Low-Complexity Channel Estimation in Large-Scale MIMO using Polynomial Expansion

Nafiseh Shariati, Emil Björnson, Mats Bengtsson, Mérouane Debbah

Low-Complexity Channel Estimation in Large-Scale MIMO using Polynomial Expansion

Nafiseh Shariati*, Emil Björnson*‡, Mats Bengtsson*, and Mérouane Debbah‡
*Signal Processing Lab, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden
‡Alcatel-Lucent Chair on Flexible Radio, SUPELEC, Gif-sur-Yvette, France
E-mail: {nafiseh, mats.bengtsson}ee.kth.se, {emil.bjornson, merouane.debbah}@supelec.fr

***Abstract*—This paper considers pilot-based channel estimation in large-scale multiple-input multiple-output (MIMO) communication systems, also known as "massive MIMO". Unlike previous works on this topic, which mainly considered the impact of inter-cell disturbance due to pilot reuse (so-called *pilot contamination*), we are concerned with the computational complexity. The conventional *minimum mean square error (MMSE)* and *minimum variance unbiased (MVU)* channel estimators rely on inverting covariance matrices, which has cubic complexity in the multiplication of number of antennas at each side. Since this is extremely expensive when there are hundreds of antennas, we propose to approximate the inversion by an L-order matrix polynomial. A set of low-complexity Bayesian channel estimators, coined Polynomial ExpAnsion CHannel (PEACH) estimators, are introduced. The coefficients of the polynomials are optimized to yield small mean square error (MSE). We show numerically that near-optimal performance is achieved with low polynomial orders. In practice, the order L can be selected to balance between complexity and MSE. Interestingly, pilot contamination is beneficial to the PEACH estimators in the sense that smaller L can be used to achieve near-optimal MSEs.**

I. Introduction

MIMO techniques can bring substantial improvements in spectral efficiency to wireless systems, by increasing the spatial reuse. While 8×8 MIMO transmission has found its way into standards such as LTE-Advanced [1], there is an increasing interest in equipping base stations with much larger antenna arrays [2]–[4]. Such large-scale MIMO, or "massive MIMO", techniques can give unprecedented spatial resolution, enabling a very dense spatial reuse that potentially can keep up with the rapidly increasing demand for wireless connectivity.

A major limiting factor in large-scale MIMO is the availability of accurate channel state information (CSI). This is since high spatial resolution can only be exploited if the propagation environment is precisely known. CSI is typically acquired by sending pilot signals and estimating the channel coefficients from the received signals [5]–[10]. The Bayesian MMSE estimator can be applied [7]–[10] if the channel statistics are known, while the MVU estimator is applied otherwise [7].

These channel estimators basically solve a linear system of equations, or equivalently multiply the received signal with an inverse of the covariance matrices, which is a mathematical operation with cubic computational complexity. Therefore, it is very computationally expensive to compute the MMSE and MVU estimates in large-scale MIMO systems. The high complexity can be avoided under ideal propagation conditions where all covariance matrices are diagonal, but large-scale MIMO channels typically have a distinct spatial correlation due to insufficient antenna spacing and richness of the propagation environment [4]. Moreover, the necessary pilot reuse in cellular networks creates spatially correlated inter-cell interference, known as *pilot contamination*, that reduces the estimation performance and the spectral efficiency [2]–[6].

Similar complexity issues appear in multiuser detection, where both the decorrelating detector and the linear MMSE detector involve matrix inversions [11]. A common low-complexity approach is reduced-rank filtering [12]. This can be achieved by *polynomial expansion (PE)*, where the matrix inverse is approximated by an L-order matrix polynomial [11]–[14]. PE-based detectors are versatile since the structure enables simple multistage hardware implementation [11] and the order L needs not to scale with the system size to achieve near-optimal performance [12]. Therefore, L is simply selected to balance between complexity and detection performance. A main problem is to select the coefficients of the polynomial to achieve high performance at small L; the optimal weights are expensive to compute [11], but alternatives based on appropriate scalings [13] and asymptotic analysis [14] exist.

Inspired by the prior works in detection, in this paper, we propose a set of low-complexity channel estimators that we call *Polynomial ExpAnsion CHannel (PEACH)* estimators. These novel estimators approximate the MMSE estimator by replacing the matrix inversion with a polynomial expansion. The coefficients of the polynomial are optimized to yield minimal MSE at any fixed polynomial order L, while keeping the low complexity. The PEACH estimators are evaluated under different propagation/interference conditions and show remarkably good performance at low polynomial orders.

Notation: Boldface (lower case) is used for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. Let $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^{-1}$ denote the transpose, the conjugate transpose, and inverse of $\mathbf{X}$, respectively. The Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$ is denoted $\mathbf{X} \otimes \mathbf{Y}$, $\mathrm{vec}(\mathbf{X})$ is the vector obtained by stacking the columns of $\mathbf{X}$, $\mathrm{tr}(\mathbf{X})$ denotes the trace, and $\|\mathbf{X}\|_F$ is the Frobenius norm. The notation $\triangleq$ denote definitions, while $\mathcal{O}(M^x)$ means that the complexity is bounded by CM^x for some $C > 0$.

E. Björnson is funded by the International Postdoc Grant 2012-228 from the Swedish Research Council. This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering).

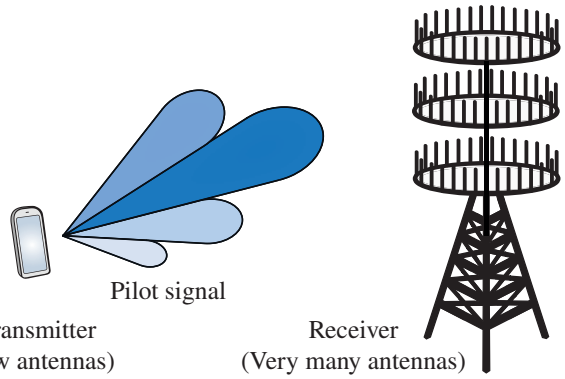

Fig. 1. Illustration of pilot signaling in a large-scale $N_t \times N_r$ MIMO system, where $N_r \gg N_t$. The complexity of conventional channel estimators is very large in these systems, which calls for low-complexity alternatives.

II. Problem Formulation

This paper revisits the problem of estimating a quasi-static flat-fading MIMO channel $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, where N_t is the number of transmit antennas and N_r is the number of receive antennas. Similar to [8]–[10], the transmitter sends a fixed predefined pilot matrix $\mathbf{P} \in \mathbb{C}^{N_t \times B}$ over $B \geq 1$ symbol times; see Fig. 1. The receiver tries to obtain $\mathbf{H}$ from the received signal $\mathbf{Y} \in \mathbb{C}^{N_r \times B}$, given by

$$\mathbf{Y} = \mathbf{H}\mathbf{P} + \mathbf{N} \tag{1}$$

where the disturbance $\mathbf{N} \in \mathbb{C}^{N_r \times B}$ is assumed to be circularly-symmetric complex Gaussian noise: $\mathrm{vec}(\mathbf{N}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{S})$. The disturbance covariance matrix $\mathbf{S} \in \mathbb{C}^{N_r B \times N_r B}$ is positive definite and can include both regular uncorrelated receiver noise and different types of interference from other systems. The analysis herein holds for any $\mathbf{S}$, but some typical special cases will be considered in Section IV-B.

The channel matrix is modeled as Rayleigh fading with $\mathrm{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, where the channel covariance matrix $\mathbf{R} \in \mathbb{C}^{N_t N_r \times N_t N_r}$ is positive semi-definite. Observe that $\mathbf{R}$ is generally *not* a scaled identity matrix, but describes the spatial propagation environment. The matrices $\mathbf{R}$ and $\mathbf{S}$ are assumed known at the transmitter and receiver.

If the channel statistics are known at the receiver, the Bayesian MMSE estimator of the MIMO channel is [7]–[10]

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}}) = \mathbf{R}\widetilde{\mathbf{P}}^H \left(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}\right)^{-1} \mathrm{vec}(\mathbf{Y}) \tag{2}$$

where $\widetilde{\mathbf{P}} \triangleq (\mathbf{P}^T \otimes \mathbf{I})$. The performance is measured in MSE $= \mathbb{E}\{\|\mathbf{H} - \widehat{\mathbf{H}}_{\mathrm{MMSE}}\|_F^2\} = \mathrm{tr}\left((\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1}\right)$.

Alternatively, if the channel distribution is unknown to the receiver, the classic MVU estimator is [7, Chapter 4]

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MVU}}) = \left(\widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}}\right)^{-1} \widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\mathrm{vec}(\mathbf{Y}). \tag{3}$$

The performance measure is then the estimation variance $\mathbb{E}\{\|\mathbf{H} - \widehat{\mathbf{H}}_{\mathrm{MVU}}\|_F^2\} = \mathrm{tr}\left((\widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1}\right)$. Obviously,

$$\mathrm{tr}\left((\mathbf{R}^{-1} + \widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1}\right) < \mathrm{tr}\left((\widetilde{\mathbf{P}}^H\mathbf{S}^{-1}\widetilde{\mathbf{P}})^{-1}\right) \tag{4}$$

for any $\mathbf{R} \neq \mathbf{0}$, thus the MMSE estimator achieves a better average estimation performance than the MVU estimator since it utilizes the channel statistics.

A. Complexity Issues in Large-Scale MIMO Systems

The main computational complexity when computing the MMSE and MVU estimators in (2) and (3) lies in solving a linear system of equations or, equivalently, in computing the matrix inversions directly.[1] Both approaches have computational complexities that scale as $\mathcal{O}(M^3)$ where $M \triangleq N_t N_r$.[2] This complexity is relatively modest in classic MIMO systems where 2×2, 4×4, or 8×8 are typical configurations.

Recently, there is an increasing interest in large-scale MIMO systems where there might be hundreds of antennas at one side of the link [2]–[4]. To excite all channel dimensions, the pilot length B should be of the same order as N_t. Large-scale systems are therefore envisioned to exploit channel reciprocity to always have $N_t < N_r$ in the channel estimation phase—N_r can even be orders of magnitude larger than N_t without degrading the estimation performance *per* antenna element.

Observe that in a potential future large-scale MIMO system with $N_r = 200$ and $N_t = 20$, the MMSE and MVU estimators would require inverting matrices of size 4000×4000 (or similarly, solving a linear system of equations with 4000 unknown variables). This massive operation needs to be redone quite often since $\mathbf{R}$ and $\mathbf{S}$ change with time. The purpose of this paper is to develop alternative channel estimators that allow for balancing between computational/hardware complexity and estimation performance.

Remark 1. *While having Gaussian channels and disturbance is a well-accepted assumption in conventional MIMO systems, the channel modeling for large-scale MIMO is still in its infancy. By increasing the number of antennas we will improve the spatial resolution of the array which eventually may invalidate the rich-scattering assumption that is behind the use of Gaussian channel distributions [4]. However, we stress that the estimators in this paper can be applied and give reasonable performance under any channel and disturbance distributions; this is since* (2) *is also the linear MMSE estimator and* (3) *is the best linear unbiased estimator (BLUE) in cases when only the first two moments of* $\mathbf{H}$ *and/or* $\mathbf{N}$ *are known [7], [10].*

III. Low-Complexity Bayesian PEACH Estimators

In this section, we propose several low-complexity Bayesian channel estimators based on the concept of polynomial expansion. To understand the main idea, we first state the following lemma which is easily proved by using standard Taylor series.

[1] If $\mathbf{R}, \mathbf{S}, \mathbf{P}$ are all diagonal matrices, the complexity can be greatly reduced by simply inverting the diagonal elements. However, such special cases are of limited practical interest, particularly in large-scale MIMO systems which are prone to non-negligible spatial channel correlation and pilot contamination.

[2] Note that $\mathcal{O}(M^3)$ refers to the complexity scaling of the classical Gaussian elimination algorithm. The complexity is reduced to $\mathcal{O}(M^{2.8074})$ by the more sophisticated Strassen's algorithm. The exponent can be further reduced, see e.g. [15], but mainly for academic purposes since extremely large matrices are required to actually benefit from such improved asymptotic behaviors.

Lemma 1. *For any Hermitian matrix $\mathbf{X} \in \mathbb{C}^{N\times N}$, with bounded eigenvalues $|\lambda_n(\mathbf{X})| < 1$ for all n, it holds that*

$$(\mathbf{I}-\mathbf{X})^{-1} = \sum_{l=0}^{\infty} \mathbf{X}^l. \tag{5}$$

Observe that the impact of $\mathbf{X}^l$ in (5) reduces with l (as $\lambda_n(\mathbf{X})^l$ for each eigenvalue). It therefore makes sense to consider L-order polynomial expansions of the matrix inverse using only the terms $l = 0, \ldots, L$. In principle, the inverse of each eigenvalue is then approximated by an L-order Taylor polynomial, thus L needs *not* scale with the matrix dimension to achieve a certain accuracy per element. Instead, L can be selected to balance low approximation error with complexity.

In order to apply Lemma 1 on matrices with any eigenvalue structure, we obtain the next result which is similar to [13].

Proposition 1. *For any positive-definite Hermitian matrix $\mathbf{X}$,*

$$\mathbf{X}^{-1} = \alpha\big(\mathbf{I} - (\mathbf{I}-\alpha\mathbf{X})\big)^{-1} \approx \alpha \sum_{l=0}^{L} (\mathbf{I}-\alpha\mathbf{X})^l \tag{6}$$

where the approximation holds with equality when $L \to \infty$ if α is selected to satisfy $0 < \alpha < \frac{2}{\max_n \lambda_n(\mathbf{X})}$.

A. Unweighted PEACH Estimator

Applying the approximation in Proposition 1 on the MMSE estimator in (2) gives the low-complexity L-order *Polynomial ExpAnsion CHannel (PEACH)* estimator

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{PEACH}}) = \mathbf{R}\widetilde{\mathbf{P}}^H \sum_{l=0}^{L} \alpha\big(\mathbf{I} - \alpha(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})\big)^l \mathrm{vec}(\mathbf{Y}) \tag{7}$$

which does not involve any channel inversions. The computational complexity of (7) is $\mathcal{O}(LM^2)$ where $M \triangleq N_t N_r$. Whenever $L \ll M$, this is a large complexity reduction as compared to $\mathcal{O}(M^3)$ for the original MMSE estimator.

Theorem 1. *The PEACH estimator in* (7) *achieves the MSE*

$$\mathrm{tr}\Big(\mathbf{R} + \mathbf{R}\widetilde{\mathbf{P}}^H \mathbf{A}_L (\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})\mathbf{A}_L^H \widetilde{\mathbf{P}}\mathbf{R} - 2\mathbf{R}\widetilde{\mathbf{P}}^H \mathbf{A}_L \widetilde{\mathbf{P}}\mathbf{R}\Big) \tag{8}$$

where $\mathbf{A}_L = \sum_{l=0}^{L} \alpha\big(\mathbf{I} - \alpha(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})\big)^l$.

Proof: Follows from direct computation of the MSE. ■

The PEACH estimator lends itself to efficient multistage hardware implementation where $(\cdot)^{l+1}\mathrm{vec}(\mathbf{Y})$ is computed from $(\cdot)^l\mathrm{vec}(\mathbf{Y})$ by multiplying the latter vector with $\big(\mathbf{I} - \alpha(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})\big)$. The hardware can be similar to the multistage detection implementation illustrated in [11, Fig. 1].

It remains to select the scaling parameter α to satisfy the condition in Proposition 1. From a complexity point of view, we easily can select α to be equal to $\frac{2}{\mathrm{tr}(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})}$. On the other hand, we are also interested in choosing α to achieve fast convergence in the polynomial expansion. Among the values that satisfy the condition in Proposition 1, the choice

$$\alpha_{\mathrm{PEACH}} = \frac{2}{\max_n \lambda_n(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}) + \min_n \lambda_n(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})} \tag{9}$$

will maximize the asymptotic convergence since the largest and smallest eigenvalue of $\big(\mathbf{I} - \alpha(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})\big)$ become symmetric around the origin [13]. Although the computation of the extreme eigenvalues is generally quite expensive, we note that these can be deduced from the setup; that is, pathloss, antenna array design, location of dominating interferers, receiver noise, etc. We therefore assume that α_{PEACH} can be computed with low complexity in this paper and refer to [13] for general techniques for efficient computation of eigenvalues.

B. Weighted PEACH Estimator

Although the PEACH estimator (7) converges to the MMSE estimator as $L \to \infty$, it is generally not the best L-order polynomial estimator at finite L. More specifically, instead of multiplying each term in the sum with α, we can assign different weights and optimize these for the specific order L. In this way, we obtain the *weighted PEACH estimator*

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\text{W-PEACH}}) = \mathbf{R}\widetilde{\mathbf{P}}^H \sum_{l=0}^{L} w_l \alpha^{l+1} \big(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}\big)^l \mathrm{vec}(\mathbf{Y}) \tag{10}$$

where $\mathbf{w} = [w_0, \ldots, w_L]^T$ are scalar weighting coefficients.[3] Observe that the α-parameter is now redundant and can be set to one. For numerical reasons, it might still be good to select

$$\alpha_{\text{W-PEACH}} \leq \frac{1}{\max_n \lambda_n(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})} \tag{11}$$

since this prevent the eigenvalues of $\alpha^{l+1}\big(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}\big)^l$ to grow unboundedly as l becomes large. This will simplify the implementation of the following theorem, which finds the weights that minimize the MSE.

Theorem 2. *The MSE $\mathbb{E}\{\|\mathbf{H} - \widehat{\mathbf{H}}_{\text{W-PEACH}}\|_F^2\}$ is minimized by*

$$\mathbf{w}_{\mathrm{opt}} = [w_0^{\mathrm{opt}} \ \ldots \ w_L^{\mathrm{opt}}]^T = \mathbf{A}^{-1}\mathbf{b} \tag{12}$$

where the ijth element of $\mathbf{A} \in \mathbb{C}^{L+1\times L+1}$ and the ith element of $\mathbf{b} \in \mathbb{C}^{L+1}$ are

$$\begin{aligned} [\mathbf{A}]_{ij} &= \alpha^{i+j}\mathrm{tr}\Big(\mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i+j-1}\widetilde{\mathbf{P}}\mathbf{R}\Big), \\ [\mathbf{b}]_i &= \alpha^i \mathrm{tr}\Big(\mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i-1}\widetilde{\mathbf{P}}\mathbf{R}\Big). \end{aligned} \tag{13}$$

The resulting MSE of the W-PEACH estimator is

$$\mathrm{tr}(\mathbf{R}) + \mathbf{w}_{\mathrm{opt}}^H \mathbf{A}\mathbf{w}_{\mathrm{opt}} - \mathbf{b}^H\mathbf{w}_{\mathrm{opt}} - \mathbf{w}_{\mathrm{opt}}^H\mathbf{b}. \tag{14}$$

Proof: Follows from differentiation of the MSE. ■

Although Theorem 2 provides the optimal weights, the computational complexity is $\mathcal{O}(M^3)$ since it involves pure matrix multiplications of the form $\mathbf{Z}^i$. This means that computing the optimal weights for the W-PEACH estimator has the same asymptotic complexity scaling as computing the original MMSE estimator. To benefit from the weighting we thus need to find a low-complexity approach to compute the weights.

[3] W-PEACH is obtained by expanding each $(\mathbf{I} - \alpha(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}))^l$ as a binomial series, collecting terms, and replacing constant factors with weights.

C. Low-Complexity Weights

Next, we propose a low-complexity algorithm to compute weights for the W-PEACH estimator. We will exploit that

$$(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S}) = \mathbb{E}\{\mathrm{vec}(\mathbf{Y})\mathrm{vec}(\mathbf{Y})^H\} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t\mathbf{y}_t^H, \tag{15}$$

where $\mathbf{y}_t = \mathrm{vec}(\mathbf{Y})$ denotes the received signal at estimation time instant t. This means that $(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})$ is closely approximated by the sample covariance matrix $\frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t^H$ if the number of samples T is large. Although one generally needs $T \gg BN_r$ to get a good approximation, we can get away with much smaller T since we will only compute traces.

For any fixed $T \geq 1$ and $i \geq 1$, we now observe that

$$\mathrm{tr}\left(\mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^i\widetilde{\mathbf{P}}\mathbf{R}\right) \tag{16}$$

$$\approx \mathrm{tr}\left(\mathbf{R}\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t^H\right)\widetilde{\mathbf{P}}\mathbf{R}\right) \tag{17}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t^H\left(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i-1}\right)\mathbf{y}_t. \tag{18}$$

Since the elements of $\mathbf{A}$ and $\mathbf{b}$ in (13) are of the form in (16), we can approximate each element using (18).[4] By computing/updating these approximations over a sliding time window of length T, we obtain Algorithm 1. At any time instant t, this algorithm computes approximations of $\mathbf{A}, \mathbf{b}$, denoted by $\widetilde{\mathbf{A}}_t, \tilde{\mathbf{b}}_t$, by using the received signals $\mathbf{y}_t, \ldots, \mathbf{y}_{t-T+1}$. These are used to compute approximate weights $\mathbf{w}_{\mathrm{approx},t}$. To reduce the amount of computations, $\widetilde{\mathbf{A}}_t, \tilde{\mathbf{b}}_t$ are obtained from $\widetilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{b}}_{t-1}$ by adding one term per element based on the current received signal $\mathbf{y}_t$ and removing the impact of the old received signal $\mathbf{y}_{t-T}$ (which is now outside the time window). The algorithm can be initialized in any way; for example, by accumulating T received signals to fill the time window.

The complexity of computing the elements in $\widetilde{\mathbf{A}}_t, \tilde{\mathbf{b}}_t$ is $\mathcal{O}(L^2M^2)$ *per time instant* since there are $(L+1)(L+1)$ elements in $\widetilde{\mathbf{A}}_t$ and $L+1$ elements in $\tilde{\mathbf{b}}_t$ to be computed, which results in $(L+2)(L+1)$ elements in total, and we need to compute a series of multiplications between vectors and matrices. Furthermore, $\mathbf{w}_{\mathrm{approx},t}$ is obtained by solving an L-dimensional system of equations, which has complexity $\mathcal{O}(L^3)$. To summarize, the W-PEACH estimator along with Algorithm 1 has a computational complexity of $\mathcal{O}(L^2M^2 + LM^2 + L^3) = \mathcal{O}(L^2M^2 + L^3)$, which is smaller than $\mathcal{O}(M^3)$ of the MMSE estimator for $L < \sqrt{M} = \sqrt{N_tN_r}$.

One additional feature of Algorithm 1 is that it can easily be extended to practical scenarios where also the true covariance matrices $\mathbf{R}$ and $(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})$ are approximated by sample covariance matrices. This would enable adaptive tracking of the slow variations in channel and disturbance statistics that appear in practice. We leave this extension for future work.

[4]Note that $b_0 = \mathrm{tr}(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H)$ needs to be treated differently since there is no $(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})$ term. For some predefined vectors $\mathbf{v}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, Algorithm 1 uses the approximation $\mathrm{tr}(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H) \approx \frac{\alpha}{T}\sum_{i=1}^{T}\mathbf{v}_i^H\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H\mathbf{v}_i$.

Algorithm 1: Low-complexity weights for W-PEACH

Input: Polynomial order L and time window T;
Input: Current time t;
Input: New and old received signals $\mathbf{y}_t, \mathbf{y}_{t-T}$;
Input: Approximations $\widetilde{\mathbf{A}}_{t-1}, \tilde{\mathbf{b}}_{t-1}$ at previous time $t-1$;

Set $[\widetilde{\mathbf{A}}_t]_{ij} = [\widetilde{\mathbf{A}}_{t-1}]_{ij}$

$$+ \frac{\alpha^{i+j}}{T}\mathbf{y}_t^H\left(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i+j-2}\right)\mathbf{y}_t$$
$$- \frac{\alpha^{i+j}}{T}\mathbf{y}_{t-T}^H\left(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i+j-2}\right)\mathbf{y}_{t-T} \quad \forall i,j$$

Set $[\tilde{\mathbf{b}}_t]_i = [\tilde{\mathbf{b}}_{t-1}]_i$

$$+ \frac{\alpha^{i}}{T}\mathbf{y}_t^H\left(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i-2}\right)\mathbf{y}_t$$
$$- \frac{\alpha^{i}}{T}\mathbf{y}_{t-T}^H\left(\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H(\widetilde{\mathbf{P}}\mathbf{R}\widetilde{\mathbf{P}}^H + \mathbf{S})^{i-2}\right)\mathbf{y}_{t-T} \quad \forall i \geq 2$$

Set $[\tilde{\mathbf{b}}_t]_1 = \frac{\alpha}{T}\sum_{i=1}^{T}\mathbf{v}_i^H\widetilde{\mathbf{P}}\mathbf{R}^2\widetilde{\mathbf{P}}^H\mathbf{v}_i$ for $\mathbf{v}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$;
Compute $\mathbf{w}_{\mathrm{approx},t} = \widetilde{\mathbf{A}}_t^{-1}\tilde{\mathbf{b}}_t$;
Output: Approximate weights $\mathbf{w}_{\mathrm{approx},t}$ at time t;

D. Summary of Computational Complexity

The complexity of the conventional estimators and proposed PEACH estimators are summarized as follows:

Channel Estimators	Computational Complexity
MMSE and MVU	$\mathcal{O}(N_t^3N_r^3)$
PEACH	$\mathcal{O}(LN_t^2N_r^2)$
W-PEACH	$\mathcal{O}(L^2N_t^2N_r^2 + L^3)$

We note that the cubic complexity scaling in N_tN_r for the conventional MMSE and MVU estimators have been reduced to squared complexity for the proposed PEACH estimators. The order L of the polynomial expansion has a clear impact on the complexity, but recall that it generally needs not scale with N_tN_r [12]. In the next section, we will illustrate that small values on L yields good performance.

IV. Numerical Evaluation

In this section, we illustrate the performance of the proposed PEACH and W-PEACH estimators. The analysis so far has been generic with respect to the disturbance covariance matrix $\mathbf{S}$. Here, we consider two scenarios: noise-limited and cellular networks with pilot contamination. We describe the latter scenario in more detail since it is one of the main challenges in the development of large-scale MIMO systems [4].

A. Noise-Limited Scenario

A commonly studied scenario is when there is only uncorrelated receiver noise; thus $\mathbf{S} = \sigma^2\mathbf{I}$ where σ^2 is the variance.

B. Pilot Contamination Scenario

A scenario that has received much attention in the large-scale MIMO literature is when there is disturbance from

simultaneous reuse of pilot signals in neighboring cells [2]–[6]. Such reuse is often necessary due to the finite channel coherence time (i.e., the time that a channel estimate can be considered accurate), but leads to a special form of interference called pilot contamination. It can be modeled as[5]

$$\mathbf{N} = \sum_{i \in \mathcal{I}} \mathbf{H}_i \mathbf{P} + \widetilde{\mathbf{N}} \tag{19}$$

where $\mathcal{I}$ is the set of interfering cells, $\mathbf{H}_i$ is the channel from the ith interfering cell to the receiver in the cell under study, and $\mathrm{vec}(\widetilde{\mathbf{N}}) \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ is uncorrelated receiver noise. If $\mathbf{H}_i$ is Rayleigh fading with $\mathrm{vec}(\mathbf{H}_i) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_i)$, then

$$\mathbf{S} = \sum_{i \in \mathcal{I}} \widetilde{\mathbf{P}} \mathbf{\Sigma}_i \widetilde{\mathbf{P}}^H + \sigma^2 \mathbf{I}. \tag{20}$$

Note that only the sum covariance matrix $\sum_{i \in \mathcal{I}} \mathbf{\Sigma}_i$ needs to be known when computing the proposed PEACH estimators.

When (20) is substituted into the PEACH and W-PEACH estimator expressions in (7) and (10) we get contaminated terms of the form $\mathbf{R}\widetilde{\mathbf{P}}^H \widetilde{\mathbf{P}} \mathbf{\Sigma}_i \widetilde{\mathbf{P}}^H$. These terms are small if $\mathbf{R}$ and $\mathbf{\Sigma}_i$ have very different span, or if $\mathrm{tr}(\mathbf{\Sigma}_i)$ is weak altogether—this is easily observed if $\widetilde{\mathbf{P}}^H \widetilde{\mathbf{P}}$ is a scaled identity matrix. Similar observations were recently made in the capacity analysis of [3] and when developing a pilot allocation algorithm in [5]. Under certain conditions, the subspaces of the useful channel and pilot contamination can be made orthogonal by coordinated pilot allocation across cells [5] or by exploiting both received pilot and data signals for channel estimation as in [6].

C. Numerical Examples

To evaluate the performance of our proposed estimators, we consider a large-scale MIMO system with $N_r = 100$ and $N_t = 10$ antennas in addition to the pilot length $B = 10$. We follow the Kronecker model [16] to describe correlation among antennas of the desired and disturbance MIMO channels. In this model, the covariance matrix of a MIMO channel is modeled as $\mathbf{R} = \mathbf{R_t} \otimes \mathbf{R_r}$ where $\mathbf{R_t} \in \mathbb{C}^{N_t \times N_t}$ and $\mathbf{R_r} \in \mathbb{C}^{N_r \times N_r}$ are the spatial covariance matrices at the transmitter and receiver sides, respectively. Following the same modeling, we have $\mathbf{\Sigma}_i = \mathbf{\Sigma}_{\mathbf{t}_i} \otimes \mathbf{\Sigma}_{\mathbf{r}_i}$ for $i \in \mathcal{I}$.

To generate covariance matrices, we use the exponential model [17]. Without loss of generality, all the covariance matrices will have diagonal elements equal to one which results in $\mathrm{tr}(\mathbf{R}) = \mathrm{tr}(\mathbf{\Sigma}_i) = N_t N_r$. We assume that there are two dominating interfering cells, $i = 1, 2$, for which the covariance matrices are weakened by the factor $0 \leq \beta_i < 1$, i.e., $\beta_i \mathbf{\Sigma}_i$. This factor represents how severe the pilot contamination part is: $\beta_i = 0$ represents the noise-limited case, while $\beta_i = 1$ represents the case when the useful channel and the ith interfering channel are equally strong. We define the normalized pilot signal-to-noise ratio (SNR) as $\gamma = \frac{\mathcal{P}_T}{\sigma^2}$ where $\mathcal{P}_T = \frac{1}{N_t}\mathrm{tr}(\mathbf{P}\mathbf{P}^H)$ is the average pilot power.

[5]Cell i can use an arbitrary pilot $\mathbf{P}_i$, but only pilots with overlapping span (i.e., $\mathbf{P}_i \mathbf{P}^H \neq \mathbf{0}$) will cause interference. Therefore, the case of a common reused pilot $\mathbf{P}_i = \mathbf{P}$ is of main interest, while extensions are straightforward.

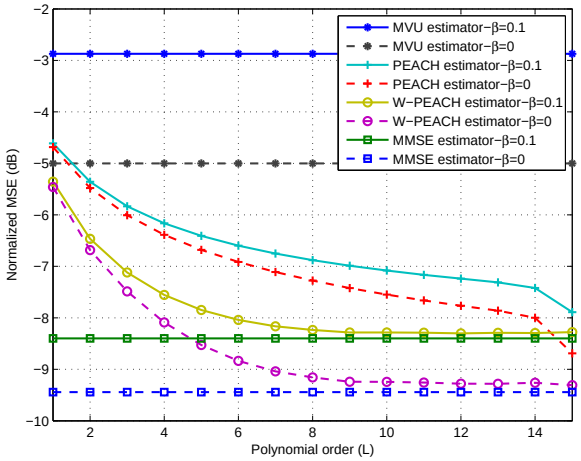

Fig. 2. MSE comparison of different estimators for different polynomial orders when $\beta \in \{0,\ 0.1\}$, which represents noise-limited and pilot contaminated scenarios.

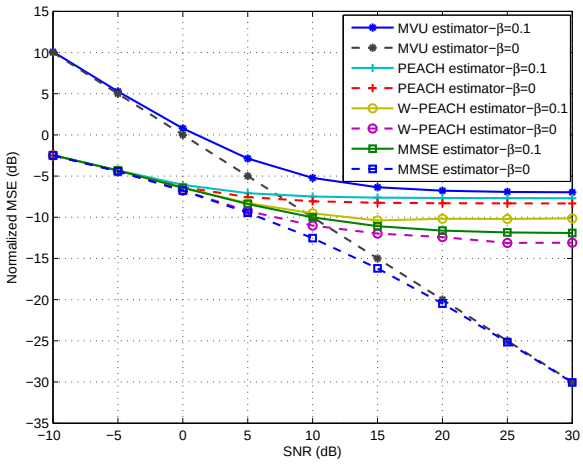

Fig. 3. MSE comparison of different estimators for different values of the SNR γ when $\beta \in \{0,\ 0.1\}$, which represents noise-limited and pilot contaminated scenarios.

We use the normalized MSE, defined as $\frac{\mathrm{MSE}}{\mathrm{tr}(\mathbf{R})}$, as performance measure. In all the figures, we compare the performance of the proposed estimators with the conventional MMSE and MVU estimators. The pilot matrix is $\mathbf{P} = \sqrt{\mathcal{P}_T}\mathbf{I}$.

In Fig. 2, MSE has been plotted as a function of the polynomial order L. The noise-limited scenario is given by $\beta = 0$, while $\beta = 0.1$ (we assume that $\beta_1 = \beta_2 = \beta$) represents the scenario where the two interfering cells have interfering channels which are 10 dB weaker than the desired channels. The SNR is $\gamma = 5$ dB. As can be seen from Fig. 2, both PEACH and W-PEACH detract by increasing L and outperform MVU for $L \geq 2$. Interestingly, W-PEACH approaches the MSE-values of the MMSE estimator very quickly, while PEACH needs a higher L than W-PEACH to get close to the MMSE curves.

In Fig. 3, we compare different estimators with or without additional interference from pilot contamination. We consider a fixed $L = 10$ and vary the SNR γ. As expected, the MSE of the MMSE and MVU estimators decay steeply to

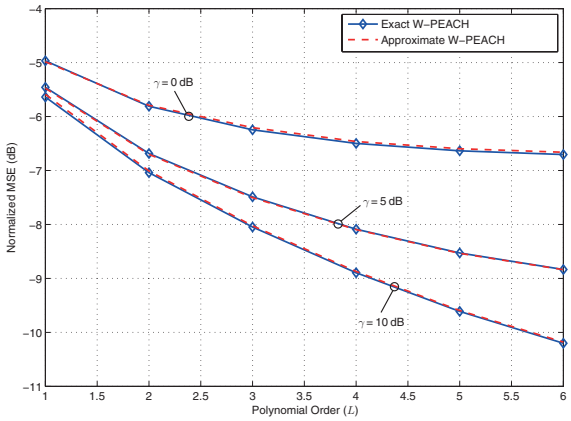

Fig. 4. Comparison of W-PEACH estimator and Approximate W-PEACH estimator in a noise-limited scenario ($\beta = 0$) for different SNR γ values.

zero when the γ increases in the noise-limited scenario, while the MSE saturates to a non-zero error floor under pilot contamination. The MSE floor when $\beta \neq 0$ is due to the fact that the normalized signal-to-interference-and-noise ratio (SINR) becomes almost constant as γ increases. To show this, we define the normalized SINR as

$$\frac{\gamma}{1 + K\beta\gamma} \tag{21}$$

where K is the number of interferers. As γ increases, the SINR in (21) approaches $\frac{1}{K\beta} > 0$, then the MSE values, which are different functions of the SINR for different estimators, approach some non-zero limits.

We observe from Fig. 3 that pilot contamination does not make significant impact on the PEACH and W-PEACH estimators; in fact, pilot contamination is beneficial in the sense that it reduces the gap to the optimal MMSE estimator. This remarkable result is explained as follows. For any fixed L, PEACH and W-PEACH will converge to a non-zero MSE when γ increases, due to the bias generated by the approximation error. Since this also happens for the MMSE and MVU estimators under pilot contamination, the relative loss of using the proposed low-complexity estimators will be smaller. Consequently, we can reduce L as β increases and still achieve near-optimal performance.

Finally, we focus on the low-complexity approach in Algorithm 1 for finding the weights. Fig. 4 considers a noise-limited scenario and a time window of length $T = 100$. Although $T \ll BN_r$, we observe that the approximate W-PEACH estimator which exploits the approximate weights from Algorithm 1 gives almost identical performance as W-PEACH with optimal weights computed according to Theorem 2. This confirms that the W-PEACH estimator is indeed a low-complexity channel estimator suitable for large-scale MIMO systems.

V. Conclusions

Large-scale MIMO techniques offer a high spatial resolution that can drastically increase the spectral and/or energy efficiency of wireless systems. Acquiring accurate CSI is a key to achieve these potential improvements in practice. Since the conventional pilot-based MMSE and MVU channel estimators have a computational complexity unsuitable for scenarios with large numbers of antennas, we proposed a set of low-complexity PEACH estimators. These are based on approximating the inversion of covariance matrices in the MMSE estimator by an L-order matrix polynomial. The proposed estimators converge to the MMSE estimator as L grows large, but by deriving the optimal coefficients for the polynomial we can obtain near-optimal MSE performance at small values on L. In practice, the order L can be selected to balance between complexity and MSE performance. Numerical results are given for noise-limited scenarios as well as under pilot contamination from pilot reuse in adjacent systems. Although pilot contamination can create an MSE floor, it is actually beneficial to the proposed estimators in the sense that smaller L can be used to achieve good performance.

References

[1] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*, 1st ed. Wiley, 2012.

[2] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

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

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

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

[6] R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," in *Proc. ITG Workshop on Smart Antennas (WSA)*, 2013.

[7] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[8] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, 2004.

[9] Y. Liu, T. Wong, and W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1486–1497, 2007.

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

[11] S. Moshavi, E. Kanterakis, and D. Schilling, "Multistage linear receivers for DS-CDMA systems," *Int. J. Wireless Information Networks*, vol. 3, no. 1, pp. 1–17, 1996.

[12] M. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1928–1946, 2001.

[13] G. Sessler and F. Jondral, "Low complexity polynomial expansion multiuser detector for CDMA systems," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1379–1391, 2005.

[14] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Proc. IEEE ISIT*, 2011.

[15] V. Williams, "Multiplying matrices faster than Coppersmith-Winograd," in *Proc. STOC*, 2012, pp. 887–898.

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

[17] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.