FDD MASSIVE MIMO CHANNEL SPATIAL COVARIANCE CONVERSION USING PROJECTION METHODS

L. Miretti⋆, R. L. G. Cavalcante, and S. Stańczak

Fraunhofer Heinrich Hertz Institute

ABSTRACT

Knowledge of second-order statistics of channels (e.g. in the form of covariance matrices) is crucial for the acquisition of downlink channel state information (CSI) in massive MIMO systems operating in the frequency division duplexing (FDD) mode. Current MIMO systems usually obtain downlink covariance information via feedback of the estimated covariance matrix from the user equipment (UE), but in the massive MIMO regime this approach is infeasible because of the unacceptably high training overhead. This paper considers instead the problem of estimating the downlink channel covariance from uplink measurements. We propose two variants of an algorithm based on projection methods in an infinite-dimensional Hilbert space that exploit channel reciprocity properties in the angular domain. The proposed schemes are evaluated via Monte Carlo simulations, and they are shown to outperform current state-of-the art solutions in terms of accuracy and complexity, for typical array geometries and duplex gaps.

Index Terms— FDD Massive MIMO, covariance matrix, angular reciprocity, projection methods, subspace estimation.

1. INTRODUCTION

In this work we address the problem of estimating the downlink (DL) spatial channel covariance matrix $\mathbf{R}^d$ in massive MIMO systems that operate in the frequency division duplexing (FDD) mode. The availability of a reliable estimate of $\mathbf{R}^d$ is a key ingredient in solving the problem of DL channel state information (CSI) acquisition, which is indeed one of the main performance bottlenecks of large-scale array systems [1].

In particular, in the massive MIMO regime, conventional DL channel estimation techniques (e.g. the ones currently implemented in LTE) require a prohibitively large estimation overhead, expressed in terms of the length of pilot sequences. In extreme cases, the training time may even exceed the channel coherence time, making a reliable CSI acquisition not possible [2]. Existing solutions are therefore typically based on downlink-uplink (UL) channel reciprocity of time division duplexing (TDD) systems. Owing to the small number of antennas at the terminals, UL channel estimates can be obtained within the channel coherence time. However, in FDD systems this reciprocity is in general not available, and the envisioned approaches to the CSI acquisition problem typically rely on the existence of a lower dimensional representation of the channel vector in the large-scale array regime. The approaches can be divided into two main categories: methods based on compressed sensing (CS) and methods based on second-order statistics. An example of the first category is a CS based technique with dictionary learning proposed in [3]. Although promising, CS techniques do not take into account the space-time correlation properties of the channel, which are often modeled

by the well-known WSS assumption (see Sect. 2). In contrast, the approaches based on second order statistics exploit these properties, and they have been shown to reduce effectively the effort for DL CSI acquisition [4–6].

Since direct estimation of $\mathbf{R}^d$ by using DL training sequences is a very challenging problem, this work proposes a novel technique to infer $\mathbf{R}^d$ from the observed UL covariance $\mathbf{R}^u$, which is easier to estimate in practice. The proposed approach has also the benefit of eliminating continuous covariance feedback from the user equipment. Related state-of-the-art solutions in literature include:

- [7] Resampling of $\mathbf{R}^u$ for a uniform linear array (ULA) at a different wavelength by using cubic splines.
- [8] (and the follow-up study [9]) Interpolation of $\mathbf{R}^d$ from $\mathbf{R}^u$ and a dictionary of stored $(\mathbf{R}^d, \mathbf{R}^u)$ pairs measured at different UE locations.
- [10] Definition of a frequency calibration matrix obtained via a truncated Fourier series representation of the so called *angular power spectrum* (APS).

The main underlying assumption of the state-of-the-art techniques and of this work is the channel reciprocity in the angular domain, which is here modeled with the frequency invariance property of the APS (see Sect. 2). Because of its crucial role establishing the connection between $\mathbf{R}^d$ and $\mathbf{R}^u$, the core part of this work is devoted to the development of an accurate technique for APS estimation given $\mathbf{R}^u$. Unlike related studies, we formalize the problem as a *convex feasibility* problem, which enable us to apply very effective solutions based on projection methods in an infinite-dimensional Hilbert space. The resulting scheme is shown to outperform the existing solutions in several aspects (see Sect. 5). In fact, it achieves estimation accuracy and flexibility comparable to [8] (the most accurate and robust algorithm considered so far) but with complexity comparable to [7] and [10] (simple dictionary-less approaches).

This study is structured as follows: In Sect. 2 we introduce the channel model. Two variants of the proposed algorithm are described in Sect. 3, while practical implementation aspects for uniform linear arrays (ULA) are detailed in Sect. 4. In Sect. 5 we evaluate the performance of the algorithms with numerical simulations, and we highlight the advantages of the proposed scheme over the competing approaches.

Notation: We use boldface to denote vectors and matrices. $(\cdot)^T$ and $(\cdot)^H$ denote respectively the transpose and Hermitian transpose, and $\|\cdot\|_F$ the Frobenius norm. $L^2[I]$ denotes the set of all square Lebesgue integrable functions over the interval $I \subset \mathbb{R}$. Given a Hilbert space, we denote by $x^{(i)} \rightharpoonup x$ a sequence $(x^{(i)})_{i \in \mathbb{N}}$ weakly convergent to a point x. We use $\Re[\cdot]$ and $\Im[\cdot]$ to denote, respectively, the real and the imaginary parts. Throughout the paper, superscripts $(\cdot)^u$ and $(\cdot)^d$ indicate respectively UL and DL matrices, vectors, or functions when we need to emphasize the dependency on the carrier frequency.

⋆L. Miretti is now with EURECOM, France.

2. SYSTEM DESCRIPTION AND PROBLEM STATEMENT

For simplicity, we consider a flat-fading MIMO channel between a base station (BTS) with $N \gg 1$ antennas and a single-antenna user equipment (UE) in a 2D (azimuth-only) scenario. However, we point out that our results can be can be extended to more general settings.

By sampling the time-variant channel vector $\mathbf{h}(t)$ at intervals corresponding to the channel coherence period T_c, a classical channel model (e.g., see [5]) assumes $\mathbf{h}[k] := \mathbf{h}(kT_c)$, $k \in \mathbb{Z}$, to be a zero-mean WSS circularly-symmetric Gaussian process that is white in time, while being correlated in the spatial domain, such that

$$\mathbf{h}[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}) \quad \text{i.i.d.} \ .$$

In typical communication models (e.g., see the 3GPP technical report [11]), the spatial covariance matrix $\mathbf{R} \in \mathbb{C}^{N \times N}$ takes the form

$$\mathbf{R} = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta, \tag{1}$$

where the vector-valued function $\mathbf{a} : [-\pi, \pi] \to \mathbb{C}^N$, with the nth coordinate function denoted by a_n, describes the array response of the BTS for a given direction of arrival/departure (DoA/DoD) $\theta \in [-\pi, \pi]$; and $\rho : [-\pi, \pi] \longrightarrow \mathbb{R}^+$ denotes the so called *angular power spectrum* (APS) determining the average received/transmitted power per unit angle. In the following, we assume ρ, $\Re[a_n]$, and $\Im[a_n]$ to belong to $L^2[-\pi, \pi]$. According to the WSS assumption, $\mathbf{R}$ is invariant over time. In practice, it is a slowly-varying parameter, since real channels can be safely assumed to be WSS just over a certain window of time T_{WSS}, which in usual scenarios is several order of magnitude larger than the channel coherence time T_c [6,12].

This paper addresses the problem of estimating the DL covariance matrix $\mathbf{R}^d$ from an observation of the UL covariance matrix $\mathbf{R}^u$. In FDD systems, $\mathbf{R}^d$ and $\mathbf{R}^u$ are different, but strongly linked through the APS which, unlike the array response, exhibits strong frequency correlation properties. Therefore, in what follows, we assume the APS to be frequency invariant, an assumption that holds for typical FDD duplex gaps (see [13,14] for further details). By this assumption and (1), we can write $\mathbf{R}^u \in \mathbb{C}^{N \times N}$ and $\mathbf{R}^d \in \mathbb{C}^{N \times N}$ as

$$\mathbf{R}^u = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta, \tag{2}$$

$$\mathbf{R}^d = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta. \tag{3}$$

In the following, we assume that the array responses $\mathbf{a}^u$ and $\mathbf{a}^d$ are known; this knowledge is cell-independent and it holds for the entire lifetime of the antenna array. In Sect. 3, to simplify the description of the proposed solutions, we assume perfect knowledge of $\mathbf{R}^u$. However, later in Sect. 4.2 and Sect. 5 we drop this assumption.

To motivate our work, let us consider DL channel estimation in the massive MIMO regime. If no knowledge on second-order statistics is available, standard techniques based on orthogonal training sequences require pilots of length at least N. However, if $\mathbf{R}^d$ is known, this knowledge can be exploited to reduce the training overhead. More precisely, if $p := rank(\mathbf{R}^d) < N$ (see [6] for a detailed analysis of this condition) the Karhuen-Loeve transform yields

$$\mathbf{h}[k] = \mathbf{U}_p \boldsymbol{\Delta}_p^{\frac{1}{2}} \mathbf{w}[k],$$

with $\mathbf{w}[k] \in \mathbb{C}^{p \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ i.i.d, and $\mathbf{U}_p \boldsymbol{\Delta}_p^{\frac{1}{2}} \in \mathbb{C}^{N \times p}$ corresponding to the p non-zero eigenvalues and respective eigenvectors of the eigen-decomposition $\mathbf{R}^d = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^H$. So if $\mathbf{R}^d$ is known, then, as shown in [4,5], it is possible to reduce the training sequence length from N to p. Moreover, in [6], this fact is exploited to reduce pilot contamination effects.

3. SPATIAL COVARIANCE CONVERSION VIA PROJECTION METHODS

We now present two variants of a novel scheme for estimating $\mathbf{R}^d$ based on the knowledge of $\mathbf{R}^u$. By recalling the model and the assumptions introduced in Sect. 2, the main idea can be summarized in two steps as follows:

1. Given $\mathbf{R}^u$, we obtain an estimate $\hat{\rho}$ of the APS from equation (2), and by exploiting known properties of ρ.

2. We compute an estimate of $\mathbf{R}^d$ from (3), by substituting ρ with its estimate $\hat{\rho}$.

In particular, in the first step, we formulate the APS estimation problem as a *convex feasibility problem*. The aforementioned two proposed variants of the scheme differ from each other in the exploitation of the known properties of the APS, which leads to two algorithms with different accuracy-complexity trade-offs.

3.1. APS estimation via projection onto a linear variety (Algorithm 1)

We rewrite (2) as a system of equations of the form

$$r_m^u = \int_{-\pi}^{\pi} \rho(\theta)g_m^u(\theta)d\theta \quad m = 1 \dots M, \quad M = 2N^2, \tag{4}$$

where $r_m^u \in \mathbb{R}$ is the mth element of $\mathbf{r}^u := \text{vec}([\Re\{\mathbf{R}^u\} \ \Im\{\mathbf{R}^u\}])$, and $g_m^u : [-\pi, \pi] \longrightarrow \mathbb{R}$ is the mth element of the corresponding vectorization of the matrix $\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H$. In general, since covariance matrices are Hermitian, the number of different equations is at most $N(N-1)$. Notice that it is possible to modify the definition of the vec$(\cdot)$ operator such that all the duplicated equations of (4) are removed, but for notation simplicity in this section this trivial operation is omitted.

Now let $\mathcal{H}$ be the Hilbert space of real functions in $L^2[-\pi, \pi]$ equipped with the inner product $\langle f, g \rangle = \int_{-\pi}^{\pi} f(\theta)g(\theta)d\theta$. By Sect. 2, ρ and g_m^u are members of $\mathcal{H}$, so that (4) can can be written as

$$r_m^u = \langle \rho, g_m^u \rangle \quad m = 1 \dots M. \tag{5}$$

The inverse problem of finding ρ given g_m^u and r_m^u, $m = 1 \dots M$, is obviously ill-posed. However, by using the set-theoretic paradigm [15–18], we propose to estimate ρ by solving

$$\text{find } \rho^* \in V := \cap_{m=1}^M V_m \neq \emptyset, \tag{6}$$

where $V_m := \{\rho \in \mathcal{H} : \langle \rho, g_m^u \rangle = r_m^u\}$ for $m = 1 \dots M$.

Among all the possible solutions of (6) (all of which are equivalent given (2)), motivated by the low complexity algorithm implementation that we show in the following, we choose the minimum norm solution

$$\hat{\rho} = \arg \min_{\rho^* \in V} \|\rho^*\|,$$

which corresponds to the orthogonal projection $P_V(0)$ of the zero vector onto the linear variety V [19, Sect. 3.10]. This projection has the following well-known closed-form expression:

$$\hat{\rho}(\theta) = \sum_{m=1}^{M} \alpha_m g_m^u(\theta), \tag{7}$$

where $\boldsymbol{\alpha} := [\alpha_1 \dots \alpha_M]$ is a solution to the linear system

$$\mathbf{r}^u = \mathbf{G}^u \boldsymbol{\alpha}, \tag{8}$$

$$\mathbf{G}^u = \begin{bmatrix} \langle g_1^u, g_1^u \rangle & \langle g_1^u, g_2^u \rangle & \cdots & \langle g_1^u, g_M^u \rangle \\ \langle g_2^u, g_1^u \rangle & \langle g_2^u, g_2^u \rangle & \cdots & \langle g_2^u, g_M^u \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_M^u, g_1^u \rangle & \langle g_M^u, g_2^u \rangle & \cdots & \langle g_M^u, g_M^u \rangle \end{bmatrix},$$

which is guaranteed to have at least one solution (notice that we do not assume linear independence of the g_i^u). Furthermore, from the

projection theorem, all solutions lead to the unique projection $\hat{\rho}$.

We obtain an estimate of $\mathbf{R^d}$ by replacing ρ in the DL equivalent of (5) with its estimate $\hat{\rho}$ obtained in (7):

$$\hat{r}^d_m = \langle \hat{\rho}, g^d_m \rangle = \sum_{l=1}^{M} \alpha_l \langle g^u_l, g^d_m \rangle \quad m = 1 \dots M, \qquad (9)$$

which can be rewritten in matrix form as

$$\hat{\mathbf{r}}^d = \mathbf{Q}\boldsymbol{\alpha},$$

where $\hat{\mathbf{r}}^d$ is an estimate of the vector $\mathbf{r}^d := \mathrm{vec}([\Re\{\mathbf{R}^d\}\ \Im\{\mathbf{R}^d\}])$, $\boldsymbol{\alpha}$ is a solution of the linear system (8) given the UL measurements $\mathbf{r}^u = \mathbf{G}^u\boldsymbol{\alpha}$ as mentioned above, and

$$\mathbf{Q} = \begin{bmatrix} \langle g^d_1, g^u_1 \rangle & \langle g^d_1, g^u_2 \rangle & \cdots & \langle g^d_1, g^u_M \rangle \\ \langle g^d_2, g^u_1 \rangle & \langle g^d_2, g^u_2 \rangle & \cdots & \langle g^d_2, g^u_M \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g^d_M, g^u_1 \rangle & \langle g^d_M, g^u_2 \rangle & \cdots & \langle g^d_M, g^u_M \rangle \end{bmatrix}.$$

It is important to underline that both $\mathbf{G}^u$ and $\mathbf{Q}$ depend only on the array geometry, and they can thus be computed or measured only once for the entire system lifetime.

3.2. Exploiting further properties of the APS (Algorithm 2)

In many applications, additional prior knowledge about the APS ρ is often available (for example, support information). If this knowledge can be expressed in terms of closed convex sets, then it is possible to narrow the set of candidate solutions of (6) to obtain more accurate APS estimates. By separating the real and imaginary part of (1), and by working in the space of real functions, the previous algorithm in Sect. 3.1 already implicitly takes into account the knowledge that ρ is real valued. In the following, we propose an extension of the previous algorithm by taking into account that, being a power spectrum, ρ is always non-negative. More precisely, we look at the problem

$$\text{find } \rho^* \in C := V \cap Z, \qquad (10)$$

where V is the linear variety defined in (6) and $Z = \{\rho \in \mathcal{H} : (\forall \theta \in [-\pi, \pi]) \quad \rho(\theta) \geq 0\}$ is the closed convex set of non-negative functions in $\mathcal{H}$. A solution to (10) can be found by applying one of the many existing iterative projection methods for convex feasibility problems available in literature. These methods typically produce a sequence $(\rho^{(i)})_{i \in \mathbb{N}} \subset \mathcal{H}$ such that $\rho^{(i)} \rightharpoonup \rho^* \in C$. In particular, we use the following fast iterative method called *extrapolated alternating projection method (EAPM)*, given by [20]

$$\rho^{(i+1)} = \rho^{(i)} + \nu K_i \left[P_V(P_Z(\rho^{(i)})) - \rho^{(i)} \right] \quad (\forall i \in \mathbb{N}), \quad (11)$$

where $\nu \in (0, 2)$ is a step size, and K_i is the extrapolation parameter defined as

$$K_i = \begin{cases} \dfrac{\|P_Z(\rho^{(i)}) - \rho^{(i)}\|^2}{\|P_V(P_Z(\rho^{(i)})) - \rho^{(i)}\|^2}, & \text{if } \rho^{(i)} \notin Z \\ 1, & \text{if } \rho^{(i)} \in Z \end{cases}.$$

The initial condition $\rho^{(0)} \in V$ can be arbitrary, and we choose $\rho^{(0)} = P_V(0)$, defined in Sect. 3.1. The projection $P_V : \mathcal{H} \to V \subset \mathcal{H}$ onto V is given by [19, Chapter 3]

$$P_V(x) = x - \sum_{m=1}^{M} \beta_m g^u_m + P_V(0),$$

with $\boldsymbol{\beta} := [\beta_1 \dots \beta_M]$ being a solution to the linear system $\mathbf{b} = \mathbf{G}^u\boldsymbol{\beta}$ where the mth element of $\mathbf{b}$ is given by $b_m = \langle x, g^u_m \rangle$ and g^u_m, G^u are defined in Sect. (3.1). The projection $P_Z : \mathcal{H} \to Z \subset \mathcal{H}$ is given by [16, p. 284]

$$P_Z(x) = \begin{cases} x(\theta), & \text{for } x(\theta) \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Now, by proceeding along the same lines as in Sect. 3.1, an estimate of $\mathbf{R^d}$ can be obtained by

$$\hat{r}^d_m = \langle \hat{\rho}, g^d_m \rangle \quad m = 1 \dots M.$$

which has the same form as (9) except that $\hat{\rho}$ results from (11).

4. IMPLEMENTATION FOR UNIFORM LINEAR ARRAY

In this section we discuss an implementation of the proposed schemes to a uniform linear array (ULA) with N antennas at the BTS. The array response of the ULA is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{j2\pi \frac{d}{\lambda} \sin\theta} & \cdots & e^{j2\pi \frac{d}{\lambda}(N-1)\sin\theta} \end{bmatrix}^T,$$

where $d \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ denote, respectively, the inter-antenna spacing and the carrier wavelength.

4.1. Analytical expressions for $\mathbf{G}^u$ and $\mathbf{Q}$

Since the ULAs are not able to distinguish among a DoA/DoD θ and its reciprocal $\theta + \pi$, we assume that the multipath components are confined to the interval $[-\pi/2, \pi/2]$, and we modify the definition of the scalar product for $\mathcal{H}$ accordingly, such that $\langle f, g \rangle = \int_{-\pi/2}^{\pi/2} f(\theta)g(\theta)d\theta$. This assumption is supported by the fact that real systems often work with a similar or even narrower cell sectorization.

For ULA, the covariance matrix is positive semi-definite Hermitian Toeplitz, so it can be completely represented by its first column. By redefining $\mathrm{vec}(\mathbf{A}) := \mathbf{a}_1$, where $\mathbf{a}_1$ indicates the first column of $\mathbf{A}$, we can prove that the matrices $\mathbf{G}^u$ and $\mathbf{Q}$ defined in Sect. 3.1 have the following analytical form expressed in terms of the Bessel function of the first kind, zero order $J_0 : \mathbb{R} \to \mathbb{R}$:

$$\mathbf{G}^u = \frac{\pi}{2N^2} \begin{bmatrix} \mathbf{G}_{\Re} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{\Im} \end{bmatrix} \quad \mathbf{Q} = \frac{\pi}{2N^2} \begin{bmatrix} \mathbf{Q}_{\Re} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\Im} \end{bmatrix},$$

where the elements corresponding to the (n, m)-entries of $\mathbf{G}_{\Re}$, $\mathbf{G}_{\Im}$, $\mathbf{Q}_{\Re}, \mathbf{Q}_{\Im} \in \mathbb{R}^{N \times N}$ are given by

$$\mathbf{G}_{\Re,nm} = J_0(x_{nm}) + J_0(y_{nm}), \quad \mathbf{Q}_{\Re,nm} = J_0(p_{nm}) + J_0(q_{nm}),$$

$$\mathbf{G}_{\Im,nm} = J_0(x_{nm}) - J_0(y_{nm}), \quad \mathbf{Q}_{\Im,nm} = J_0(p_{nm}) - J_0(q_{nm}),$$

where

$$x_{nm} = 2\pi \frac{d}{\lambda^u}(n - m), \quad p_{nm} = 2\pi d \left(\frac{n-1}{\lambda^d} - \frac{m-1}{\lambda^u} \right),$$

$$y_{nm} = 2\pi \frac{d}{\lambda^u}(n + m - 2), \quad q_{nm} = 2\pi d \left(\frac{n-1}{\lambda^d} + \frac{m-1}{\lambda^u} \right).$$

The proof is omitted because of the space limitation.

4.2. Imperfect $\mathbf{R}^u$ knowledge

In this section we analyze a scenario in which the BTS has access only to a UL sample covariance $\bar{\mathbf{C}}^u := \frac{1}{K} \sum_{k=1}^{K} \hat{\mathbf{h}}^u[k](\hat{\mathbf{h}}^u[k])^H$ computed from a limited number K of channel estimates defined as $\hat{\mathbf{h}}^u[k] = \mathbf{h}^u[k] + \mathbf{z}[k], \mathbf{z}[k] \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I})$ i.i.d. The noise power is obtained by setting a given per-antenna received $SNR := \frac{\mathbf{E}[|h_n|^2]}{\sigma_z^2}$. Let $\mathcal{H}_M$ be the Hilbert space of all $N \times N$ Hermitian matrices whose inner product is defined by $\langle A, B \rangle = \mathrm{trace}(B^H A)$, and let $\mathcal{C}, \mathcal{T}$ be the subsets of $\mathcal{H}_M$ composed respectively by positive semi-definite (PSD) and Toeplitz matrices. The matrix $\bar{\mathbf{C}}^u$ is shown in [12] to be a sufficient statistic for estimating $\mathbf{C}^u := \mathbf{E}[\hat{\mathbf{h}}^u[k](\hat{\mathbf{h}}^u[k])^H] = \mathbf{R}^u + \sigma_z^2\mathbf{I}$, and in [21] the matrix $\bar{\mathbf{R}}^u := \bar{\mathbf{C}}^u - \sigma_z^2\mathbf{I}$ is used to obtain the maximum-likelihood (ML)-PSD estimate of $\mathbf{R}^u$ by projecting it onto $\mathcal{C}$. The direct feeding of either $\bar{\mathbf{R}}^u$ or the ML-PSD estimate as input to the proposed algorithms may result in poor performance because the Toeplitz assumption imposed by the ULA is not satisfied.
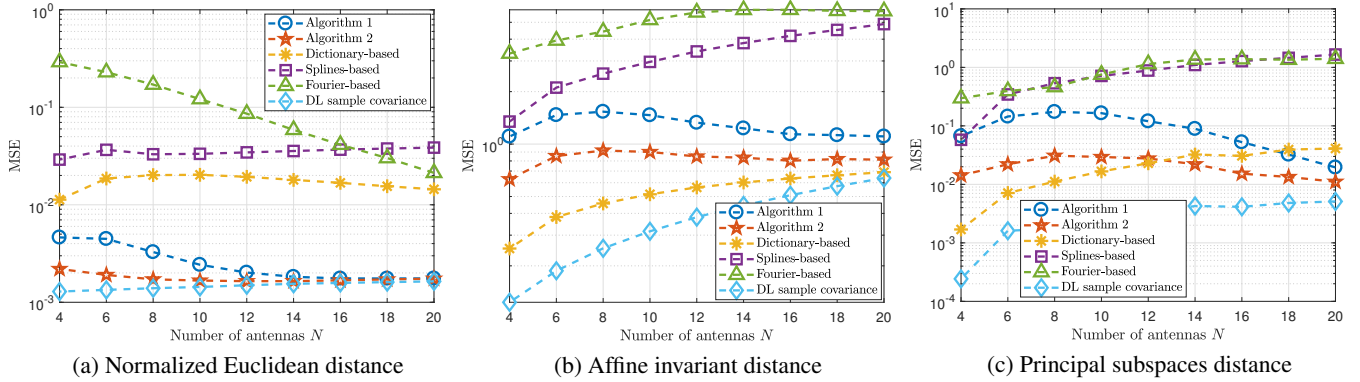
(a) Normalized Euclidean distance   (b) Affine invariant distance   (c) Principal subspaces distance

Fig. 1. Simulation results: comparison of different DL covariance estimators vs number of BTS antennas N.

To overcome this problem, we propose to feed as input the projection of $\bar{\mathbf{R}}^u$ onto the set $\mathcal{T}_+ := \mathcal{C} \cap \mathcal{T}$, which imposes the desired Toeplitz structure. More precisely, we use $\hat{\mathbf{R}}^u = \arg\min_{\mathbf{X} \in \mathcal{T}_+} \|\mathbf{X} - \bar{\mathbf{R}}^u\|_F$ as the input. Since the projections on $\mathcal{C}$ and $\mathcal{T}$ are known [22] and easy to compute, it is possible to compute $\hat{\mathbf{R}}^u$ by applying standard methods such as the Dykstra's or Haugazeau's algorithm [23, Chapter 29]. In this work we use the approach described in [22].

5. SIMULATIONS AND FINAL REMARKS

We simulate a typical model for the APS in cellular environments based on the well-known *geometry-based stochastic channel model* (GSCM) [13], where ρ is assumed to be composed of a weighted superposition of probability density functions, supported by the intuition that the multipath components mainly originate from a set of Q clusters of scatterers surrounding the BTS and the UE:

$$\rho(\theta) = \sum_{q=1}^{Q} f_q(\theta) \alpha_q.$$

As an example, in the following we assume Q uniformly drawn from $\{1, 2, 3, 4, 5\}$, Gaussian distributions $f_q \sim \mathcal{N}\left(\phi_q, \Delta_q^2\right)$ with ϕ_q uniformly drawn from $[-\pi/3, \pi/3]$ and standard deviation (also called *angular spread*) Δ_q uniformly drawn from $[3°, 8°]$, weights α_q uniformly drawn from $[0, 1]$ and further normalized such that $\sum_{q=1}^{Q} \alpha_q = P_{RX}$, where P_{RX} indicates the total received power. These statistical quantities are introduced to emulate the effect of different scattering patterns corresponding to random user locations.

A ULA is assumed for the BTS operating at UL/DL carrier wavelengths of $\lambda = 3 \cdot 10^8 / f$ with $f = 1.8$ Ghz and 1.9 Ghz respectively. The antenna spacing d is set to half UL wavelength. Channel realizations conditioned on $\mathbf{R}$ computed from (1) are given by $\mathbf{h} = \mathbf{R}^{\frac{1}{2}} \mathbf{w}$, with $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The BTS is assumed to have access only to a UL sample covariance matrix computed from $K = 1000$ noisy channel estimates as described in Sect. 4.2. The performance of the two algorithms defined in Sect. 3.1 and 3.2 are compared with the algorithms proposed in [7], [9], and [10], referred, respectively, to *splines*-based, *dictionary*-based, and *Fourier*-based. The DL sample covariance, obtained with the same number of samples and SNR as for the UL, is used as a baseline. For fairness, all the sample covariances used in this comparison are corrected with the Toeplitzation procedure outlined in Sect. 4.2. The accuracy of an estimate $\hat{\mathbf{R}}$ of $\mathbf{R}$ is evaluated in terms of the mean square error $MSE := \mathbf{E}[e^2(\mathbf{R}, \hat{\mathbf{R}})]$, where $e(\cdot, \cdot)$ is a given error metric. In particular, we consider:

- The normalized Euclidean distance
  $e(\mathbf{R}, \hat{\mathbf{R}}) := \|\mathbf{R} - \hat{\mathbf{R}}\|_F / \|\mathbf{R}\|_F$.
- [8, 24] The affine invariant distance in the Riemannian space of PSD matrices $e(\mathbf{R}, \hat{\mathbf{R}}) := \| \log(\mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1} \mathbf{R}^{\frac{1}{2}}) \|_F$.
- [24] The Grassmanian distance between the principal subspaces $\mathbf{U}_p, \hat{\mathbf{U}}_p$ defined from $\mathbf{R}, \hat{\mathbf{R}}$ by considering their eigenvectors corresponding to the minimum number p of largest eigenvalues λ_n satisfying $\sum_{n=1}^{p} \lambda_n / \sum_{n=1}^{N} \lambda_n \geq 95\%$. The metric is then $e(\mathbf{R}, \hat{\mathbf{R}}) := \sqrt{\sum_{n=1}^{p} \gamma_n^2}$, where $\cos(\gamma_n)$ are the eigenvalues of $\mathbf{U}_p^H \hat{\mathbf{U}}_p$. This metric is particularly meaningful for the massive MIMO channel estimation problem, where a reliable signal subspace knowledge plays a crucial role.

The statistical mean is then obtained by Monte-Carlo simulations. For every Monte-Carlo run, a new APS and a SNR level $\in [10, 30]$ (dB) are drawn.

Figure 1 compares the algorithms for different numbers of BTS antennas N. Both the proposed algorithms approach the performance of the DL sample covariance estimator, as the number of constraints in the convex feasibility problem grows with N. The performance of both algorithms are comparable or better (depending on the metric and on the number of antennas) than the *dictionary*-based method, which in principle can achieve extremely high accuracy given that the dictionary is sufficiently large (here we used 1000 entries). However, the proposed algorithms are dictionary-less, thus not requiring any overhead for dictionary acquisition. *Algorithm 1* has the same low complexity of the *Fourier*-based method, but it achieves much better accuracy. Compared to *Algorithm 1*, *Algorithm 2* shows better performance, especially in the low N region, where the prior information about the positivity of the APS becomes important. However, the performance gains are achieved at the cost of higher complexity, which is due to the fact that the algorithm requires the numerical evaluation of integrals of the form $\int_{-\pi}^{\pi} x(\theta) d\theta$.

In summary, we have shown that the set-theoretic approach can be applied effectively to the problem of channel spatial covariance conversion. Compared to the competing approaches, the two variants of the proposed scheme are shown to achieve the high accuracy of the *dictionary*-based method, but with the low complexity of the *Fourier*-based and the *splines*-based methods.

6. REFERENCES

[1] H. Huh, A. M. Tulino, and G. Caire, "Network mimo with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[3] Y. Ding and B. D. Rao, "Channel estimation using joint dictionary learning in FDD massive mimo systems," in *IEEE Global Conference on Signal and Information Processing (Global-SIP)*, 2015, pp. 185–189.

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

[5] A. Adhikary, J. Nam, J.-Y. 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.

[6] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 942–953, 2014.

[7] M. Jordan, A. Dimofte, X. Gong, and G. Ascheid, "Conversion from uplink to downlink spatio-temporal correlation with cubic splines," in *IEEE 69th Vehicular Technology Conference*, 2009, pp. 1–5.

[8] A. Decurninge, M. Guillaud, and D.T.M. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *IEEE Globecom*, 2015, pp. 1–6.

[9] A. Decurninge, M. Guillaud, and D.T.M. Slock, "Riemannian coding for covariance interpolation in massive MIMO frequency division duplex systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2016, pp. 1–5.

[10] Ying-Chang Liang and Francois P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 222–232, 2001.

[11] 3GPP, "Spatial channel model for multiple input multiple output MIMO simulations (release 14)," Tech. Rep. TR 25.996 V14.0.0, 3rd Generation Partnership Project (3GPP), 2017.

[12] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, 2017.

[13] A.F. Molish, *Wireless Communications*, Wiley, second edition, 2010.

[14] K. Hugl, K. Kalliola, and J. Laurila, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," *Proc. COST 273 Technical Document TD (02)*, vol. 66, pp. 7, 2002.

[15] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[16] H. Stark, Y. Yang, and Y. Yang, *Vector space projections: a numerical approach to signal and image processing, neural nets, and optics*, John Wiley & Sons, Inc., 1998.

[17] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical functional analysis and optimization*, vol. 25, pp. 593–617, 2005.

[18] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 97–123, 2011.

[19] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, jan 1998.

[20] H.H. Bauschke, P.L. Combettes, and S.G. Kruk, "Extrapolation algorithm for affine-convex feasibility problems," *Numerical Algorithms*, vol. 41, no. 3, pp. 239–274, 2006.

[21] D. Neumann, M. Joham, L. Weiland, and W. Utschick, "Low-complexity computation of LMMSE channel estimates in massive MIMO," in *Proceedings of the 19th International ITG Workshop on Smart Antennas (WSA 2015)*. VDE, 2015, pp. 1–6.

[22] K.M. Grigoriadis, A.E. Frazho, and R.E. Skelton, "Application of alternating convex projection methods for computation of positive toeplitz matrices," *IEEE transactions on signal processing*, vol. 42, no. 7, pp. 1873–1875, 1994.

[23] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, vol. 408, Springer, 2011.

[24] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.