

Massive MIMO Channel Estimation Considering Pilot Contamination and Spatially Correlated Channels

Felipe A. P. de Figueiredo, Dimas A. M. Lemes, Claudio F. Dias, and Gustavo Fraidenraich

In this letter, we present a study on linear channel estimators and their respective mean square error (MSE) expressions acknowledging spatially correlated channels and pilot contamination. We also investigate the impact of imperfect channel covariance matrix knowledge.

Introduction: In real propagation environments, channels are spatially correlated [1], which means that the elements of the channel are, to some extent, correlated. Generally, the channel covariance matrices exhibit spatial correlation substantiating in the different diagonal elements and non-zero off-diagonal elements. This letter studies linear channel estimators for multi-cell multi-user massive MIMO systems with spatially correlated channels and pilot contamination.

System Model: In this work, we consider a multi-cell multi-user system with L cells where each one of the cells has a BS at its center with M co-located antennas and K single antenna users. We consider correlated Rayleigh fading channels, and therefore, the $M \times 1$ channel vector from the k th user in the l th cell to the M antennas at the i th BS is defined by $\mathbf{g}_{ilk} = [g_{ilk1}, g_{ilk2}, \dots, g_{ilkM}]^T \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{R}_{ilk})$, where $\mathbf{R}_{ilk} \in \mathbb{C}^{M \times M}$ is the positive semi-definite channel covariance matrix. It is important to notice that in our case, \mathbf{R}_{ilk} is not a scaled identity matrix, but describes the spatial propagation environment and array geometry (*i.e.*, it describes macroscopic effects, which include the average path-loss in different spatial directions and the spatial channel correlation).

We assume that users in different cells transmit at the same time-frequency resources (a typical scenario in massive MIMO) and that the pilot reuse factor is one, the worst possible use case scenario. The N -length pilot sequence sent by the k th user, $\phi_k \in \mathbb{C}^N$, exhibits the following property, $\phi_k^H \phi_k = 1, \forall k$. Extending it to the $N \times K$ matrix containing all the K pilots, $\Phi = [\phi_1, \phi_2, \dots, \phi_K]$, it becomes $\Phi^H \Phi = \mathbf{I}_M$. The received uplink training sequences at the i th BS can be represented as a $M \times N$ matrix, which is defined as

$$\mathbf{Y}_i = \sqrt{p} \sum_{l=1}^L \mathbf{G}_{il} \Phi^H + \mathbf{N}_i, \quad (1)$$

where p is the pilot power or average pilot signal to noise ratio (SNR) and \mathbf{N}_i is a $M \times N$ noise matrix with independent and identically distributed elements following $\mathcal{CN}(0, 1)$.

LS Channel Estimation: A sufficient statistic for estimating the channel vector, \mathbf{g}_{ilk} , at the i th BS is given by

$$\hat{\mathbf{g}}_{ilk}^{\text{LS}} = \mathbf{z}_{ik} = \frac{1}{\sqrt{p}} \mathbf{Y}_i \phi_k = \sum_{l=1}^L \mathbf{g}_{ilk} + \mathbf{w}_{ik}, \quad (2)$$

where $\mathbf{w}_{ik} = \frac{1}{\sqrt{p}} \mathbf{N}_i \phi_k \sim \mathcal{CN}(\mathbf{0}_M, \frac{1}{p} \mathbf{I}_M)$ and $\hat{\mathbf{g}}_{ilk}^{\text{LS}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{Q}_{ik})$, with $\mathbf{Q}_{ik} = \sum_{l=1}^L \mathbf{R}_{ilk} + \frac{1}{p} \mathbf{I}_M$. The estimation error vector, $\tilde{\mathbf{g}}_{ilk}^{\text{LS}} = \hat{\mathbf{g}}_{ilk}^{\text{LS}} - \mathbf{g}_{ilk}$, is distributed as $\tilde{\mathbf{g}}_{ilk}^{\text{LS}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{Q}_{ik} - \mathbf{R}_{ilk})$ and is not independent of $\hat{\mathbf{g}}_{ilk}^{\text{LS}}$, having a covariance matrix defined as $\text{cov}(\hat{\mathbf{g}}_{ilk}^{\text{LS}}, \tilde{\mathbf{g}}_{ilk}^{\text{LS}}) = \mathbf{Q}_{ik} - \mathbf{R}_{ilk}$. The MSE per antenna of the LS estimator is given by

$$\eta_{ilk}^{\text{LS}} = \frac{1}{M} \mathbb{E}\{\|\hat{\mathbf{g}}_{ilk}^{\text{LS}} - \mathbf{g}_{ilk}\|^2\} = \frac{1}{M} \text{Tr}[\mathbf{Q}_{ik} - \mathbf{R}_{ilk}], \quad (3)$$

where $\text{Tr}[\cdot]$ is the Trace operator. As can be seen, the LS estimator does not rely on any prior information on the channel statistics, such as the large-scale fading coefficients. Additionally, this estimator is known to have inferior performance than the MMSE estimator [2].

Remark 1: If the elements of \mathbf{g}_{ilk} are i.i.d. circularly-symmetric complex normal variables for all i, l , and k , then $\mathbf{Q}_{ik} = \zeta_{ik} \mathbf{I}_M = (\sum_{l=1}^L \beta_{ilk} + \frac{1}{p}) \mathbf{I}_M$, and $\mathbf{R}_{ilk} = \beta_{ilk} \mathbf{I}_M$, and therefore, $\eta_{ilk}^{\text{LS}} = \zeta_{ik} -$

β_{ilk} , where $\beta_{ilk} = \frac{1}{M} \text{Tr}[\mathbf{R}_{ilk}]$ is the normalized trace that defines the average large-scale fading between the k th user in the i th cell and the i th BS.

Remark 2: Due to pilot contamination, $\eta_{ilk}^{\text{LS}} \rightarrow \frac{1}{M} \text{Tr}[\sum_{l=1, l \neq i}^L \mathbf{R}_{ilk}]$ as $p \rightarrow \infty$. And if \mathbf{g}_{ilk} is a i.i.d. complex Gaussian vector, then $\eta_{ilk}^{\text{LS}} \rightarrow \sum_{l=1, l \neq i}^L \beta_{ilk}$ as $p \rightarrow \infty$.

Remark 3: From (2) we see that $\hat{\mathbf{g}}_{ilk}^{\text{LS}} = \hat{\mathbf{g}}_{ilk}^{\text{LS}}, \forall l$, which means that the channel estimates are parallel vectors and therefore, the BS is unable to separate these users that transmitted the same pilot sequence.

MMSE Channel Estimation: The MMSE channel estimator of \mathbf{g}_{ilk} , $\forall k$, based on the observation \mathbf{Y}_i at the i th BS is defined by

$$\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \frac{\mathbf{R}_{ilk}}{\mathbf{Q}_{ik}} \mathbf{z}_{ik} = \mathbf{R}_{ilk} \mathbf{Q}_{ik}^{-1} \mathbf{z}_{ik}. \quad (4)$$

Due to the MMSE properties under the Gaussian model, the channel estimate $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}}$ and the estimation error, $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE}} = \mathbf{g}_{ilk} - \hat{\mathbf{g}}_{ilk}^{\text{MMSE}}$, are independent random vectors distributed as $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{R}_{ilk} \mathbf{Q}_{ik}^{-1} \mathbf{R}_{ilk})$ and $\tilde{\mathbf{g}}_{ilk}^{\text{MMSE}} \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{R}_{ilk} (\mathbf{I}_M - \mathbf{Q}_{ik}^{-1} \mathbf{R}_{ilk}))$. The estimator $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}}$ is uncorrelated with the de-spread received vector, \mathbf{z}_{ik} , and is consequently independent of it as both are jointly complex Gaussian distributed. The MSE per antenna of the MMSE estimator is given by

$$\eta_{ilk}^{\text{MMSE}} = \frac{1}{M} \mathbb{E}\{\|\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} - \mathbf{g}_{ilk}\|^2\} = \frac{1}{M} \text{Tr}[\mathbf{R}_{ilk} - \mathbf{R}_{ilk} \mathbf{Q}_{ik}^{-1} \mathbf{R}_{ilk}]. \quad (5)$$

Remark 4: If the elements of \mathbf{g}_{ilk} are i.i.d. circularly-symmetric complex normal variables for all i, l , and k , then $\mathbf{Q}_{ik} = \zeta_{ik} \mathbf{I}_M = (\sum_{l=1}^L \beta_{ilk} + \frac{1}{p}) \mathbf{I}_M$, and $\mathbf{R}_{ilk} = \beta_{ilk} \mathbf{I}_M$, and therefore, $\eta_{ilk}^{\text{MMSE}} = \beta_{ilk} - \frac{\beta_{ilk}^2}{\zeta_{ik}}$.

Remark 5: Due to pilot contamination, $\eta_{ilk}^{\text{MMSE}} \rightarrow \frac{1}{M} \text{Tr}[\mathbf{R}_{ilk} - \frac{\mathbf{R}_{ilk}^2}{\sum_{l=1}^L \mathbf{R}_{ilk}}]$ as $p \rightarrow \infty$. And if \mathbf{g}_{ilk} is a i.i.d. complex Gaussian vector, then $\eta_{ilk}^{\text{MMSE}} \rightarrow \beta_{ilk} \left(1 - \frac{\beta_{ilk}}{\sum_{l=1}^L \beta_{ilk}}\right)$ as $p \rightarrow \infty$.

Remark 6: If \mathbf{R}_{ilk} is invertible, then from (4) we see that $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \mathbf{R}_{ilk} \mathbf{R}_{ilk}^{-1} \hat{\mathbf{g}}_{ilk}^{\text{MMSE}}, \forall l$. In [3], the authors show that if $\mathbf{R}_{ilk}, \forall l$ are mutually asymptotically linearly independent, then the channels are not parallel vectors and consequently, the BS is able to separate users transmitting the same pilot sequence.

Remark 7: If \mathbf{g}_{ilk} is an i.i.d. complex Gaussian vector, then, again from (4), we see that $\hat{\mathbf{g}}_{ilk}^{\text{MMSE}} = \frac{\beta_{ilk}}{\beta_{ilk}} \hat{\mathbf{g}}_{ilk}^{\text{MMSE}}, \forall l$, meaning that the channel estimates are parallel vectors that only differ by the scaling factor, $\frac{\beta_{ilk}}{\beta_{ilk}}$, and therefore, the BS is also unable to separate users transmitting the same pilot sequence.

Approximate MMSE Estimation: In general, the acquisition of the covariance matrices, $\mathbf{R}_{ilk}, \forall i, l, k$, is a daunting task as it involves the estimation of LK matrices. However, a simple but yet effective solution to this problem comes from the observation of (4) and the finding that the sum of all covariance matrices plus the inverse of the pilot power, *i.e.*, \mathbf{Q}_{ik} , can be estimated. Therefore, we estimate \mathbf{Q}_{ik} and replace it back into (4). The classical approach to this estimation problem is to approximate the covariance matrix with the sample covariance matrix. The estimation of \mathbf{Q}_{ik} is based on the fact that $\mathbb{E}[\mathbf{z}_{ik} \mathbf{z}_{ik}^H] = \mathbf{Q}_{ik}$, which can be approximated by the sample covariance matrix,

$$\hat{\mathbf{Q}}_{ik} = \frac{1}{N_Q} \sum_{n=1}^{N_Q} \mathbf{z}_{ik}(n) \mathbf{z}_{ik}^H(n), \quad (6)$$

where $\mathbf{z}_{ik}(n), n = 1, \dots, N_Q$ are the N_Q different observations of (2). Additionally we see that $\mathbb{E}[\hat{\mathbf{Q}}_{ik}] = \mathbf{Q}_{ik}$. Note that this estimate is obtained from the several observations of the de-spread pilot signals, \mathbf{z}_{ik} , used for channel estimation, and thus, no extra pilots are necessary. The sample covariance matrix almost surely converges to the true covariance

matrix as $N_Q \rightarrow \infty$

$$\lim_{N_Q \rightarrow \infty} \hat{\mathbf{Q}}_{ik} = \lim_{N_Q \rightarrow \infty} \frac{1}{N_Q} \sum_{n=1}^{N_Q} \mathbf{z}_{ik}(n) \mathbf{z}_{ik}^H(n) \rightarrow \mathbf{Q}_{ik}, \quad (7)$$

which follows directly from the law of large numbers and the fact that the channels are assumed ergodic. Considering $[\hat{\mathbf{Q}}_{ik}]_j$ and $[\mathbf{Q}_{ik}]_j$ as the j th columns of the $\hat{\mathbf{Q}}_{ik}$ and \mathbf{Q}_{ik} matrices respectively, then

$$\text{cov} \left([\hat{\mathbf{Q}}_{ik}]_j \right) = \frac{1}{N_Q} [\mathbf{Q}_{ik}]_j [\mathbf{Q}_{ik}]_j^T. \quad (8)$$

Remark 8: If \mathbf{g}_{ilk} is an i.i.d. complex Gaussian vector, then $[\mathbf{Q}_{ik}]_j = [\zeta_{ik} \ 0]^T$, and therefore, $\text{cov}([\hat{\mathbf{Q}}_{ik}]_j) = \frac{1}{N_Q} \zeta_{ik}^2 \mathbf{I}_M$.

The errors in all the M^2 elements of $\hat{\mathbf{Q}}_{ik}$ harm its eigenstructure, making its eigenvalues and eigenvectors unaligned with those of \mathbf{Q}_{ik} [5]. Hence, it has a great impact on the system performance, as the MMSE channel estimator takes advantage of the eigenstructure of \mathbf{Q}_{ik} to acquire better channel estimates. Therefore, in order to overcome such issues, we estimate the covariance matrix as the following convex combination scheme, as suggested in [5],

$$\hat{\mathbf{Q}}_{ik}(\eta) = \eta \hat{\mathbf{Q}}_{ik} + (1 - \eta) \hat{\mathbf{Q}}_{ik}^{\text{diag}}, \quad (9)$$

between (6) and its diagonalized version, $\hat{\mathbf{Q}}_{ik}^{\text{diag}}$. This kind of regularization turns $\hat{\mathbf{Q}}_{ik}$ into a full-rank matrix for any value of $\eta < 1$, even for the case where $N_Q < M$, and it underestimates the values of the unreliable off-diagonal elements.

In this work, we focus on the performance assessment when we estimate \mathbf{Q}_{ik} and do not consider the estimation of the individual \mathbf{R}_{ilk} , $\forall i, l, k$. In [4] the authors propose a specific training phase for estimating \mathbf{R}_{ilk} . By replacing (6) into (4) and treating $\hat{\mathbf{Q}}_{ik}(\eta)$ as the true covariance matrix, we can then approximate the MMSE estimate of \mathbf{g}_{ilk} as

$$\hat{\mathbf{g}}_{ilk}^{\text{approx. MMSE}} = \frac{\mathbf{R}_{ilk}}{\hat{\mathbf{Q}}_{ik}(\eta)} \mathbf{z}_{ik} = \mathbf{R}_{ilk} \hat{\mathbf{Q}}_{ik}^{-1}(\eta) \mathbf{z}_{ik}. \quad (10)$$

Assuming that \mathbf{z}_{ik} is independent of $\hat{\mathbf{Q}}_{ik}(\eta)$, *i.e.*, \mathbf{z}_{ik} is not used to estimate the covariance matrices, $\hat{\mathbf{Q}}_{ik}(\eta)$, and that N_Q is large enough to produce good estimates of \mathbf{Q}_{ik} , then the MSE per antenna of this estimator is given by

$$\begin{aligned} \eta_{ik}^{\text{approx. MMSE}} &= \frac{1}{M} \mathbb{E} \{ \|\hat{\mathbf{g}}_{ilk}^{\text{approx. MMSE}} - \mathbf{g}_{ilk}\|^2 \} \\ &= \frac{1}{M} \text{Tr} \left[\mathbf{R}_{ilk} - \mathbf{R}_{ilk} \hat{\mathbf{Q}}_{ik}^{-1}(\eta) \mathbf{R}_{ilk} \right], \end{aligned} \quad (11)$$

which tends to that of the actual MMSE estimator as $N_Q \rightarrow \infty$. The regularization factor η can be selected so that the MSE per antenna is minimized.

After all the previous derived MSE equations, we can define the following Lemma.

Lemma 1: Considering the channel estimator $\hat{\mathbf{g}}_{ilk} = \mathbf{A}_{ik} \mathbf{z}_{ik}$, then the MSE per antenna of the estimation is given by

$$\begin{aligned} \frac{1}{M} \mathbb{E} \{ \|\mathbf{g}_{ilk} - \mathbf{A}_{ik} \mathbf{z}_{ik}\|^2 \} &= \frac{1}{M} \text{Tr} \left[(\mathbf{I}_M - \mathbf{A}_{ik} - \mathbf{A}_{ik}^H) \mathbf{R}_{ilk} \right] \\ &\quad + \frac{1}{M} \text{Tr} \left[\mathbf{A}_{ik} \mathbf{Q}_{ik} \mathbf{A}_{ik}^H \right], \end{aligned} \quad (12)$$

where \mathbf{A}_{ik} is a deterministic matrix defined as

$$\mathbf{A}_{ik} = \begin{cases} \mathbf{I}_M, & \text{LS estimator} \\ \mathbf{R}_{ilk} \mathbf{Q}_{ik}^{-1}, & \text{MMSE estimator} \\ \mathbf{R}_{ilk} \hat{\mathbf{Q}}_{ik}^{-1}(\eta), & \text{Approx. MMSE estimator} \end{cases} \quad (13)$$

Proof: The proof of (12) is obtained through the direct calculation of the MSE as calculated earlier.

Simulation Results: For our simulations, an adequate correlation model is the one in which the channels are spatially correlated, and all eigenvalues of the correlation matrix are non-zero. Thus, we adopt the Exponential correlation model described in [3] with correlation factor, $r = 0.5$, and large-scale fading variations along the array with $\sigma = 4$. We assume the same challenging symmetric setup of Fig. 3 in [3] with $K = 2$

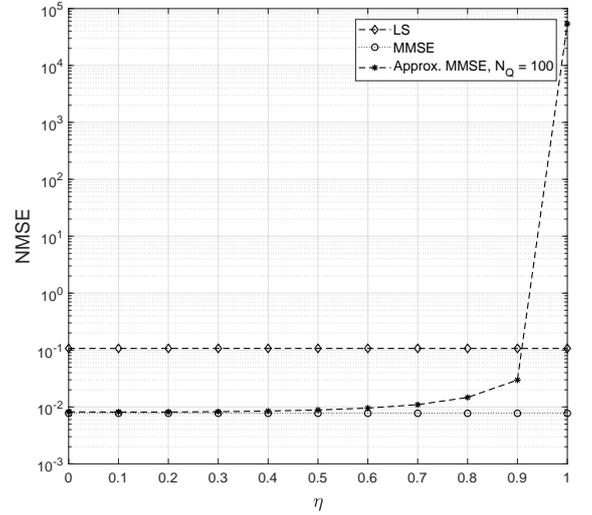


Fig. 1. Normalized MSE per antenna versus the variation of η .

UEs per cell, $L = 4$ cells, coherence block, τ_c , of 200 channel uses, and that each device transmits with a power of 100 mW. For all results, except the one in Fig. 3, we consider $M = 100$ antennas. Except for the results shown in Fig. 1, $\eta = 0.5$. As described in [3], the pilot contamination is very high in that setup. All figures presented next plot the normalized MSE (NMSE) per antenna, $\mathbb{E} \{ \|\mathbf{g}_{ilk} - \mathbf{A}_{ik} \mathbf{z}_{ik}\|^2 \} / M \text{Tr} \left[\mathbf{R}_{ilk} \right]$.

Fig. 1 depicts the NMSE per antenna versus the regularization factor, η . As expected, the LS estimator results in a constant and the highest MSE value. Although exhibiting the highest MSE value, it is important to highlight that the LS estimator does not account for the large-scale fading coefficients knowledge. As the same in LS estimator, the MMSE estimator also presents constant MSE along all considered η values. However, differently from the LS estimator, it has the lowest MSE among the studied estimators. Both estimators, LS and MMSE, present constant MSE as they do not depend on η . On the other hand, the approximated MMSE estimator performance depends on the considered η value. From η values ranging from 0 up to 0.4, the approximated MMSE estimator has MSE values quite similar to the MMSE estimator. The resulting performance is due to the correlation elements between the channels (*i.e.*, off-diagonal elements) that has smaller weights when compared to the diagonal elements. In η ranging from 0.5 up to 0.9, the performance is worse than that of the MMSE estimator but, on the other hand, is still better than the LS. For η values greater than 0.9, the MSE of the approximated MMSE estimator presents inferior performance, becoming worse than the other estimators. This is due to the fact that the covariance matrix is not a full-rank matrix anymore. Another important conclusion

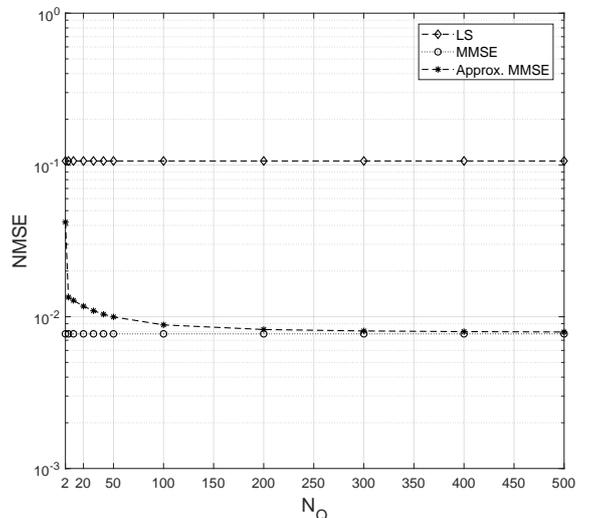


Fig. 2. Normalized MSE per antenna versus the variation of N_Q .

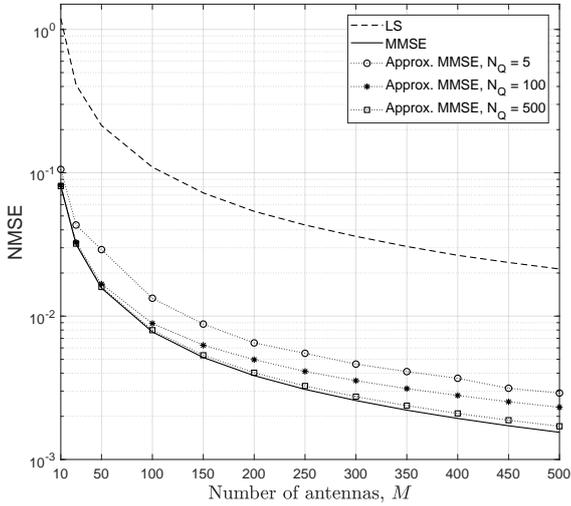


Fig. 3. Normalized MSE per antenna versus the number of antennas, M .

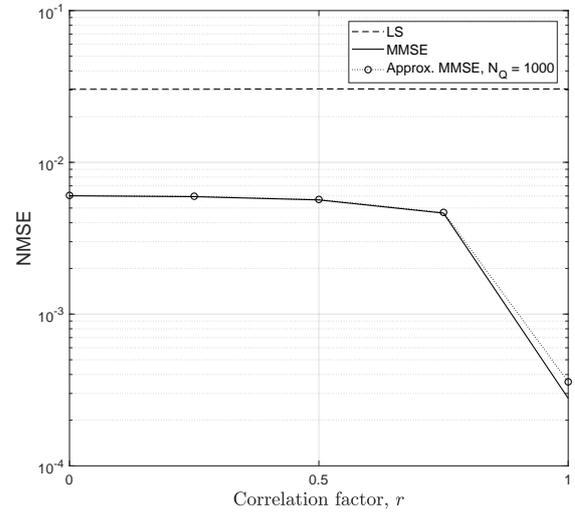


Fig. 5. Normalized MSE per antenna versus the correlation factor.

that can be drawn from the figure is that for the correlation model adopted here, the estimation of only the diagonal elements of the covariance matrix would suffice to have a good channel estimator.

Fig. 2 depicts the performance of the channel estimators in terms of the NMSE per antenna versus the variation of the number of channel observations, N_Q . As expected, both the MMSE and the approximated MMSE estimators have better performance than the LS for all considered N_Q values. A critical remark is that LS estimator does not need any prior channel information, and the MMSE estimator assumes perfect knowledge of the channel statistics. As a result, the MSE is constant over all N_Q values for the LS and MMSE. On the other hand, the approximated MMSE estimator depends on the number of observations used to estimate the covariance matrix and, consequently, presents performance dependable on N_Q . As can be noticed, the performance of the approximated MMSE asymptotically tends to that of the MMSE estimator as N_Q increases.

Fig. 3 presents the performance of the channel estimators in terms of the NMSE per antenna versus the number of antennas. As can be seen, the MMSE and the approximated MMSE estimators have better performance than the LS estimator for all considered number of antennas. The NMSE of all three estimators decreases as the number of antennas also increases. It is also essential to highlight the fact that, as expected, while N_Q increases, the approximated MMSE performance tends to that of the MMSE estimator.

Fig. 4 presents the performance of the channel estimators in terms of the NMSE per antenna versus the uplink transmit power, p . Again,

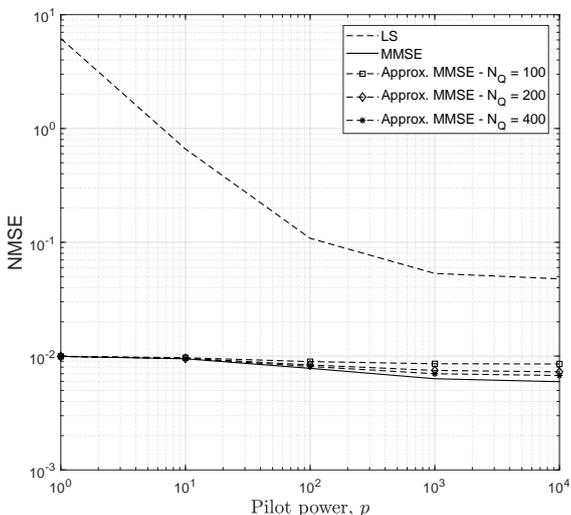


Fig. 4. Normalized MSE per antenna versus the variation of the pilot power.

the LS estimator is the worst one among the studied estimators. The result shows that the MMSE estimator is the best one for all values of p . The performance of the approximated MMSE becomes better as N_Q increases, providing as consistent estimates of the covariance matrices as the MMSE estimator. While p increases, the MMSE performs better, although the same behavior cannot be observed with the approximated MMSE estimator if N_Q is insufficient.

Figure 5 depicts the NMSE for the LS, MMSE and the approximated MMSE estimators as a function of the correlation factor. The covariance matrix is estimated averaging $N_Q = 1000$ channel observations. For the LS estimator, the NMSE is constant for all range of correlation factor. For the MMSE and approximated MMSE, the NMSE decreases as the correlation factor, r , increases. Note that since N_Q is large, it produces a good estimation and consequently the approximated MMSE performs closely to the MMSE estimator. However, when the correlation between the channels becomes higher, the MMSE estimator is better than the approximated MMSE since the MMSE has the knowledge of the channel statistics. Moreover, if we analyse (5), it is possible to conclude that both terms become similar as the correlation factor increases, which consequently causes the NMSE to decrease. The same analysis is applicable to (11).

Conclusion: We derived analytical MSE expressions for perfect and imperfect channel covariance matrices. These expressions are useful in the design and performance evaluation of systems. We have also discussed and shown how a BS can estimate the channel covariance matrix necessary for the approximated MMSE estimation.

Felipe A. P. de Figueiredo (*Instituto Nacional de Telecomunicações- INATEL, Brazil*), Dimas A. M. Lemes (*UNICAMP, Brazil*), Cláudio F. Dias (*UNICAMP, Brazil*) and Gustavo Fraidenraich (*UNICAMP, Brazil*)
E-mail: felipe.pereira@ugent.be

References

- Xiang Gao et. al., *Massive MIMO Performance Evaluation Based on Measured Propagation Data*, IEEE Trans. Wireless Commun., vol. 14, no. 7, pp. 3899-3911, July 2015.
- Felipe A. P. de Figueiredo et. al., *Channel estimation for massive MIMO TDD systems assuming pilot contamination and flat fading*, EURASIP Journal on Wireless Communications and Networking, January 2018.
- Emil Bjornson, et al., *Massive MIMO has Unlimited Capacity*, IEEE Trans. Wireless Commun., vol. 17, no. 1, pp. 574-590, January 2018.
- H. Yin, et al., *A coordinated approach to channel estimation in large-scale multiple-antenna systems*, IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 264-273, February 2013.
- Nafiseh Shariati, et al., *Low-Complexity Polynomial Channel Estimation in Large-Scale MIMO With Arbitrary Statistics*, IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, October 2014.