SNOPS: Short Non-Orthogonal Pilot Sequences for Downlink Channel State Estimation in FDD Massive MIMO

Beatrice Tomasi, Alexis Decurninge, Maxime Guillaud
Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies Co. Ltd.
France Research Center, 20 quai du Point du Jour, 92100 Boulogne Billancourt, France
email: {beatrice.tomasi,alexis.decurninge,maxime.guillaud}@huawei.com

Abstract—Channel state information (CSI) acquisition is a significant bottleneck in the design of Massive MIMO wireless systems, due to the length of the training sequences required to distinguish the antennas (in the downlink) and the users (for the uplink where a given spectral resource can be shared by a large number of users). In this article, we focus on the downlink CSI estimation case. Considering the presence of spatial correlation at the base transceiver station (BTS) side, and assuming that the per-user channel statistics are known, we seek to exploit this correlation to minimize the length of the pilot sequences. We introduce a scheme relying on non-orthogonal pilot sequences and feedback from the user terminal (UT), which enables the BTS to estimate all downlink channels. Thanks to the relaxed orthogonality assumption on the pilots, the length of the obtained pilot sequences can be strictly lower than the number of antennas at the BTS, while the CSI estimation error is kept arbitrarily small. We introduce two algorithms to dynamically design the required pilot sequences, analyze and validate the performance of the proposed CSI estimation method through numerical simulations using a realistic scenario based on the one-ring channel model.

I. INTRODUCTION

Channel state information (CSI) acquisition represents an important problem in the multi-user Massive MIMO (Multiple-Input Multiple-Output) scenario [1]. Accurate downlink CSI is required in order to obtain the large multiplexing gain expected from massive MIMO systems and achieve the rates shown e.g. in [2]. It is well known that, in the presence of i.i.d. channels, it is necessary to make the length of the pilot sequences at least as large as the total number of transmit antennas, in order to avoid the effect known as pilot contamination [3]. Depending on the coherence time of the channel, the transmission of long training sequences instead of data-bearing symbols can represent a significant loss in spectral efficiency. This issue is exacerbated in frequency-division duplex (FDD) systems, where downlink CSI can not be simply obtained by reciprocity [4] at the BTS; in that scenario, the large number of BTS antennas makes the use of orthogonal pilot sequences especially unwelcome.

In the context of Massive MIMO however, the channels exhibit a large degree of correlation [5]. In fact, a denser antenna array improves the spatial resolution, and makes the received signal more spatially correlated, to the point of resulting in a rank deficient spatial correlation matrix [6]. This correlation can potentially help reduce the required training overhead. The problem of pilot design for MIMO correlated channels has been studied previously, for example in [7], [8], [9] and [10]. However, in these works the pilot optimization is done for a point-to-point MIMO system. The single-user assumption was lifted in [11], where it is proposed to schedule uplink CSI acquisition across the users such that the terminals can be separated in space; pilot sequences orthogonality between users is therefore not required, yielding shorter training sequences. However, it is not clear how close to perfect separation in space a practical system can operate, considering realistic propagation conditions, and a finite number of users to choose from at the scheduling stage. An alternative approach was introduced in [12], where channel sparsity in the time, frequency and angular domains is assumed, and short pilots are obtained through compressed sensing techniques.

Another aspect of the problem, related to mitigating pilot contamination in the context of multi-cell Massive MIMO, has been considered in [6]; in this work, the same (orthogonal) pilot sequences are reused across the cells, and the problem of assigning each user to one of the pilot sequences is considered.

In [13], the design of non-orthogonal pilots based on per-user spatial covariance information has been addressed for *uplink* multi-user Massive MIMO CSI estimation; the object of the present article is to introduce a downlink counterpart to this approach. In the sequel, we seek to optimize the design of the training sequences transmitted by the BTS during downlink CSI estimation. The salient features of the proposed approach are:

- The BTS designs a set of non-orthogonal pilots based on statistical CSI, and transmits them.
- Each UT feeds back (typically after quantization) the corresponding received sequence to the BTS.
- The BTS reconstructs the CSI based on UT feedback and prior covariance information.

As will be seen, this enables to dynamically control the accuracy of CSI estimation, as well as to use short, non-orthogonal pilot sequences (SNOPS), suitably adapted to the
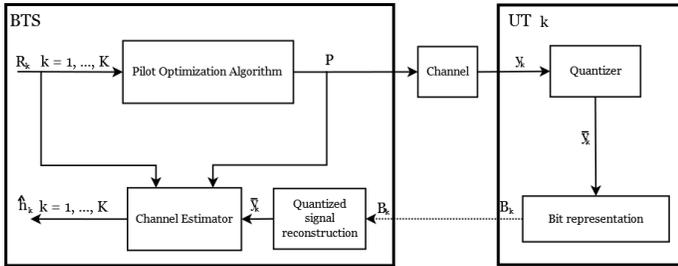
Fig. 1. Functional representation of the proposed CSI acquisition approach. Only one among the K UTs is shown.

channel statistics.

This article is organized as follows: we first introduce the system model and the pilot-based downlink CSI estimation model in Section II. In Section III, we introduce two algorithms to design pilot sequences with the objective of minimizing their length or their total energy, while simultaneously ensuring that the channel estimation error variance is upper-bounded by a per-user constant. Finally, in Section IV, we show through simulations using a realistic scenario based on the one-ring channel model that the proposed technique can yield pilot sequences of length significantly smaller than the number of antennas at the BTS.

II. SYSTEM MODEL

A. Channel model

We consider a communication system with M antenna elements at the BTS and K single-antenna UTs, with $M > K$. In order to incorporate spatial correlation in the model, it is assumed that the column vector of the flat-fading downlink channel coefficients between the M antennas at the BTS and the k-th single-antenna UT, $\mathbf{h}_k \in \mathbb{C}^M$ can be expressed as the product between a spatial correlation matrix $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{M \times r_k}$ and a vector $\boldsymbol{\eta}_k \in \mathbb{C}^{r_k}$ of complex zero-mean Gaussian i.i.d. random variables with covariance $\mathbf{I}_{r_k}$ that represents the fast fading process, i.e.,

$$\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}} \boldsymbol{\eta}_k. \tag{1}$$

In this model, r_k is the rank of the spatial covariance matrix, $\mathbf{R}_k$, and is lower or equal to the number of antenna elements in the BTS, M. We will decompose each per-user covariance matrix as $\mathbf{R}_k^{\frac{1}{2}} = \mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}}$, where $\boldsymbol{\Lambda}_k$ and $\mathbf{U}_k$ are respectively the diagonal matrix containing the r_k non-zero eigenvalues of $\mathbf{R}_k$, also incorporating a path loss coefficient, and the matrix of the associated eigenvectors. In this study, $\mathbf{R}_k$ is assumed to be constant (stationary channel process) and known at the BTS, i.e., statistical CSI is available[1].

[1] The issue of estimating the per-user BTS-side spatial covariance is outside of the scope of this paper. One approach is to apply classical covariance estimation and tracking methods, based on the estimated instantaneous CSI. Another method, based on a learned dictionary of uplink/downlink covariance matrix pairs, suitable for FDD systems, was introduced in [14].

B. Pilot-Based CSI Estimation Approach

Since we are only concerned with downlink CSI estimation, we focus on the transmission of pilot sequences by the BTS. Let us assume that pilot sequences of overall length T are simultaneously broadcasted by all M antennas; we let matrix $\mathbf{P} \in \mathbb{C}^{T \times M}$ denote the complete set of pilot sequences used in the system. The signal received at UT k over the corresponding T time instants, denoted by vector $\mathbf{y}_k \in \mathbb{C}^T$, can be expressed as:

$$\mathbf{y}_k = \mathbf{P}\mathbf{h}_k + \mathbf{w}_k \tag{2}$$

where $\mathbf{w}_k$ represents the per-user additive white Gaussian noise at UT k, with covariance $\sigma_{n,k}^2 \mathbf{I}_T$. Each UT will then quantize the received signal and transmit it, e.g., in the form of a binary representation B_k, back to the BTS. We assume that the feedback link is perfect, and therefore the BTS can perfectly recover the quantized version of $\mathbf{y}_k$, for all $k = 1, \ldots, K$, denoted as

$$\bar{\mathbf{y}}_k = \mathbf{y}_k + \mathbf{z}_k = \mathbf{P}\mathbf{h}_k + \mathbf{w}_k + \mathbf{z}_k \tag{3}$$

where $\mathbf{z}_k$ represents the per-user quantization error, which is assumed Gaussian with covariance $\sigma_{Q,k}^2 \mathbf{I}_T$ and independent from the other variables in the problem. The BTS can then compute the channel estimate $\hat{\mathbf{h}}_k$ based on the combined knowledge of $\bar{\mathbf{y}}_k$, $\mathbf{R}_k$ and $\mathbf{P}$, using a simple linear MMSE approach (more details are provided in Section II-C below).

A pictorial representation of the above steps is shown in Fig. 1. Some remarks about this approach are in order:

- The situation of interest in this paper is $T < M$; in this case, $\mathbf{P}$ does not admit a left inverse, and it is not possible to recover exactly $\mathbf{h}_k$ from $\mathbf{y}_k$ only, even in the noise-free case. Therefore, according to the proposed model, the UT does not directly try to estimate the channel.
- On the other hand, the BTS is in a position to estimate $\mathbf{h}_k$ thanks to the statistical CSI assumed available there. In fact, in a Massive MIMO scenario the downlink CSI is used for multi-user precoding, thus making its availability more important at the BTS.
- Due to the above, the channel statistics (in the form of spatial covariance matrices) are required at the BTS only.
- Having short pilot sequences is doubly advantageous, since it improves the spectral efficiency of the downlink pilot transmission, and reduces the dimension of the uplink feedback variable $\bar{\mathbf{y}}_k$.
- Although mathematically similar (but not identical), the uplink [13] and downlink (presented here) CSI estimation methods differ significantly in terms of system design: the former requires the feed-forward coded transmission of the *pilot sequences* $\mathbf{p}_k$ from the BTS to each UT; while the latter involves the coded feed-back of the sequence $\mathbf{y}_k$ as received by UT k to the BTS.

C. LMMSE Estimation Error Analysis

We now introduce the considered Linear Minimum Mean Square Error (LMMSE) channel estimator and derive the

estimation error covariance matrix of each user, as a function of the choice of the common pilot sequences. For a given choice of the pilot matrix $\mathbf{P}$, the LMMSE estimator of the fast fading coefficients between user k and the BTS array, $\hat{\boldsymbol{\eta}}_k$, can be written as

$$\hat{\boldsymbol{\eta}}_k = \mathbf{C}_{\boldsymbol{\eta}_k \bar{\mathbf{y}}_k} \mathbf{C}_{\bar{\mathbf{y}}_k}^{-1} \bar{\mathbf{y}}_k \tag{4}$$

where $\mathbf{C}_{\boldsymbol{\eta}_k \bar{\mathbf{y}}_k} = \mathbf{R}_k^{\frac{H}{2}} \mathbf{P}^H$, and $\mathbf{C}_{\bar{\mathbf{y}}_k} = \mathbf{P}\mathbf{R}_k\mathbf{P}^H + \sigma_k^2 \mathbf{I}_T$, where $\sigma_k^2 = \sigma_{n,k}^2 + \sigma_{Q,k}^2$ represents the variance of the combination of the thermal and quantization noise terms. Considering (1), the CSI estimate is obtained for each user as

$$\hat{\mathbf{h}}_k = \mathbf{R}_k^{\frac{1}{2}} \hat{\boldsymbol{\eta}}_k. \tag{5}$$

The covariance matrix of the estimation error for $\boldsymbol{\eta}_k$ is given by [15]

$$\mathbf{C}_{\mathbf{e},\boldsymbol{\eta}_k}(\mathbf{P}) = \mathbb{E}[(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_k)(\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_k)^H] \tag{6}$$

$$= \mathbf{I}_{r_k} - \mathbf{R}_k^{\frac{H}{2}} \mathbf{P}^H (\mathbf{P}\mathbf{R}_k\mathbf{P}^H + \sigma_k^2\mathbf{I}_T)^{-1} \mathbf{P}\mathbf{R}_k^{\frac{1}{2}} \tag{7}$$

$$= \left(\mathbf{I}_{r_k} + \frac{\mathbf{R}_k^{\frac{H}{2}} \mathbf{P}^H \mathbf{P}\mathbf{R}_k^{\frac{1}{2}}}{\sigma_k^2} \right)^{-1} \tag{8}$$

where from eq. (7) to eq. (8) we used the Woodbury identity [16]. The covariance matrix of the estimation error on $\boldsymbol{h}_k$ is therefore

$$\mathbf{C}_{\mathbf{e},k}(\mathbf{P}) = \mathbb{E}[(\hat{\mathbf{h}}_k - \mathbf{h}_k)(\hat{\mathbf{h}}_k - \mathbf{h}_k)^H] \tag{9}$$

$$= \mathbf{U}_k \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{C}_{\mathbf{e},\boldsymbol{\eta}_k}(\mathbf{P}) \boldsymbol{\Lambda}_k^{\frac{1}{2}} \mathbf{U}_k^H \tag{10}$$

$$= \mathbf{U}_k \left(\boldsymbol{\Lambda}_k^{-1} + \frac{\mathbf{U}_k^H \mathbf{P}^H \mathbf{P}\mathbf{U}_k}{\sigma_k^2} \right)^{-1} \mathbf{U}_k^H. \tag{11}$$

III. MINIMUM LENGTH PILOT SEQUENCES UNDER ESTIMATION ERROR CONSTRAINTS

In order to control the accuracy of the channel estimation process for user k, we assume that we wish to uniformly bound the estimation error on all dimensions of $\boldsymbol{h}_k$ by a given constant $\epsilon_k > 0$. This can be done by requiring that all the eigenvalues of $\mathbf{C}_{\mathbf{e},k}$ are lower than or equal to a per-user constant ϵ_k, which we denote[2] as $\mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}_M$. In this section, we address the design of short pilot sequences under such estimation error constraints.

A. Minimum pilot length without energy constraint

We consider the problem of minimizing the length T of the pilots subject to the per-user maximum estimation error constraints outlined above, first assuming unbounded transmission energy:

$$\min_{\mathbf{P}\in\mathbb{C}^{T\times M}} \quad T \tag{12}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k}(\mathbf{P}) \preceq \epsilon_k\mathbf{I}_M \quad \forall k = 1,\dots,K.$$

We prove the following result in Appendix A:

---

[2] For two positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \preceq \mathbf{B}$ is a shorthand notation for the condition that $\mathbf{B} - \mathbf{A}$ is positive semidefinite.

Proposition 1. *If for each user k the r_k nonzero eigenvalues of $\mathbf{R}_k$ are greater than ϵ_k, i.e., $\boldsymbol{\Lambda}_k \succeq \epsilon_k\boldsymbol{I}$, then the minimum T from (12) is* $\max_{k=1,\dots,K} r_k$.

In other terms, in the regime of low ϵ_k (high accuracy), the minimum pilot length needed for downlink CSI acquisition is limited by the user whose covariance has the highest rank.

B. Energy-Constrained Pilot Length Minimization

We now revisit the optimization problem (12) by introducing a maximum pilot energy constraint, i.e. we seek the minimum T for which there exists a $T \times M$ matrix $\mathbf{P}$ that satisfies $\mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k\mathbf{I}$ and that satisfies a per-antenna maximum pilot energy constraint $E_{\max}$:

$$\min_{\mathbf{P}\in\mathbb{C}^{T\times M}} \quad T \tag{13}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k\mathbf{I}_M \quad \forall k = 1,\dots,K$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1,\dots,M$$

where $\mathbf{X} = \mathbf{P}^H\mathbf{P} \in \mathbb{C}^{M\times M}$ and X_{mm} denotes the m-th diagonal element of $\mathbf{X}$. Note that minimizing T is equivalent to minimizing $\text{rank}(\mathbf{X})$, therefore (13) becomes:

$$\min_{\mathbf{X}\succeq 0} \quad \text{rank}(\mathbf{X}) \tag{14}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k\mathbf{I}_M \quad \forall k = 1,\dots,K$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1,\dots,M.$$

As a consequence of Proposition 1, we get the following Corollary.

Corollary 1. *Let $\lambda_{1,k},\dots,\lambda_{r_k,k}$ denote the r_k nonzero eigenvalues of $\mathbf{R}_k$. The optimum T from (13) is lower bounded by*

$$\max_{k=1,\dots,K} \sum_{i=1}^{r_k} \mathbb{1}\{\lambda_{i,k} \geq \epsilon_k\}.$$

Proof: The constraints of (13) are more restrictive than the constraints of (12), therefore the minimum T from (13) is greater than the optimum T from (12). Using a similar argument, the minimum T from (12) is greater than the optimum T from (12) where we projected each error covariance matrix on the eigenspace composed by eigenvectors whose eigenvalues are greater than ϵ_k. The lower bound is then a consequence of Proposition 1. ■

The problem (13) can efficiently be solved approximately through heuristics, e.g., following the method outlined in [17] as follows. Considering a regularized smooth surrogate of the rank function, namely $\log\det(\mathbf{X} + \delta\mathbf{I})$ for some small δ, (14) becomes:

$$\min_{\mathbf{X}\succeq 0} \quad \log\det(\mathbf{X} + \delta\mathbf{I}) \tag{15}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k\mathbf{I}_M \quad \forall k = 1,\dots,K$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1,\dots,M.$$

The objective function in (15) is concave, however it is smooth on the positive definite cone; a possible way to

approximately solve this problem is to iteratively minimize a locally linearized version of the objective function, i.e. solve

$$\mathbf{X}_{t+1} = \arg \min_{\mathbf{X} \succeq \mathbf{0}} \quad \mathrm{Tr}\left((\mathbf{X}_t + \delta \mathbf{I})^{-1} \mathbf{X}\right) \tag{16}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}_M \quad \forall k = 1, \dots, K$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1, \dots, M$$

until convergence to some $\mathbf{X}^*$. We suggest to initialize the algorithm by choosing $\mathbf{X}_0$ as the rank-1, all-ones matrix $\mathbf{1}_{M \times M}$.

Let us now focus on the constraints on $\mathbf{C}_{\mathbf{e},k}$, $k = 1, \dots, K$. Using (11) and the fact that $\mathbf{C}_{\mathbf{e},k}$ is positive semidefinite, we obtain

$$\mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}_M \Leftrightarrow \mathbf{U}_k^H \mathbf{X} \mathbf{U}_k \succeq \left(\epsilon_k^{-1} \mathbf{I}_{r_k} - \mathbf{\Lambda}_k^{-1}\right) \sigma_k^2 \tag{17}$$
$$\forall k = 1, \dots, K.$$

Note that since these constraints are convex, (16) is a convex optimization problem that can be solved numerically.

Due to the various approximations involved in transforming (14) into (16), $\mathbf{X}^*$ might not be strictly rank-deficient, but it can have some very small eigenvalues instead. It is therefore necessary to apply some thresholding on these eigenvalues to recover a strictly rank-deficient solution. Let us denote by $\boldsymbol{e}_m$ the eigenvector associated to the m-th eigenvalue v_m of $\mathbf{X}^*$, $m = 1 \dots M$, with $v_1 \geq \dots \geq v_M \geq 0$. We then let

$$T = \max_{i=1\dots M} i \tag{18}$$
$$\text{s.t.} \quad v_i \geq \epsilon_s$$

for a suitably chosen (small) ϵ_s, and obtain the matrix of optimized training sequences as $\mathbf{P} = \mathbf{V}^{\frac{1}{2}}[\boldsymbol{e}_1, \dots, \boldsymbol{e}_T]^T$, where $\mathbf{V}$ is the diagonal matrix having $v_1 \dots v_T$ as diagonal coefficients.

The proposed algorithm is shown in Algorithm 1.

Algorithm 1 Minimum length downlink pilot sequence computation with estimation error and per-antenna energy constraints.

Initialize $\mathbf{X}_0 \leftarrow \mathbf{1}_{M \times M}$.
repeat
$$\mathbf{X}_{t+1} \leftarrow \arg \min_{\mathbf{X} \succeq \mathbf{0}} \mathrm{Tr}\left((\mathbf{X}_t + \delta \mathbf{I})^{-1} \mathbf{X}\right)$$
$$\text{s.t.} \quad \mathbf{U}_k^H \mathbf{X} \mathbf{U}_k \succeq \left(\epsilon_k^{-1} \mathbf{I}_{r_k} - \mathbf{\Lambda}_k^{-1}\right) \sigma_k^2 \quad \forall k$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1, \dots, M$$
until convergence to $\mathbf{X}^*$.
Compute T according to (18).
Output: $\mathbf{P} = \mathbf{V}^{\frac{1}{2}}[\boldsymbol{e}_1, \dots, \boldsymbol{e}_T]^T$.

C. Pilot Energy Minimization

We now propose an algorithm where we search the pilots minimizing the total energy dedicated by the BTS to pilot sequences, i.e. $\mathrm{Tr}(\mathbf{X})$, under the same CSI accuracy constraints as before. Note that the trace criterion is also a proxy criterion for the rank (see [17]), incorporating both the energy minimization and the pilot length minimization objectives.

This leads to solve the convex optimization problem presented in Algorithm 2.

Algorithm 2 Pilot energy minimization with estimation error and per-antenna energy constraints.

Compute

$$\mathbf{X}^* \leftarrow \arg \min_{\mathbf{X} \succeq \mathbf{0}} \quad \mathrm{Tr}(\mathbf{X}) \tag{19}$$
$$\text{s.t.} \quad \mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}_M \quad \forall k = 1, \dots, K$$
$$X_{mm} \leq E_{\max} \quad \forall m = 1, \dots, M$$

Compute T according to (18).
Output: $\mathbf{P} = \mathbf{V}^{\frac{1}{2}}[\boldsymbol{e}_1, \dots, \boldsymbol{e}_T]^T$.

IV. NUMERICAL RESULTS

The algorithms presented in this paper have been evaluated through numerical simulations, according to the scenario outlined below: the BTS is equipped with a uniform circular array (UCA) of diameter 2 m, consisting of $M = 32$ antennas, and serves $K = 8$ UTs randomly distributed around the BTS at a distance between 250 and 750 m. For each UT, it is assumed that 200 scatterers are distributed randomly on a disc of radius 50 m centered on each terminal are causing fast fading (one-ring channel model [18]). The covariance matrices are generated by a ray-tracing procedure, with a central frequency of 3.5 GHz; according to this model, the support of the angle of arrivals associated to a given UT is limited, which yields covariance matrices with few large eigenvalues. We have applied a threshold to these eigenvalues to obtain the ranks r_k that ensure that at least 99% of the energy of the full-rank matrix is captured by the rank-deficient model. Distance-dependent path-loss is also incorporated, following the 3D-UMi LoS (dense urban micro-cell, line-of-sight) scenario from [19]. The noise variance (incorporating thermal noise and quantization noise, as detailed in (3)) is chosen identically across all users as $\sigma_k^2 = -110$ dBm.

Algorithms 1 and 2 have been evaluated numerically. For each realization of the covariance matrices, we applied the proposed algorithm to compute $\mathbf{P}$. Since the effect of the estimation error $\hat{\boldsymbol{h}}_k - \boldsymbol{h}_k$ must be interpreted relatively to the channel variance, we introduce a global accuracy parameter ϵ^*, and set the per-user constraints as a function of the path-loss, according to

$$\epsilon_k = \epsilon^* \mathrm{Tr}\{\mathbf{R}_k\}, \tag{20}$$

where Tr denotes the trace operator; note that $\mathrm{Tr}\{\mathbf{R}_k\} = \mathbb{E}\left[||\boldsymbol{h}_k||^2\right]$. The solutions to (16) and (19) were obtained via the numerical solver CVX [20], and parameter δ was set to 10^{-2}. Moreover, for thresholding we consider $\epsilon_s = 10^{-5} \max(v_1, \dots, v_M)$.

Fig. 2 depicts the total transmit energy contained in the pilot sequences transmitted by the BTS versus the corresponding length T, as obtained by Algorithm 1 in light gray and Algorithm 2 in dark gray. The maximum energy constraint per antenna, $E_{\max}$, is set to 41 dB. The vertical lines represent the average lower bound on the minimum rank of the
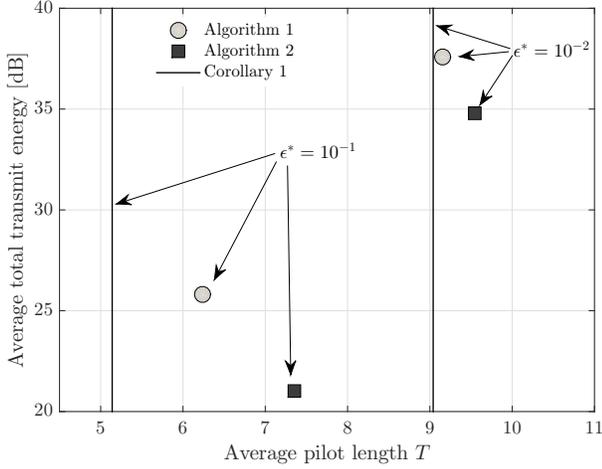
Fig. 2. Average total transmit energy and average pilot length obtained by Algorithms 1 and 2 in a system with $M = 32$ antennas at the BTS.

Fig. 3. Empirical frequency of the maximum value of the ratio between the maximum eigenvalue of the error covariance matrix and the accuracy constraint for Algorithms 1 and 2, when $\epsilon^* = 10^{-1}$.

pilot sequence matrix derived in Corollary 1. All metrics are averaged over 100 realizations of the random user locations. It can be seen that Algorithm 1 performs closer to the optimal (the loss of optimality being attributable to a combination of the per-antenna energy constraint and the influence of the user eigenvalues below the precision threshold ϵ_k that are not taken into account in the lower bound), while Algorithm 2 has better (lower) energy figures, at the cost of slightly longer sequences. It is noticeable that the obtained pilot sequences are significantly shorter (average length lower than 10 for $\epsilon^* = 10^{-2}$) than what would be necessary in the absence of statistical CSI, i.e. $T = M = 32$.

Because of the thresholding operation detailed in (18), for both Algorithm 1 and Algorithm 2, $\mathbf{X} = \mathbf{P}^H \mathbf{P}$ is only approximately equal to $\mathbf{X}^*$. Therefore, the obtained pilot sequences can slightly violate the maximum error variance constraints $\mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}_M$. In order to verify that this has no detrimental effect on the global performance, let us consider the quantity $\max_k \lambda_{\max}(\mathbf{C}_{\mathbf{e},k})/\epsilon_k$, which should always be below 1 in the absence of the thresholding operation. Figs. 3 and 4 show the empirical distributions of these quantities for parameter $\epsilon^* = 10^{-1}$ and $\epsilon^* = 10^{-2}$, respectively. These results show that the departure from the ideal value of 1 is extremely small, and therefore the effect of the thresholding operation on the final CSI estimation error is negligible.

V. CONCLUSIONS

In this paper we addressed the problem of downlink CSI acquisition in a MIMO FDD single-cell cellular network. We proposed two algorithms for pilot sequence design that exploit the higher spatial correlation of the channel coefficients between the BTS antennas and the UTs. Pilot sequences shorter than the current state of the art solutions can be employed to estimate the channel coefficients with a desired accuracy. We evaluated analytical bounds for the rank of pilot matrices and we evaluated numerically the proposed algorithms.
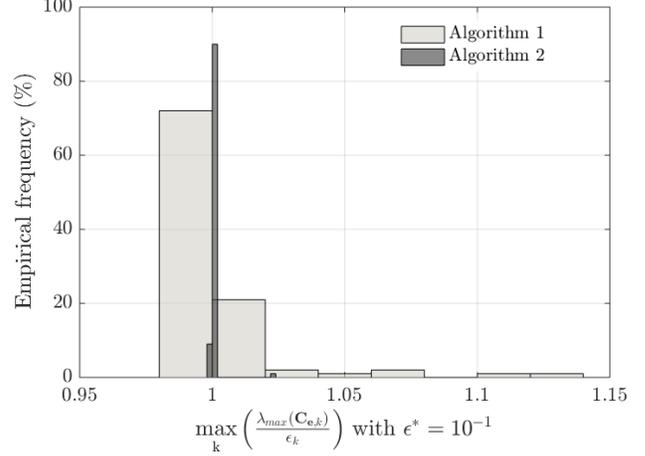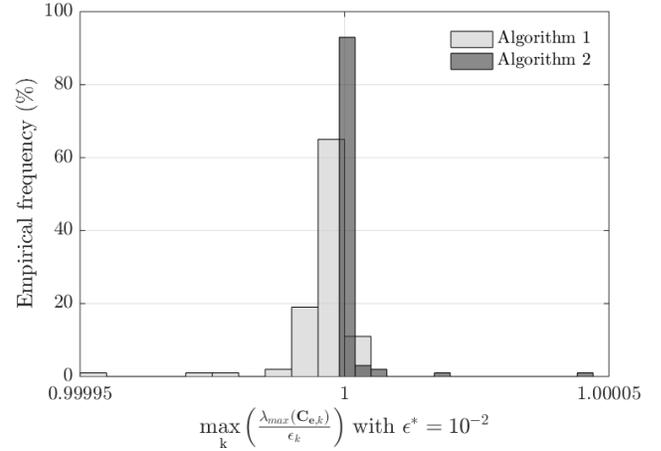

Fig. 4. Empirical frequency of the maximum value of the ratio between the maximum eigenvalue of the error covariance matrix and the accuracy constraint for Algorithms 1 and 2, when $\epsilon^* = 10^{-2}$.

APPENDIX A
PROOF OF PROPOSITION 1

Considering eq. (11), let us denote $\mathbf{Z}_k = \mathbf{P}\mathbf{U}_k$. Proposition 1 follows as consequence of the following propositions proved in this Appendix.

Proposition 2. *Assume that $\mathbf{\Lambda}_k \succeq \epsilon_k \mathbf{I}$ for all users $k = 1, \ldots, K$, then solving the minimization problem (12) is equivalent to solving*

$$\min_{\mathbf{P} \in \mathbb{C}^{T \times M}} \quad T \tag{21}$$
$$\text{s.t.} \quad \text{rank}(\mathbf{Z}_k) = r_k \quad \forall k = 1, \ldots, K.$$

Proof: Let us first recall that $\mathbf{C}_{\mathbf{e},k} \preceq \epsilon_k \mathbf{I}$ is equivalent to $\mathbf{\Lambda}_k^{-1} - \frac{\mathbf{I}}{\epsilon_k} + \frac{\mathbf{Z}_k^H \mathbf{Z}_k}{\sigma_k^2} \succeq 0$. The matrix $\mathbf{D}_k = \frac{\mathbf{I}}{\epsilon_k} - \mathbf{\Lambda}_k^{-1}$ is diagonal of size $r_k \times r_k$ with positive coefficients.

Let us first suppose that there exists $\mathbf{P}$ such that $\mathbf{C_{e,k}} \preceq \epsilon_k \mathbf{I}$. Then for any $\boldsymbol{x} \in \mathbb{C}^{r_k}$, $\|\mathbf{Z}_k \boldsymbol{x}\|^2 - \sigma_k^2 \|\mathbf{D}_k \boldsymbol{x}\|^2 \geq 0$. Then, for any $\boldsymbol{x} \neq 0$, it holds $\|\mathbf{Z}_k \boldsymbol{x}\| \geq \sigma_k^2 \|\mathbf{D}_k \boldsymbol{x}\|^2 > 0$, i.e. $\mathbf{Z}_k$ is full rank, equal to r_k.

Conversely, let us suppose that there exists $\mathbf{P}$ such that $\mathbf{Z}_k$ is full rank. Then, let us consider $\mathbf{P}_2 = \alpha \mathbf{P}$ with some power level $\alpha > 0$. We indicate as $\mathbf{Z}_{2,k} = \mathbf{P}_2 \mathbf{U}_k$. It holds that for any $\boldsymbol{x} \in \mathbb{C}^{r_k}$, $\|\mathbf{Z}_{2,k} \boldsymbol{x}\|^2 - \sigma_k^2 \|\mathbf{D}_k \boldsymbol{x}\|^2 \geq \alpha^2 \lambda_{min}(\mathbf{Z}_k^H \mathbf{Z}_k) - \sigma_k^2 \lambda_{max}(\mathbf{D}_k^2)$. Since $\mathbf{Z}_k$ is full rank, it holds $\lambda_{min}(\mathbf{Z}_k^H \mathbf{Z}_k) > 0$ for all $1 \leq k \leq K$. Then by choosing α such that $\alpha^2 \geq \max_{1 \leq k \leq K} \frac{\sigma_k^2 \lambda_{max}(\mathbf{D}_k^2)}{\lambda_{min}(\mathbf{Z}_k^H \mathbf{Z}_k)}$, it holds that $\frac{\mathbf{Z}_{2,k}^H \mathbf{Z}_{2,k}}{\sigma_k^2} - \mathbf{D}_k \succeq 0 \quad \forall k$. ∎

Proposition 3. *The minimum T attained in* (21) *is equal to* $\max_{1 \leq k \leq K} r_k$.

Proof: It is straightforward to prove that if $\text{rank}(\mathbf{Z}_k) = r_k$ then $r_k = \text{rank}(\mathbf{Z}_k) = \text{rank}(\mathbf{PU}_k) \leq \min(T, r_k)$. Therefore, $T \geq r_k$, for all $k = 1, \ldots, K$, thus, $T \geq \max(r_1, \ldots, r_K)$.

In order to show the proposition, we will now show the existence of a pilot sequence of length $\max(r_1, \ldots r_K)$ such that $\text{rank}(\mathbf{Z}_k) = r_k$ for all $1 \leq k \leq K$.
If we denote $\text{span}(\mathbf{U}_k)^\perp$ the subspace of $\mathbb{C}^M$ orthogonal to the subspace spanned by $\mathbf{U}_k$, it holds

$$r_k \geq \text{rank}(\mathbf{Z}_k) = \text{rank}(\mathbf{PU}_k) \geq \text{rank}(\mathbf{P}_{r_k}\mathbf{U}_k)$$
$$\geq r_k - \dim(\text{span}(\mathbf{P}_{r_k}^T) \cap \text{span}(\mathbf{U}_k)^\perp)$$

where $\mathbf{P}_\ell$ denotes the concatenation of the ℓ first rows of $\mathbf{P}$. It is then sufficient to prove the existence of $\mathbf{P}$ such that for each k, $\text{rank}(\mathbf{P}_{r_k}\mathbf{U}_k) = r_k$, i.e. $\text{span}(\mathbf{P}_{r_k}^T) \cap \text{span}(\mathbf{U}_k)^\perp = \emptyset$. With such a $\mathbf{P}$, it will hold that $\text{rank}(\mathbf{Z}_k) = r_k$ for all k.
We then recursively build the pilot sequence (i.e. the rows of matrix $\mathbf{P}$). For any step $\ell \geq 1$, let $\text{span}(\mathbf{P}_{\ell-1})^\perp$ denote the subspace of $\mathbb{C}^M$ orthogonal to $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{\ell-1}$. We take the convention $\text{span}(\mathbf{P}_0)^\perp = \mathbb{C}^M$, so that the construction is valid for $\ell = 1$. Let us suppose that we built $\ell - 1$ pilots. Then, we choose $\boldsymbol{p}_\ell \in \text{span}(\mathbf{P}_{\ell-1})^\perp$ such that for all users k satisfying $r_k \geq \ell$, $\text{span}(\mathbf{P}_{r_k}) \cap \text{span}(\mathbf{U}_k)^\perp = \emptyset$. After $\ell = \max(r_1, \ldots, r_K)$ steps, it holds that, for each k, $\text{span}(\mathbf{P}_{r_k}^T) \cap \text{span}(\mathbf{U}_k)^\perp = \emptyset$ by construction.
It now remains to show the existence of such $\boldsymbol{p}_\ell$. For each k satisfying $r_k \geq \ell$, it holds $\dim(\text{span}(\mathbf{U}_k)^\perp) = M - r_k < M - \ell + 1 = \dim(\text{span}(\mathbf{P}_{\ell-1})^\perp)$.
We deduce that for each k such that $r_k \geq \ell$, the subspace $\text{span}(\mathbf{U}_k)^\perp$ has an empty interior in $\text{span}(\mathbf{P}_{\ell-1})^\perp$ (see e.g. [21]), then the union on k (denoted by $\mathbf{V}_\ell$) also has an empty interior. Therefore, the interior of $\text{span}(\mathbf{P}_{\ell-1})^\perp \setminus \mathbf{V}_\ell$ is not empty. This means there exists some $\boldsymbol{p}_\ell \in \text{span}(\mathbf{P}_{\ell-1})^\perp$ that is not in this union which ends the proof. ∎

REFERENCES

[1] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[2] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications,*, vol. 31, no. 2, pp. 160–171, 2013.

[3] H. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 3464–3467.

[4] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. Eighth International Symposium on Signal Processing and Its Applications (ISSPA '05)*, Sydney, Australia, 2005.

[5] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2012.

[6] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journ. Sel. Areas in Comm.*, vol. 31, no. 2, pp. 264–273, 2013.

[7] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[8] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, 2010.

[9] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, 2014.

[10] N. Shariati, J. Wang, and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Transactions on Signal Processing,*, vol. 62, no. 1, pp. 107–120, 2014.

[11] A. Adhikary, J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[12] 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*, vol. 63, no. 23, pp. 6169–6183, 2015.

[13] B. Tomasi and M. Guillaud, "Pilot length optimization for spatially correlated multi-user MIMO channel estimation," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2015.

[14] A. Decurninge, M. Guillaud, and D. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, San Diego, CA, USA, 2015.

[15] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[16] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Feb. 2007.

[17] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proc. American Control Conference (ACC)*, 2004, pp. 3273–3278.

[18] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, 2000.

[19] 3GPP Technical Specification Group Radio Access Network, "Study on 3D channel model for LTE (release 12)," Tech. Rep. TR 36.873, 2015.

[20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx.

[21] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*. Scott Foresman & Co, Glenview, IL, 1967.