Covariance Estimation with Projected Data: Applications to CSI Covariance Acquisition and Tracking

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

***Abstract*—We consider the problem of covariance estimation with projected or missing data, and in particular the application to spatial channel covariance estimation in a multi-user Massive MIMO wireless communication system with arbitrary (possibly time-varying and/or non-orthogonal) pilot sequences. We introduce batch and online estimators based on the expectation-maximization (EM) approach, and provide sufficient conditions for their asymptotic (for large sample sizes) unbiasedness. We analyze their application to both uplink and downlink Massive MIMO, and provide numerical performance benchmarks.**

I. Introduction

Channel state information (CSI) acquisition represents an important problem in the multi-user Massive MIMO (Multiple-Input Multiple-Output) scenario [1], where accurate 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]. In this context, the channels typically exhibit a large degree of spatial (across the antennas) correlation [3]. A number of recent works on Massive MIMO have made use of the assumption that the spatial correlation of the wireless channel state process (sometimes called statistical CSI, or channel distribution information, CDI) is available to the device in charge of the baseband signal processing. The assumed channel models are generally such that statistical CSI can be identified to the second-order statistics, through a spatial covariance matrix associated to each channel state process. In particular, an abundant literature is dedicated to the topic of reducing the amount of reference symbols dedicated to the estimation of instantaneous CSI in multi-user systems, for which a variety of approaches have been proposed [4]–[11]; all these techniques have in common the idea that the prior information contained in the statistics can substantially reduce the amount of reference symbols dedicated to CSI estimation, by allowing either a denser reuse of identical pilot sequences, or the use of non-orthogonal pilot sequences without sacrificing estimation accuracy.

Most of these works however, do not cover the topic of how the required statistical CSI is obtained. Although a vast statistical literature on covariance estimation is available, focusing among others on large dimensional analysis [12], many of the existing approaches do not directly apply to the problem at hand if one takes into account the dynamic user scheduling and pilot sequence allocation (see [4]–[11]) resulting from evolving channel statistics.

Furthermore, since the role of statistical CSI is to mitigate or remove the effect of pilot contamination, the available statistical CSI itself should not be contaminated. One possible solution to this issue is to have dedicated pilots for covariance estimation, which can be less frequent but require a lower reuse factor to reduce or suppress pilot contamination; note however that this is not the only possible approach [13], [14].

In this article, we argue that statistical CSI acquisition in Massive MIMO should be formulated as a problem of covariance estimation with missing data (see e.g. [15]). This particular point of view has been adopted in the context of subspace estimaton and/or tracking [16], [17]. We propose in the following practical maximum likelihood (ML) approaches based on the expectation-maximization (EM) algorithm. This approach has the advantage of 1) gracefully handling the case of scheduling and dynamic pilot sequence allocation, and 2) providing asymptotically contamination-free covariance estimates without requiring dedicated pilot sequences.

The article is organized as follows. In Section II, we introduce the channel model; Section III introduces the proposed EM-based approach, discusses its asymptotic properties, and introduces special cases as well as an online version of the estimator. The application of these results to the context of multi-user Massive MIMO is discussed in Section IV. Section V introduces numerical simulation results.

II. Problem Statement

We address the covariance estimation of a multivariate zero-mean circular Gaussian random vector $\boldsymbol{h} \in \mathbb{C}^N$. We suppose that we observe T noisy projections of independent identically distributed samples $\mathbf{h}_1, \dots, \mathbf{h}_T$:

$$\mathbf{y}_t = \mathbf{P}_t^H \mathbf{h}_t + \mathbf{n}_t \tag{1}$$

where $\mathbf{y}_t \in \mathbb{C}^{L_t}$ is the observed random vector at time t, L_t represents the dimension of the projection at time t (where typically $L_t < N$), $\mathbf{P}_t \in \mathbb{C}^{N \times L_t}$ contains the projection directions and $\mathbf{n}_t \in \mathbb{C}^{L_t}$ representing the noise is modeled by a zero-mean Gaussian random vector of known covariance

ISBN 978-0-9928626-7-1 © EURASIP 2017

equal to $\sigma^2\mathbf{I}$. Let $\mathbf{R}$ denote the covariance of the random vector $\mathbf{h}$. Then, for each $t = 1, \dots, T$,

$$\mathbf{y}_t = \mathbf{P}_t^H\mathbf{h}_t + \mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{P}_t^H\mathbf{R}\mathbf{P}_t + \sigma^2\mathbf{I}_{L_t}).$$

We will denote in the following $\mathbf{\Sigma}_t(\mathbf{R}) = \mathbf{P}_t^H\mathbf{R}\mathbf{P}_t + \sigma^2\mathbf{I}_{L_t}$. In the context of the considered application to wireless communications, $\mathbf{h}$ will represent the channel state and $\mathbf{P}_t$ the pilot sequence of length L_t used during the fading interval indexed by t.

By independence of the realizations $\mathbf{y}_1, \dots, \mathbf{y}_T$, the maximum likelihood estimator of $\mathbf{R}$ is

$$\hat{\mathbf{R}}_T = \arg \min_{\substack{\mathbf{R} \succeq 0 \\ \mathbf{R} \in \mathcal{S}}} \mathcal{L}_T(\mathbf{R}) \tag{2}$$

where $\mathcal{S}$ is a linear subspace of $\mathbb{C}^{N \times N}$ and $\mathcal{L}_T$ is proportional to the negative log-likelihood of $\mathbf{y}_1, \dots, \mathbf{y}_T$. $\mathcal{S}$ can be used to impose a structure on $\hat{\mathbf{R}}_T$, e.g. to incorporate prior model information about the independence of certain coefficients in $\mathbf{h}$. Since the density of each circular Gaussian random vector $\mathbf{y}_t$ is given by $f_t(\mathbf{y}_t) = \frac{1}{\pi^N \det(\mathbf{\Sigma}_t(\mathbf{R}))} e^{-\mathbf{y}_t^H\mathbf{\Sigma}_t(\mathbf{R})^{-1}\mathbf{y}_t}$, we consider $\mathcal{L}_T$ as

$$\mathcal{L}_T(\mathbf{R}) = \sum_{t=1}^{T} \log\det(\mathbf{\Sigma}_t(\mathbf{R})) + \mathbf{y}_t^H\mathbf{\Sigma}_t(\mathbf{R})^{-1}\mathbf{y}_t.$$

We now introduce approaches to solve (2).

III. ML covariance estimator based on Expectation Maximization

A direct minimization of $\mathcal{L}_T$ thanks to gradient descent techniques has been proposed in the context of missing data and factor analysis; see e.g. [15]. However, this approach requires the use of second-order derivatives of the likelihood which may be intractable for large N. A way to compute a local minimum of the negative log-likelihood is to consider the EM algorithm [18]. Indeed, we are in the situation where we observe only a part $\mathbf{y}_t$ of the total sample $(\mathbf{h}_t^T, \mathbf{n}_t^T)^T$. The EM algorithm allows then to compute the maximum likelihood of such partial observations (see [19] which uses an EM algorithm in the context of missing observations).

The EM algorithm is an iterative scheme whereby at each iteration, the covariance estimate $\mathbf{R}^{(n)}$ is updated into $\mathbf{R}^{(n+1)}$ as follows:

E-step

The E-step consists in computing the expectation of the negative log-likelihood of the total sample $\mathbf{h}_1, \dots, \mathbf{h}_T$ (we omit the noise terms $\mathbf{n}_1, \dots, \mathbf{n}_T$ since their distribution do not depend on $\mathbf{R}$), namely

$$\mathcal{L}_{\mathbf{h}}(\mathbf{R}) = T\log\det(\mathbf{R}) + \sum_{t=1}^{T} \mathbf{h}_t^H\mathbf{R}^{-1}\mathbf{h}_t \tag{3}$$

with respect to the unobserved part of the data sample only (note that if $\mathbf{R}$ is rank-deficient, the inversion and the $\log\det$ operator may be adapted by considering the restriction to the range of $\mathbf{R}$). In other terms, we compute $\mathbb{E}[\mathcal{L}_{\mathbf{h}}(\mathbf{R})|\mathbf{y}_t, \mathbf{R}^{(n)}] = T\log\det(\mathbf{R}) + \sum_{t=1}^{T} \mathbb{E}[\mathbf{h}_t^H\mathbf{R}^{-1}\mathbf{h}_t|\mathbf{y}_t, \mathbf{R}^{(n)}]$ where $\mathbb{E}[\cdot|\mathbf{y}_t, \mathbf{R}^{(n)}]$ denotes the expectation conditioned on the observation $\mathbf{y}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}(\mathbf{R}^{(n)}))$.

Since $\mathbf{h}_t$ and $\mathbf{y}_t$ are jointly Gaussian, the random vector $(\mathbf{h}_t^T, \mathbf{y}_t^T)^T$ is Gaussian with zero mean and covariance equal to

$$\begin{pmatrix} \mathbf{R}^{(n)} & \mathbf{R}^{(n)}\mathbf{P}_t \\ \mathbf{P}_t^H\mathbf{R}^{(n)} & \mathbf{P}_t^H\mathbf{R}^{(n)}\mathbf{P}_t + \sigma^2\mathbf{I} \end{pmatrix}.$$

Therefore, the random vector $(\mathbf{h}_t|\mathbf{y}_t, \mathbf{R}^{(n)})$ is also Gaussian with mean $\boldsymbol{\mu}_t = \mathbf{R}^{(n)}\mathbf{P}_t\mathbf{\Sigma}_t(\mathbf{R}^{(n)})^{-1}\mathbf{y}_t$ and covariance $\mathbf{C}_t = \mathbf{R}^{(n)} - \mathbf{R}^{(n)}\mathbf{P}_t\mathbf{\Sigma}_t(\mathbf{R}^{(n)})^{-1}\mathbf{P}_t^H\mathbf{R}^{(n)}$. Then, it holds $\mathbb{E}[\mathbf{h}_t^H\mathbf{R}^{-1}\mathbf{h}_t|\mathbf{y}_t, \mathbf{R}^{(n)}] = \mathrm{Tr}[\mathbf{R}^{-1}(\mathbf{R}^{(n)} + \mathbf{\Delta}_t^{(n)})]$ where

$$\begin{cases} \mathbf{\Delta}_t^{(n)} = \mathbf{M}_t^{(n)}(\mathbf{y}_t\mathbf{y}_t^H - \mathbf{\Sigma}_t(\mathbf{R}^{(n)}))(\mathbf{M}_t^{(n)})^H \\ \mathbf{M}_t^{(n)} = \mathbf{R}^{(n)}\mathbf{P}_t^H\mathbf{\Sigma}_t(\mathbf{R}^{(n)})^{-1}. \end{cases} \tag{4}$$

M-step

The M-step consists in minimizing the conditional expectation computed for the E-step (thus maximizing the conditional expectation of the likelihood) with respect to the parameter $\mathbf{R}$ i.e.

$$\mathbf{R}^{(n+1)} = \arg \min_{\substack{\mathbf{R} \succeq 0 \\ \mathbf{R} \in \mathcal{S}}} \log\det(\mathbf{R}) + \mathrm{Tr}\left[\mathbf{R}^{-1}\left(\mathbf{R}^{(n)} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Delta}_t^{(n)}\right)\right]. \tag{5}$$

If the matrix $\left(T\mathbf{R}^{(n)} + \sum_{t=1}^{T}\mathbf{\Delta}_t^{(n)}\right)$ is not positive definite, the minimum is undefined because the minimized function is unbounded. We will therefore give an approximation of the M-step by letting $(.)_+$ denote the projection operator on the space of definite positive matrices, and Π denote the projection onto the subspace $\mathcal{S}$ and define

$$\mathbf{R}^{(n+1)} = \left(\Pi\left(\mathbf{R}^{(n)} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Delta}_t^{(n)}\right)\right)_+. \tag{6}$$

A. Accelerated EM

Since EM algorithm is a first-order scheme, its convergence towards a local minimum may be slow. Then, we used an off-the-shelf scheme presented in [20] named squared iterative method (SQUAREM) in order to accelerate the convergence. The idea is to compute two iterations of the EM algorithm and to apply a Cauchy-Barzilai-Borwein method to compute the next iterate (see Algorithm 1 where the function *EM update* refers to eq. (6)).

B. Asymptotic properties of the EM estimator

The asymptotic properties of the maximum likelihood estimate $\hat{\mathbf{R}}_T$ are related to standard asymptotic properties for independent non identically distributed random variables; see [21]. Let us give the following theorem which states that the proposed EM algorithm is asymptotically unbiased if the projections are chosen appropriately (a sketch of the proof is provided in Appendix A)

Algorithm 1 Accelerated EM.

```
while convergence is not achieved
    $\mathbf{R}_1 =$ EM update($\mathbf{R}^{(n)}$) according to (6).
    $\mathbf{R}_2 =$ EM update($\mathbf{R}_1$).
    $\mathbf{L} = \mathbf{R}_1 - \mathbf{R}^{(n)}$.
    $\mathbf{L}_2 = \mathbf{R}_2 - \mathbf{R}_1 - \mathbf{L}$.
    $\ell = -\frac{\|\mathbf{L}\|_F}{\|\mathbf{L}_2\|_F}$.
    while likelihood has not increased
        $\mathbf{R}^{(n+1)} = (\Pi(\mathbf{R}^{(n)} - 2\ell\mathbf{L} + \ell^2\mathbf{L}_2))_+$.
        $\ell \leftarrow \frac{\ell-1}{2}$.
    end
end
```

Theorem 1. *Suppose $\sigma^2 > 0$ and that there exists $P_{\min}, P_{\max} > 0$ such that $P_{\min}^2 \mathbf{I}_{L_t} \preceq \mathbf{P}_t^H \mathbf{P}_t \preceq P_{\max}^2 \mathbf{I}_{L_t}$ for all t. Suppose moreover that $\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \|\mathbf{P}_t^H \mathbf{x}\|^4 > 0$ holds for any $\mathbf{x} \neq 0$. Then,*

$$\hat{\mathbf{R}}_T \xrightarrow[T\to\infty]{} \mathbf{R} \ \ a.s.$$

Remark that this result shows that a incorporating sufficient randomness in the choice of $\mathbf{P}_t$ is sufficient to achieve consistent covariance estimation.

C. Online version of EM

Since N may be large, each iteration of the EM algorithm may be costly. Moreover, received signals $\mathbf{y}_t$ may be considered as a process. For these reasons, an online version of the EM algorithm is of interest in order to track the covariance. Some online versions of EM have been proposed, e.g. an algorithm based on a stochastic version of EM using the knowledge of Fisher information matrix (see [22]). In this section, we present an online version following [23], for its simple derivation in our setup. Indeed, the idea is to replace the expectation step by an approximation of the likelihood conditional expectation $\hat{\mathcal{L}}_{t+1}(\mathbf{R}) = (1-\gamma_{t+1})\hat{\mathcal{L}}_t(\mathbf{R}) + \gamma_{t+1}\mathbb{E}[\mathcal{L}_{\mathbf{h}_{t+1}}(\mathbf{h}_{t+1};\mathbf{R})|\mathbf{y}_{t+1}, \mathbf{R}_t]$ where $\mathcal{L}_{\mathbf{h}_{t+1}}$ is the negative log-likelihood of $\mathbf{h}_{t+1}$ (see eq. (3)). Therefore, if we define step parameters $(\gamma_t)_{t=1\ldots T}$, each iteration results in

$$\left\{\begin{array}{l} \mathbf{S}_{t+1} = (1-\gamma_{t+1})\mathbf{S}_t + \gamma_{t+1}(\mathbf{R}_t + \boldsymbol{\Delta}_{t+1}(\mathbf{y}_{t+1}, \mathbf{R}_t)) \\ \mathbf{R}_{t+1} = (\Pi(\mathbf{S}_{t+1}))_+ \end{array}\right. \tag{7}$$

with

$$\left\{\begin{array}{l} \boldsymbol{\Delta}_{t+1}(\mathbf{y}_{t+1}, \mathbf{R}_t) = \mathbf{M}_t(\mathbf{y}_{t+1}\mathbf{y}_{t+1}^H - \boldsymbol{\Sigma}_t(\mathbf{R}_t))\mathbf{M}_t^H \\ \mathbf{M}_t = \mathbf{R}_t \mathbf{P}_t^H \boldsymbol{\Sigma}_t(\mathbf{R}_t)^{-1}. \end{array}\right. \tag{8}$$

This algorithm converges to a local minimum under mild assumptions (see [23]). In particular, the convergence proof requires that $\sum \gamma_t = \infty$ and $\sum \gamma_t^2 < \infty$, as is the case for most stochastic gradient descent. However, in practice, we do not have access to infinitely many channel realizations sharing the same covariance; we will therefore consider $\gamma_t = \gamma$, which yields a tracking covariance estimator, which does not converge asymptotically but rather behaves like a stochastic process. In this case, the choice of γ directly impacts the asymptotic variance of the process.

D. Special case of a fixed basis

We may assume that the projection directions are a subset of a fixed basis of vectors. Let us denote by $\mathbf{f}_1, \ldots, \mathbf{f}_N$ a basis of N orthogonal unit-norm vectors in $\mathbb{C}^N$. Then, for each time slot t, the projections directions are drawn from a permutation σ_t, i.e. $\mathbf{P}_t = (\mathbf{f}_{\sigma_t(1)}, \ldots, \mathbf{f}_{\sigma_t(L_t)})$ (in the multi-user MIMO scenario, σ_t would result from the scheduling decision at instant t). Therefore, $\mathbf{y}_t = \boldsymbol{\Pi}_t^H \mathbf{F}^H \mathbf{h}_t + \mathbf{n}_t$, with $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_N)$ and $\boldsymbol{\Pi}_t$ the subset of the permutation matrix corresponding to σ_t. For the sake of notational simplicity, we assume in this section that $\mathcal{S} = \mathbb{C}^{N\times N}$.

1) Full observation (i.e. orthogonal) case: The full observation case corresponds to the case where $\boldsymbol{\Pi}_t = \mathbf{I}$ for all t. The ML estimate is then given in closed form by $\hat{\mathbf{R}}_T = \mathbf{F}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{y}_t\mathbf{y}_t^H - \sigma^2\mathbf{I}\right)_+ \mathbf{F}^H$.

Proof: Let us denote $\lambda_1 \leq \cdots \leq \lambda_N$ the eigenvalues of $\frac{1}{T}\sum_{t=1}^T \mathbf{y}_t\mathbf{y}_t^H$ and $\mathbf{U}$ its corresponding eigenvector matrix. Similarly, denote $\mu_1 \leq \cdots \leq \mu_N$ the eigenvalues of $\mathbf{R} + \sigma^2\mathbf{I}_N$ and $\mathbf{V}$ its corresponding eigenvectors. Then, since $\mathbf{F}$ is a unitary matrix, it holds after some calculus

$$\mathbf{F}^H\hat{\mathbf{R}}\mathbf{F} + \sigma^2\mathbf{I}_N = \arg \min_{\substack{\mu_i \geq \sigma^2 \\ \mathbf{V}\mathbf{V}^H = \mathbf{I}_n}} \sum_{i=1}^N \log(\mu_i)$$
$$+\mathrm{Tr}\left[\mathbf{V}\mathrm{diag}(\mu_i)_{i=1\ldots N}^{-1}\mathbf{V}^H\mathbf{U}\mathrm{diag}(\lambda_i)_{i=1\ldots N}\mathbf{U}^H\right].$$

Let us denote $\mathbf{A} = \mathbf{U}^H\mathbf{V}$. Therefore $\mathbf{A}$ is a unitary matrix and $\mathrm{Tr}\left[\mathbf{V}\mathrm{diag}(\mu_i)_{i=1\ldots N}^{-1}\mathbf{V}^H\mathbf{U}\mathrm{diag}(\lambda_i)_{i=1\ldots N}\mathbf{U}^H\right] \geq \sum_{i=1}^N \frac{\lambda_i}{\mu_i}$.

Therefore, the minimum is obtained for $\mathbf{V} = \mathbf{U}$ and μ_i satisfying $\arg\min_{\mu_i\geq\sigma^2}\sum_{i=1}^N \log(\mu_i) + \frac{\lambda_i}{\mu_i}$, i.e. $\mu_i = \max(\sigma^2, \lambda_i)$ which ends the proof. ■

2) Partial observation (i.e. non-orthogonal) case: In this case, if $\mathbf{F} = \mathbf{I}_N$, the ML is given by

$$\hat{\mathbf{R}}_T = \arg\min_{\mathbf{R}\succeq 0} \sum_t \quad \log\det(\boldsymbol{\Pi}_t^H\mathbf{R}\boldsymbol{\Pi}_t + \sigma^2\mathbf{I})$$
$$+\mathbf{y}_t(\boldsymbol{\Pi}_t^H\mathbf{R}\boldsymbol{\Pi}_t + \sigma^2\mathbf{I})^{-1}\mathbf{y}_t.$$

This case corresponds to the case of covariance estimation with missing data considered e.g. in [19].

IV. Massive MIMO channel covariance estimation

In this section, the estimation framework introduced before is applied to covariance estimation and tracking in the context of pilot-aided CSI acquisition. Following the Massive MIMO framework where the base station (BS) is equipped with many antennas, we consider the estimation of BS-side channel correlation, i.e. transmitter-side correlation in the downlink case, and receiver-side correlation in the uplink case.

A. Downlink case

In the downlink CSI acquisition scenario, we consider single-antenna users receiving a signal sent by a N-antennas BS corresponding to a pilot matrix $\mathbf{P}_t$ of size $N \times L_t$ at each slot $t = 1\ldots T$; covariance estimation can be performed by each terminal in the system based on the received signal $\mathbf{y}_t$.

ISBN 978-0-9928626-7-1 © EURASIP 2017

In this context, $\mathcal{S}$ becomes equal to $\mathcal{C}^{N\times N}$ with the notations of eq. (2), i.e. the projector Π becomes the identity mapping.

We may remark that the EM iteration may be rewritten as a function of the MMSE estimate of $\mathbf{h}$. Indeed, this MMSE estimate is expressed by $\hat{\mathbf{h}}_t(\mathbf{R}) = \mathbf{R}\mathbf{P}_t(\mathbf{P}_t^H\mathbf{R}\mathbf{P}_t + \sigma^2\mathbf{I})^{-1}\mathbf{y}_t$. Then, the EM iteration step may be written as $\mathbf{R}^{(n+1)} = (\mathbf{R}^{(n)} + \hat{\mathbf{h}}_t(\mathbf{R}^{(n)})\hat{\mathbf{h}}_t(\mathbf{R}^{(n)})^H - \mathbb{E}_n[\hat{\mathbf{h}}_t(\mathbf{R}^{(n)})\hat{\mathbf{h}}_t(\mathbf{R}^{(n)})^H])_+$ where $\mathbb{E}_n$ denotes the expectation with $\mathbf{y}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_t(\mathbf{R}^{(n)}))$.

Note also that in the case of orthogonal pilots with reuse, a fixed association between users and pilot sequences should be avoided since it can lead to biased covariance estimates (in this case, the hypotheses of Theorem 1 do not hold); however this can be easily circumvented by incorporating some randomness in the pilot allocation to make it evolve over time, e.g. through a pseudo-random permutation.

B. Uplink case

Conversely to the downlink case, in the uplink the signal received at a BS contains pilot signals coming from multiple terminals, and it is necessary to consider the joint estimation of the channel covariances of K users. Let us consider an uplink CSI acquisition scenario where pilots are sent simultaneously from all single-antenna terminals to the BS in order to estimate the channel coefficients. Let $\mathbf{p}_{k,t} \in \mathbb{C}^{L_t}$ denote the pilot sequence of length L_t symbols transmitted by terminal k. The signal received at the BS, $\mathbf{Y}_t = [\mathbf{y}(1), \ldots, \mathbf{y}(L_t)] \in \mathbb{C}^{N\times L_t}$, is obtained as

$$\mathbf{Y}_t = \mathbf{H}_t\mathbf{P}_t + \mathbf{N}_t, \tag{9}$$

where $\mathbf{H}_t = [\mathbf{h}_{1,t}, \ldots, \mathbf{h}_{K,t}]$ is the column concatenation of the channel vectors from the K terminals to the N antennas at the BS, $\mathbf{P}_t = [\mathbf{p}_{1,t}^T; \ldots; \mathbf{p}_{K,t}^T] \in \mathbb{C}^{K\times L_t}$ is the matrix containing the training sequences sent by the User Terminals (UTs), and $\mathbf{N}_t \in \mathbb{C}^{N\times L_t}$ represents additive noise.

Vectorizing (9) yields

$$\mathbf{y}_t := \mathrm{vec}(\mathbf{Y}_t) = \tilde{\mathbf{P}}_t^H \mathrm{vec}(\mathbf{H}_t) + \mathrm{vec}(\mathbf{N}_t), \tag{10}$$

where $\tilde{\mathbf{P}}_t = (\mathbf{P}_t^* \otimes \mathbf{I}_N)$ and $(.)^*$ denotes the complex conjugate operator. The vector $\mathbf{h}_t := \mathrm{vec}(\mathbf{H}_t)$ is modelled as a Gaussian random vector of covariance

$$\tilde{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{R}_K \end{pmatrix}$$

which corresponds to a structured model where the fading processes are known to be independent across the users. With the notations of eq. (2), the subspace $\mathcal{S}$ becomes the space of such $\tilde{\mathbf{R}}$ and the projector Π the linear operator setting the off-diagonal blocks of size $N \times N$ to 0. Then, the EM iteration step is given by

$$\mathbf{R}^{(n+1)} = \left(\Pi\left(\mathbf{R}^{(n)} + \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Delta}_t^{(n)}\right)\right)_+$$

where $\mathbf{\Delta}_t^{(n)}$ is given by eq. (8).

V. Numerical results

We have compared the different estimators acording to the following scenario. We consider a covariance $\mathbf{R}$ drawn from a Wishart distribution with r degrees of freedom, i.e. there exists a matrix $\mathbf{X} \in \mathbb{C}^{N\times r}$ with independent zero-mean circular Gaussian entries of variance 1 such that $\mathbf{R} = \mathbf{X}\mathbf{X}^H$. The per-user pilot sequences (columns of $\mathbf{P}_t$) are independently drawn from a Haar distribution on the space of unit-norm vectors. Furthermore, let us assume that the dimension of $\mathbf{P}_t$ is independent of t, i.e. $L_t = L$. Finally, we assume that $\sigma^2 = 10^{-2}$ which corresponds to a Signal-to-Noise ratio equal to 20dB. With these parameters, we have compared different estimators: 1) the sample covariance estimator (SCM) assuming full observations of channel realizations (which is therefore a lower bound in terms of estimation error), i.e. $\hat{\mathbf{R}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{h}_t\mathbf{h}_t^H$, 2) the batch EM estimator (Algorithm 1), 3) the tracking EM estimator (7)–(8) with different values of the tracking parameter $\gamma = 0.1$ and $\gamma = 0.01$, where we chose as criterion of comparison between $\mathbf{R}$ and its estimate $\hat{\mathbf{R}}_T$ the estimation error measured by $\frac{\|\mathbf{R}-\hat{\mathbf{R}}_T\|_F}{\|\mathbf{R}\|^{\frac{1}{2}}\|\hat{\mathbf{R}}_T\|_F^{\frac{1}{2}}}$.

In Figs. 1 and 2, we represent these estimation errors with respect to the number of samples T. We average the estimation error among $N_{MC} = 10$ Monte Carlo runs. We may observe that the EM estimator is already close to the SCM as soon as the pilot length L is large enough. The "convergence" of the two tracking estimators is obviously slower than for the batch estimators and do not decrease below a floor corresponding to the asymptotic variance of the tracking process. However, we may remark that the lower the tracking step γ, the lower the asymptotic variance.

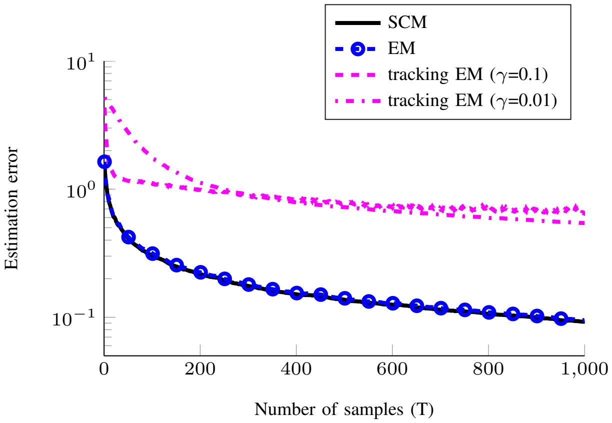

Fig. 1. Estimation error with respect to the number of samples ($L = 10, r = 10, N = 64$).

VI. Conclusion

We have discussed the application of covariance estimation with projected or missing data to the problem of spatial channel covariance estimation in a multi-user Massive MIMO wireless communication system, including time-varying and/or non-orthogonal pilot sequences, for both uplink and downlink

ISBN 978-0-9928626-7-1 © EURASIP 2017

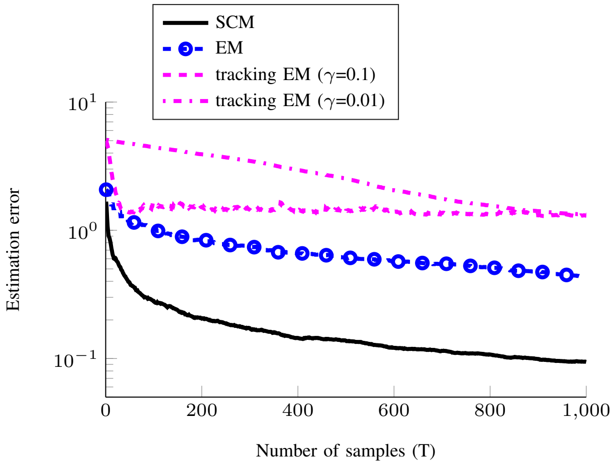

Fig. 2. Estimation error with respect to the number of samples ($L = 2, r = 10, N = 64$).

cases. We introduced batch and online estimators based on the expectation-maximization (EM) approach, and provided sufficient conditions for their asymptotic unbiasedness.

Appendix A
Proof of Theorem 1

Let us check the hypotheses of [21, Thm. 1] and verify that the theorem applies. Since the result is given for real random variables, we simply remark that $\left(\text{Re}(\mathbf{z})^T, \text{Im}(\mathbf{z})^T\right)^T$ is a zero mean Gaussian with covariance $\frac{1}{2}\begin{pmatrix} \text{Re}(\mathbf{R}) & -\text{Im}(\mathbf{R}) \\ \text{Im}(\mathbf{R}) & \text{Re}(\mathbf{R}) \end{pmatrix}$. Note that the negative log-likelihood function $\mathcal{L}_T$ satisfies $\frac{1}{T}\mathcal{L}_T(\mathbf{S}) \geq \log(P_{\min}^2 \lambda_{\min}(\mathbf{S}))$ for any $\mathbf{S}$ where $\lambda_{\min}(\mathbf{S})$ denotes the minimum eigenvalue of $\mathbf{S}$. Then, for any $C > 0$, there exists $B > 0$ such that for T large enough, if $\mathbf{S} \succeq B\mathbf{I}_N$, then $\frac{1}{T}\mathcal{L}_T(\mathbf{S}) \geq C$. Moreover, it holds $\frac{1}{T}\mathcal{L}_T(\mathbf{I}_N) \leq \log(P_{\max}^2 + \sigma^2) + \frac{P_{\max}^2}{\sigma^2 T}\sum_{t=1}^T \mathbf{h}_t^H \mathbf{h}_t + 1$. Since $\frac{1}{T}\sum_{t=1}^T \mathbf{h}_t^H \mathbf{h}_t \rightarrow \mathbb{E}[\|\mathbf{h}_t\|^2]$ almost surely (a.s.), $\frac{1}{T}\mathcal{L}_T(\mathbf{I}_N)$ is majored independently from T a.s. Therefore, since the maximum likelihood $\hat{\mathbf{R}}_T$ minimizes $\mathcal{L}_T$, we can restrict the parameter space for T large enough to $\Theta = \{\mathbf{R} \text{ s.t. } \mathbf{R} \preceq M\mathbf{I}\}$. Assumptions (C1) and (C2) are then easily verified. Furthermore, the likelihood function difference satisfies $|\mathcal{L}_t(\mathbf{S}) - \mathcal{L}_t(\mathbf{R})| \leq 2\log(MP_{\max}^2 + \sigma^2) + 2P_{\max}^2\sigma^{-2}\mathbf{h}_t^H\mathbf{h}_t + 2$ which implies (C3') and (C5).

Let us consider $\mathbf{S} \neq \mathbf{R}$ and note $h(\mathbf{A}) := -\log\det(\mathbf{A}) + \text{Tr}(\mathbf{A}) - N$. Then the expected log-likelihood difference is $\mathbb{E}[\mathcal{L}_t(\mathbf{S})] - \mathbb{E}[\mathcal{L}_t(\mathbf{R})] = h(\mathbf{A}_t)$, with $\mathbf{A}_t = \mathbf{\Sigma}_t(\mathbf{R})\mathbf{\Sigma}_t(\mathbf{S})^{-1}$. Remark that h is a convex function attaining its maximum in $\mathbf{I}$ with an Hessian matrix equal to the information matrix. On the other hand, let $g : \mathbf{A} \mapsto \|\mathbf{A} - \mathbf{I}\|^2$, it holds $\mathbf{A}_t \preceq \frac{P_{\max}}{\sigma^2}\mathbf{R} + \mathbf{I}$, i.e. $\mathbf{A}_t$ belongs to a compact and for any $\mathbf{A}_t$ in this compact, $\nabla^2 h(\mathbf{A}_t) = \mathbf{A}_t^{-1} \otimes \mathbf{A}_t^{-T}$ with $M_1 = \frac{\sigma^2}{\sigma^2 + P_{\max}\lambda_{\max}(\mathbf{R})}$. Since $g(\mathbf{I}) = 0$ and $\nabla g(\mathbf{I}) = 0$, we can integrate the inequality, i.e. for any $\mathbf{A}_t$ in this compact $\frac{1}{T}\sum_{t=1}^T h(\mathbf{A}_t) \geq \sum_{i=1}^N \lambda_i^2 M_2 \left(\frac{1}{T}\sum_{t=1}^T \|\mathbf{P}_t^H \mathbf{x}_i\|^4\right) > 0$ with λ_i and $\mathbf{x}_i$ the eigenvalues and eigenvectors of $\mathbf{R} - \mathbf{S}$ respectively and $M_2 = \frac{M_1}{\sigma^2 + P_{\max}\lambda_{\max}(\mathbf{S})}$ which ensures that (C4') is also verified.

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 Select. Areas in Comm.*,, vol. 31, no. 2, pp. 160–171, 2013.

[3] 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.

[4] 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.

[5] 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.

[6] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[7] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Transactions on Wireless Communications*, 2017.

[8] D. Neumann, M. Joham, and W. Utschick, "CDI precoding for massive MIMO," in *Proc. International ITG Conference on Systems, Communications and Coding (SCC)*, Feb. 2015.

[9] S. E. Hajri, M. Assaad, and G. Caire, "Scheduling in massive MIMO: User clustering and pilot assignment," in *Proc. Allerton Conference on Communication, Control, and Computing*, sep 2016.

[10] E. Björnson, L. Sanguinetti, and M. Debbah. (2016) Massive MIMO with imperfect channel covariance information. [Online]. Available: https://arxiv.org/pdf/1612.04128v2.pdf

[11] 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.

[12] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[13] P. Ferrand, A. Decurninge, M. Guillaud, and L. G. Ordóñez, "Efficient channel state information acquisition in massive MIMO systems using non-orthogonal pilots," in *Proc. Workshop on Smart Antennas (WSA)*, Mar. 2017.

[14] D. Neumann, M. Joham, and W. Utschick. (2017) Covariance matrix estimation in massive MIMO. [Online]. Available: https://arxiv.org/abs/1705.02895

[15] C. Finkbeiner, "Estimation for the multiple factor model when data are missing," *Psychometrika*, vol. 44, pp. 409–420, 1979.

[16] Y. Chi, Y. C. Eldar, and R. Calderbank, "Petrels: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. on Signal Processing*, vol. 61, no. 23, pp. 5947–5959, 2013.

[17] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. on Signal Processing*, vol. 65, no. 2, pp. 303–318, 2016.

[18] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *Journal of the Royal Statistic Society, Series B*, vol. 39, pp. 1–38, 1977.

[19] M. Jamshidian and P. Bentler, "ML estimation of mean and covariance structures with missing data using complete data routines," *Journal of Educational and Behavioral Statistics*, vol. 24, pp. 21–41, 1999.

[20] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any EM algorithm," *Scandinavian Journal of Statistics*, vol. 35, pp. 335–353, 2008.

[21] B. Hoadley, "Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case," *Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1977–1991, 1971.

[22] D. Titterington, "Recursive parameter estimation using incomplete data," *Journal of the Royal Stat. Soc., Series B*, vol. 46, pp. 257–267, 1984.

[23] O. Cappé and E. Moulines, "On-line expectation-maximization algorithm for latent data models," *Journal of the Royal Statistic Society, Series B*, vol. 71, no. 3, pp. 593–613, 2009.

ISBN 978-0-9928626-7-1 © EURASIP 2017