225, 2016.

[41] Omar El Ayach, Sridhar Rajagopal, Shadi Abu-Surra, Zhouyue Pi, and Robert W Heath. Spatially sparse precoding in millimeter wave mimo systems. *IEEE Transactions on Wireless Communications*, 13(3):1499–1513, 2014.

[42] Akbar M. Sayeed and John H. Brady. *Millimeter-Wave MIMO Transceivers: Theory, Design and Implementation*, pages 231–253. John Wiley & Sons, Ltd, 2016.

[43] Werner H Greub. *Linear algebra*, volume 23. Springer Science & Business Media, 1975.

[44] D. C. Youla. A normal form for a matrix under the unitary congruence group. *Canadian Journal of Mathematics*, 13:694–704, 1961.

[45] Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[46] Roger A Horn and Charles R Johnson. *Matrix Analysis: Second Edition*. Cambridge university press, 2012.

Luc Le Magoarou is a researcher at b<>com (Rennes, France). He received the Ph.D. in signal processing and the M.Sc. in electrical engineering, both from the National Institute of Applied Sciences (INSA Rennes, France), in 2016 and 2013 respectively. During his Ph.D., he was with Inria (Rennes, France), in the PANAMA research group. His main research interests are signal processing and machine learning, currently applied to MIMO communication systems.

Stéphane Paquelet received the B.Sc. degree from the Ecole Polytechnique, Paris, France, in 1996, and the M.Sc. degree from Telecom Paris, Paris, France, in 1998. He successively worked in the fields of cryptology and signal processing for electronic warfare with Thales and led UWB R&D with Mitsubishi Electric from 2002 to 2007, where he proposed two pioneering transceivers for short-range/high data rates and large-range/low data rates, including telemetry. He developed multi-standard reconfigurable RF-IC at Renesas Design France until 2014. From 2014 to 2019, he led wireless communications activities for IRT b<>com, where he now leads the artificial intelligence laboratory.