

Blind Pilot Decontamination

Ralf R. Müller

Dept Electronics & Telecommunications
Norwegian University of Science & Technology
7491 Trondheim, Norway

Mikko Vehkaperä

Dept Signal Processing & Acoustics
Aalto University
02150 Espoo, Finland

Laura Cottatellucci

Institute Eurecom
450 route des Chappes
06410 Biot, France

Abstract—We propose a nonlinear channel estimation method based on a subspace projection that is suitable for asymmetric antenna array systems. We show that the so-called pilot contamination problem reported in [1], is an artefact of (inappropriate) linear channel estimation and does not occur in cellular systems with power-controlled handoff when the channel estimation method proposed in this paper is used. We intuitively explain our result by establishing an isomorphism between a massive MIMO system and a spread-spectrum system with unknown signature sequences.

Keywords—Multiple antennas, multiple-input multiple-output (MIMO) systems, massive MIMO, spread-spectrum, channel estimation, principal component analysis

I. INTRODUCTION

In [2], a multiple antenna system was proposed that mimics the idea of spread-spectrum. Like a large processing gain can be realized in a spread-spectrum system by massive use of radio spectrum, a large array gain is realized by a massive use of antennas elements. This system design has attracted frequent attention recently, see e.g. [3], and it is commonly referred to as *massive MIMO*. Its advantage over the old spread-spectrum idea lies in the fact that antennas can be manufactured in arbitrarily high numbers, while radio spectrum is limited. Since the array gain grows unboundedly with the number of antenna elements at the access point, multiuser interference can be overcome regardless of the powers and number of the interfering users.

In [1], however, a pessimistic conclusion about the performance of massive MIMO in cellular systems was reached. Based on the implicit assumption of linear channel estimation in [1, Eq. (5)], it was concluded that the array gain can be achieved only for data detection, but not for channel estimation. As a result, pilot interference from interfering cells would limit the ability to estimate the channel sufficiently accurate regardless of the number of antenna elements at the access point. This effect, commonly referred to as *pilot contamination*, was believed by many researchers, e.g. [4]–[6] to be a fundamental effect, despite the lack of a solid proof that it cannot be overcome.

Using Bayesian channel estimation, [7] found that pilot contamination can vanish under certain conditions of the channel covariance. In this paper, we show that pilot contamination is, in fact, not a fundamental effect, but a shortcoming of linear channel estimation. We show, that the array gain can easily be utilized to also have the accuracy of channel estimation to grow unboundedly with the number of antennas. Furthermore, we show that this can be achieved with polynomial complexity in the number of antenna elements.

In Section II, we introduce the system model. In Section III, we propose an algorithm for nonlinear channel estimation utilizing the array gain. In Sections IV and V, we investigate the performance of this algorithm by analytic and simulative means, respectively. Finally, Section VI concludes the paper. Technical derivations are placed in the two appendices.

II. SYSTEM MODEL

Consider a wireless communication channel in the up-link. In order to ease notation, let the channel bandwidth be smaller than the coherence bandwidth. Channels whose physical bandwidth is wider than the coherence bandwidth can be decomposed into equivalent parallel narrowband channels by means of orthogonal frequency division multiplexing or related techniques.

Let the frequency-flat, block-fading, narrowband channel from T transmit antennas to $R > T$ receive antennas be described by the matrix equation

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \quad (1)$$

where $\mathbf{X} \in \mathbb{C}^{T \times C}$ is the transmitted data (eventually multiplexed with pilot symbols), $C \geq R^1$ is the coherence time in multiples of the symbol interval, $\mathbf{H} \in \mathbb{C}^{R \times T}$ is the channel matrix of unknown propagation coefficients, $\mathbf{Y} \in \mathbb{C}^{R \times C}$ is the received signal, and $\mathbf{Z} \in \mathbb{C}^{R \times C}$ is additive noise. Furthermore, we assume that channel, data, and noise have zero mean, i.e. $\mathbb{E}\mathbf{X} = \mathbb{E}\mathbf{H} = \mathbb{E}\mathbf{Z} = \mathbf{0}$. The noise accounts for both thermal noise and interference from other cells and is, in general, neither white nor Gaussian.

Note that (1) can also be understood as a code-division multiple-access (CDMA) system with the columns of \mathbf{H} denoting the spreading sequences and R denoting the processing gain. It is well-known that CDMA can be demodulated without knowledge of the spreading sequences by means of blind algorithms, see e.g. [8]. Those algorithms can also be applied in massive MIMO systems as proposed in [9]. In the following section, we introduce an algorithm, which we consider particularly suited for cellular massive MIMO.

III. PROPOSED ALGORITHM

Before going into the details of the proposed algorithm, we start with the idea behind the proposed procedure. Consider the channel model (1) for a single active transmit antenna, i.e. $T = 1$ and look for the matched filter \mathbf{m}^\dagger such that the

¹The assumption $C \geq R$ is made to simplify the exposition. In fact, all the formulas presented in the following hold for $C < R$, as well, although their derivations might require modifications.

signal-to-noise ratio (SNR) at its output is maximum. In white noise, maximizing the SNR is equivalent to maximizing the total received power normalized to the power gain of the filter. Thus, the optimum filter is given by

$$\mathbf{m}^* = \underset{\mathbf{m}}{\operatorname{argmax}} \frac{\mathbf{m}^\dagger \mathbf{Y} \mathbf{Y}^\dagger \mathbf{m}}{\mathbf{m}^\dagger \mathbf{m}}. \quad (2)$$

It is a well-known result of linear algebra that the vector \mathbf{m}^* maximizing the right hand side of (2) is that eigenvector of $\mathbf{Y} \mathbf{Y}^\dagger$ that corresponds to the largest eigenvalue of $\mathbf{Y} \mathbf{Y}^\dagger$. Having found the optimum algorithm for a single transmitter and white noise, we now apply this idea to multiple transmit antennas and analyze its performance in colored noise.

Consider the singular value decomposition

$$\mathbf{Y} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\dagger \quad (3)$$

with unitary matrices $\mathbf{U} \in \mathbb{C}^{R \times R}$ and $\mathbf{V} \in \mathbb{C}^{C \times C}$ and the $R \times C$ diagonal matrix $\mathbf{\Sigma}$ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R$ sorted in non-increasing order. Now, decompose the matrix of left singular vectors

$$\mathbf{U} = [\mathbf{S} | \mathbf{N}] \quad (4)$$

into the signal space basis $\mathbf{S} \in \mathbb{C}^{R \times T}$ and the null space basis $\mathbf{N} \in \mathbb{C}^{R \times (R-T)}$. Now, we project the received signal onto the signal subspace and get

$$\tilde{\mathbf{Y}} = \mathbf{S}^\dagger \mathbf{Y}. \quad (5)$$

As will become clear in the following, there is no need to explicitly calculate the full singular value decomposition (3). Only the basis of the signal subspace \mathbf{S} is needed and there are efficient algorithms available to exclusively calculate \mathbf{S} .

Consider now the massive MIMO case, i.e. $R \gg T$: That means that the T -dimensional signal subspace is much smaller than the R -dimensional full space, which the noise lives in. White noise is evenly distributed in all dimensions of the full space. Thus, the influence of white noise onto the signal subspace becomes negligible as $R \rightarrow \infty$. Using the algorithm above, we can achieve an array gain even without the need for estimating the channel coefficients. In fact, the channel can be estimated solely after projection onto the signal subspace when the dominant part of the white noise has already been suppressed.

In practice, it might be even a good idea not to estimate the channel matrix \mathbf{H} , at all. Instead, one might directly consider the subspace channel

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{H}} \mathbf{X} + \tilde{\mathbf{Z}} \quad (6)$$

and estimate the much smaller subspace channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{T \times T}$. Although the data dependent projection (5) implies that the noise $\tilde{\mathbf{Z}} \in \mathbb{C}^{T \times C}$ is not independent from the data \mathbf{X} , neglecting this dependence is an admissible approximation that becomes exact as the number of receive antennas R grows large.

In addition to white noise, there is co-channel interference from L neighboring cells. This interference is typically not white. In fact, it is the more colored, the smaller the ratio

$$\alpha = \frac{T}{R} \quad (7)$$

which will be called *load* in the following. In the limit of zero load, i.e. $\alpha \rightarrow 0$, the subspace spanned by co-channel interference is orthogonal to the signal subspace. Moreover, any R -dimensional channel vector from any transmitter, be it in the cell of interest or in a neighboring cell, to the receive array in the cell of interest is orthogonal to any other channel vector. That means that in the limit $R \rightarrow \infty$, the $(L+1)T$ largest singular values of the received signal matrix \mathbf{Y} become identical to the Euclidean norms of the $(L+1)T$ channel vectors.

If we could identify which singular values correspond to channel vectors from inside the cell as opposed to channel vectors from transmitters in neighboring cells, we could remove the interference from neighboring cells by subspace projection. Note that for $R \rightarrow \infty$, the system has infinite diversity, thus the effect of short-term fading (Rayleigh fading) vanishes. Thus, the norm of a channel vector is solely determined by path loss and long-term fading (shadowing). In a cellular system with power-controlled handoff strategy, the norm of channel vectors from neighboring cells can never be greater than the norm of channel vectors from the cell of interest. We conclude that the identification of singular values belonging to transmitters within the cell of interest is possible by means of ordering them by magnitude. Admittedly, such orderings might be inaccurate for a small fraction of transmitters that happen to experience similar channel conditions to more than a single access point. In practice, this ambiguity can be overcome by a smart choice of frequency re-use patterns that ensure a certain margin of power separation between users within the cell of interest and users from neighboring cells.

IV. PERFORMANCE ANALYSIS

We have demonstrated above, that the proposed algorithm works in principle in massive MIMO systems as the number of receive antennas grows much larger than the product of transmit antennas and neighboring cells. In this section, we will focus on the question, how large is large enough in practice.

A. Approximate Analysis

We start with a performance analysis for L finite, $R, T \rightarrow \infty$, and $0 \neq \alpha \ll 1$. This regime is the classical massive MIMO setting and we will see that it leads to explicit and intuitive design guidelines.

We decompose the impairment process

$$\mathbf{Z} = \mathbf{W} + \mathbf{H}_1 \mathbf{X}_1 \quad (8)$$

into white noise \mathbf{W} and interference from L neighboring cells where data $\mathbf{X}_1 \in \mathbb{C}^{LT \times R}$ is transmitted and received in the cell of interest through the channel $\mathbf{H}_1 \in \mathbb{C}^{R \times LT}$. We define the normalized coherence time

$$\kappa = \frac{C}{R}. \quad (9)$$

Furthermore, we assume that the elements of \mathbf{W} are independent and identically distributed (iid) with zero-mean and variance W .

In the large antenna limit $R \rightarrow \infty$, the singular values of \mathbf{W}/\sqrt{CW} follow the Marchenko-Pastur law, i.e.

$$p_\sigma(x) = \frac{\sqrt{\frac{4}{\kappa} - (x - 1 - \frac{1}{\kappa})^2}}{\pi x} \quad (10)$$

for $1/\sqrt{\kappa} - 1 < x < 1/\sqrt{\kappa} + 1$. In the worst case, the T largest singular values of the noise affect the signal of interest. The power of white noise being present in $\tilde{\mathbf{Y}}$ is thus at most

$$TCW \left(1 + \frac{1}{\sqrt{\kappa}}\right)^2. \quad (11)$$

Let the entries of the data signal \mathbf{X} be iid with zero mean and variance P . Furthermore, let the entries of the channel matrix \mathbf{H} be also iid with zero mean, but have unit variance. The total power of the signal of interest at the receiver is thus $TRCP$ and the signal-to-noise ratio in $\tilde{\mathbf{Y}}$ is lower bounded by

$$\text{SNR} \geq \frac{P}{W} \frac{R}{\left(1 + \frac{1}{\sqrt{\kappa}}\right)^2}. \quad (12)$$

The signal-to-noise ratio scales linearly with the number of receive antennas R and can be made as large as desired by adding more and more receive antennas. The influence of the coherence time $C \geq R$ onto the signal-to-noise ratio is at most a factor of 4 and plays only a minor role.

In addition to white noise, there is co-channel interference from neighboring cells. The co-channel interference is not white, but highly concentrated in certain subspaces. It might look hopeless to try to suppress this co-channel interference without explicit knowledge of a basis spanning the interference subspace. However, a phase-transition of spectra of large random matrices comes to our aid. The empirical distribution of the squared singular values of the normalized signal of interest, i.e. $\mathbf{H}\mathbf{X}/\sqrt{TR}$, is shown in Appendix A to converge, as $R \rightarrow \infty$, to a limit distribution which for $\alpha \ll 1$ is supported in the interval

$$\mathcal{P} = \left[\frac{\kappa P}{\alpha} - 2P\sqrt{\frac{\kappa^2 + \kappa}{\alpha}}; \frac{\kappa P}{\alpha} + 2P\sqrt{\frac{\kappa^2 + \kappa}{\alpha}} \right]. \quad (13)$$

Let the entries of the matrix of interfering signals be iid with zero mean and variance P and let the entries of the matrix of interfering channels be iid with zero mean and variance I/P such that the ratio I/P accounts for the relative attenuation between intercell users and out-of-cell users. Then, the empirical distribution of the squared singular values of the normalized co-channel interference, i.e. $\mathbf{H}_I\mathbf{X}_I/\sqrt{TR}$, also converges to a limit distribution. For $\alpha \ll 1$, it is supported in the interval

$$\mathcal{I} = \left[\frac{\kappa I}{\alpha} - 2I\sqrt{L\frac{\kappa^2 + \kappa}{\alpha}}; \frac{\kappa I}{\alpha} + 2I\sqrt{L\frac{\kappa^2 + \kappa}{\alpha}} \right] \quad (14)$$

If the two supporting intervals do not overlap, i.e.

$$\mathcal{P} \cap \mathcal{I} = \emptyset \quad (15)$$

or equivalently

$$\frac{P}{I} > \frac{1 + 2\sqrt{\alpha L \left(1 + \frac{1}{\kappa}\right)}}{1 - 2\sqrt{\alpha \left(1 + \frac{1}{\kappa}\right)}} \quad (16)$$

$$\stackrel{\alpha \ll 1}{\approx} 1 + 2 \left(1 + \sqrt{L}\right) \sqrt{\alpha \left(1 + \frac{1}{\kappa}\right)}, \quad (17)$$

the singular value distribution of the sum of the signal of interest and the interference converges, as $R \rightarrow \infty$, to a limit distribution that is composed of two separate non-overlapping bulks [10]. Note that in the limit $\alpha \rightarrow 0$, the signal bulk always separates from the interference bulk as long as $P/I > 1$. Therefore, the signal subspace and the interference subspace can be identified blindly. The interference can be nulled out and pilot contamination does *not* happen. We note that the bulk separation was calculated in the absence of white noise. In the presence of strong white noise, the actual bulk separation is slightly smaller.

B. Exact Large System Analysis

In this section, we still consider a larger system where the number of transmit antennas T and the number of receive antennas R is infinite, but their ratio α is not very small. The obtained results will be more accurate, but implicit and less intuitive than those of the previous section.

Combining (1) and (8), we get

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{H}_I\mathbf{X}_I + \mathbf{W}. \quad (18)$$

Let us denote the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^\dagger$ as $\mathcal{P}_{\mathbf{Y}\mathbf{Y}^\dagger}(x)$. In Appendix B, we show that this asymptotic eigenvalue distribution obeys

$$\begin{aligned} s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 = & \\ & - \frac{PTC\alpha (s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 - \kappa) \mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s)}{\alpha\kappa - PTC (s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 - \kappa) \mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s)} \\ & - \frac{ILTC\alpha (s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 - \kappa) \mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s)}{\alpha\kappa - ITC (s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 - \kappa) \mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s)} \\ & - \frac{WC (s\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) + 1 - \kappa) \mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s)}{\kappa} \end{aligned} \quad (19)$$

with

$$\mathcal{G}_{\mathbf{Y}\mathbf{Y}^\dagger}(s) = \int \frac{d\mathcal{P}_{\mathbf{Y}\mathbf{Y}^\dagger}(x)}{x - s} \quad (20)$$

denoting its Stieltjes transform.

V. NUMERICAL RESULTS

In this section, we provide Monte-Carlo simulation results for the uncoded bit error rate (BER) of quaternary phase-shift keying (QPSK) in flat Rayleigh fading and compare the proposed algorithm to the linear channel and data estimation considered in [1]. First, we consider the case of high SNR in Fig. 1. Although, this case is not relevant in practice, it shows that the BER drops to arbitrarily low values if the co-channel interference is below the threshold provided by random matrix theory (RMT) in (16). Thus, it confirms that the pilot contamination problem is overcome, in principle.

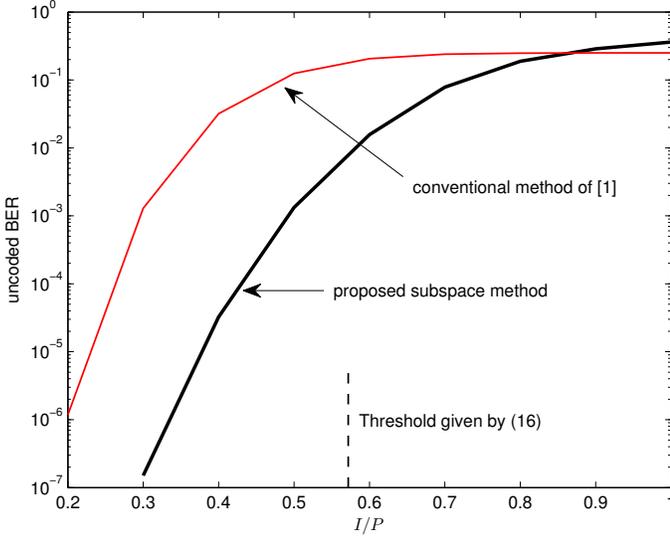


Fig. 1. BER for 1 pilot symbol per transmit antenna and cell, $R = 400$ receive antennas, $T = 4$ transmit antennas, $L = 2$ neighboring cells, coherence time of $C = 1000$ symbols, and signal-to-noise ratio $P/W = 100$ (20 dB). Single-user matched filtering for data detection.

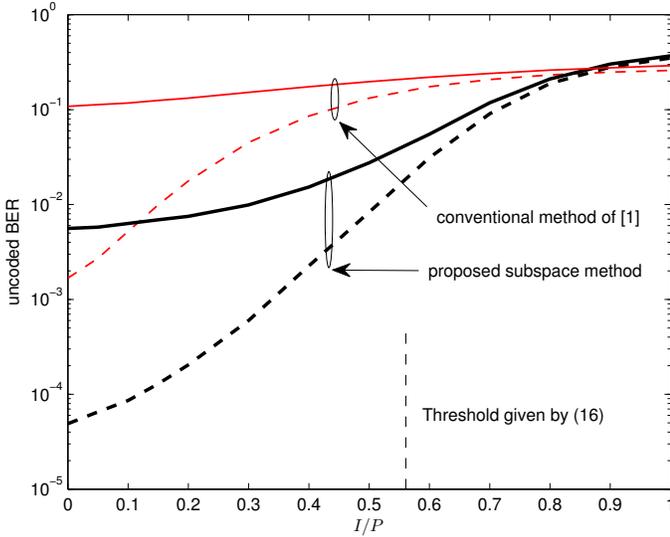


Fig. 2. BER for $R = 200$ receive antennas, $T = 2$ transmit antennas, $L = 2$ neighboring cells, coherence time of $C = 400$ symbols, and signal-to-noise ratio $P/W = 0.1$ (-10 dB). Single-user matched filtering for data detection. The solid and dashed lines refer to 1 and 10 pilot symbols per transmit antenna and cell, respectively.

The practically relevant case of low SNR is depicted in Fig. 2. Again, the proposed algorithm achieves significant performance gains below the RMT threshold when compared to linear channel estimation. While the RMT threshold is an exact threshold for the asymptotic case $T \rightarrow \infty$ and $R \gg T$, the figure also shows the largest ratio I/P which did not lead to separation of the singular values in the simulations. As expected, the singular values spread out somewhat wider for a finite number of antennas as compared to the asymptotically large case.

The asymptotic eigenvalue distribution is compared to a simulated histogram in Fig. 3. The histogram matches well

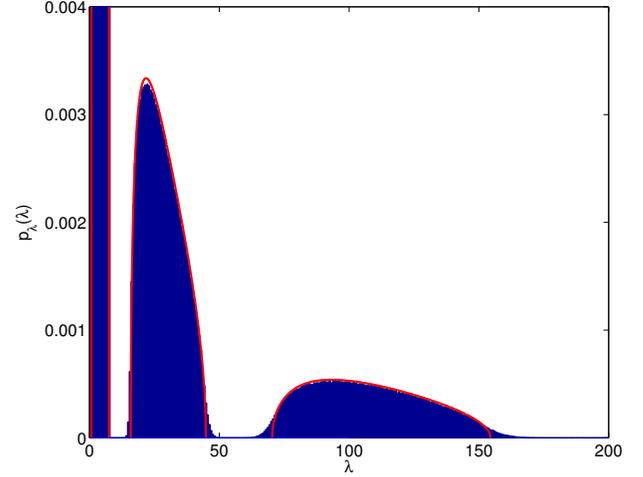


Fig. 3. Eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^\dagger/R$ for $R = 300$ receive antennas, $T = 10$ transmit antennas, coherence time $C = 1000$, $L = 2$ neighboring cells, unit noise power ($W = 1$), $P/W = 0.1$ (-10 dB), and $P/I = 4$. The red line shows the asymptotic result of (19).

with the asymptotic result. However, the bulk separation is less pronounced in the finite case.

VI. SUMMARY AND CONCLUSIONS

We proposed a practical algorithm with polynomial complexity to avoid pilot contamination in cellular systems with power controlled handoff and appropriate frequency-reuse patterns. The dominant complexity of this algorithm is a singular value decomposition of received signal block. We conclude, that pilot contamination is not a fundamental effect, but an artefact of linear channel estimation.

APPENDIX A

It is shown in [11, Eq. (31)], that the asymptotic eigenvalue distribution of $\mathbf{X}^\dagger \mathbf{H}^\dagger \mathbf{H} \mathbf{X} / TR$ has a Stieltjes transform $G(s)$ fulfilling

$$s^2 \kappa^2 G^3(s) + s \kappa (\alpha + 1 - 2\kappa) G^2(s) + (s\alpha + (\kappa - 1)(\kappa - \alpha)) G(s) - \alpha = 0. \quad (21)$$

The support of the distribution is given by the interval $[x_1; x_2]$ where x_1 and x_2 are the two largest nonnegative solutions to the equation [11, Eq. (37)]

$$4x^3 - \left(10\kappa + 10 + \frac{10\kappa}{\alpha} - \alpha - \frac{\kappa^2}{\alpha} - \frac{1}{\alpha}\right) x^2 + 2 \left(4\kappa^2 + 4 + \frac{4\kappa^2}{\alpha^2} - 2\kappa - \frac{2\kappa^2}{\alpha} - \frac{2\kappa}{\alpha}\right) x - \frac{\kappa^3}{\alpha} - \frac{\kappa^3}{\alpha^2} - \kappa\alpha - \frac{\kappa}{\alpha^2} - \alpha - \frac{1}{\alpha} = 0. \quad (22)$$

For $\alpha \ll 1$, this can be approximated by

$$4x^3 - (10\kappa - \kappa^2 - 1) \frac{x^2}{\alpha} + 2(4\kappa^2 - \kappa^3 - \kappa - 8\kappa\alpha - 8\kappa^2\alpha) \frac{x}{\alpha^2} + \frac{(\kappa - 1)^2}{\alpha^3} (\kappa^2 - 4\kappa\alpha - 4\kappa^2\alpha) = 0. \quad (23)$$

It can easily be verified that (23) has the following three roots:

$$x_1 = \frac{\kappa}{\alpha} - 2\sqrt{\frac{\kappa^2 + \kappa}{\alpha}} \quad (24)$$

$$x_2 = \frac{\kappa}{\alpha} + 2\sqrt{\frac{\kappa^2 + \kappa}{\alpha}} \quad (25)$$

$$x_3 = -\frac{(\kappa - 1)^2}{4\alpha} < 0. \quad (26)$$

This completes the derivation.

APPENDIX B

Consider the random matrix

$$D = \sum_{k=1}^K a_k B_k C_k \quad (27)$$

with $a_k \in \mathbb{R}$, $B_k \in \mathbb{C}^{n \times m_k}$ and $C_k \in \mathbb{C}^{m_k \times n}$ being random matrices with iid. zero-mean entries with variance $1/m_k$ and $1/n$, respectively. First, we will derive the asymptotic eigenvalue distribution of DD^\dagger in terms of its Stieltjes transform $G_{DD^\dagger}(s)$. Let $D_k = B_k C_k$. From [11, Eq. (31)], we have

$$-s^2 G_{D_k D_k^\dagger}^3(s) - s(\rho_k - 1) G_{D_k D_k^\dagger}^2(s) + s\rho_k G_{D_k D_k^\dagger}(s) = \rho_k. \quad (28)$$

with

$$\rho_k = \frac{m_k}{n} \quad (29)$$

With [10, Lemma 1], we get

$$-s\tilde{G}_{D_k}^3(s) - (\rho_k - 1)\tilde{G}_{D_k}^2(s) + s\rho_k\tilde{G}_{D_k}(s) = \rho_k \quad (30)$$

with $\tilde{G}_{D_k}(s)$ denoting the Stieltjes transform of the symmetrized singular value distribution of D_k . The definition of the R-transform [12] gives

$$\tilde{R}_{D_k}(w) = \frac{\rho_k w}{\rho_k - w^2} \quad (31)$$

and additive free convolution implies

$$\tilde{R}_D(w) = \sum_{k=1}^K \frac{a_k^2 \rho_k w}{\rho_k - a_k^2 w^2}. \quad (32)$$

It follows straightforwardly from the definition of the R-transform that

$$\frac{1}{\tilde{G}_D(s)} = -s + \tilde{R}_D(-\tilde{G}_D(s)) \quad (33)$$

and with [10, Lemma 1] that

$$\frac{1}{G_{DD^\dagger}(s)} = -s + \sqrt{s} \tilde{R}_D(-\sqrt{s} G_{DD^\dagger}(s)) \quad (34)$$

$$= -s - \sum_{k=1}^K \frac{a_k^2 \rho_k s G_{DD^\dagger}(s)}{\rho_k - a_k^2 s G_{DD^\dagger}^2(s)}. \quad (35)$$

Next we consider the decomposition

$$D = \begin{bmatrix} E \\ F \end{bmatrix} \quad (36)$$

with $E \in \mathbb{C}^{\beta n \times n}$. From [13, Theorem 14.10], we have

$$R_{EE^\dagger}(w) = R_{DD^\dagger}(\beta w). \quad (37)$$

In the Stieltjes domain, this R-transform relation translates into

$$\beta G_{EE^\dagger}(s) = G_{DD^\dagger}\left(s + \frac{\beta - 1}{\beta G_{EE^\dagger}(s)}\right). \quad (38)$$

Thus, we find with (35)

$$\frac{1}{\beta G_{EE^\dagger}(s)} = -s - \frac{\beta - 1}{\beta G_{EE^\dagger}(s)} \quad (39)$$

$$- \sum_{k=1}^K \frac{a_k^2 \rho_k \left(s + \frac{\beta - 1}{\beta G_{EE^\dagger}(s)}\right) \beta G_{EE^\dagger}(s)}{\rho_k - a_k^2 \left(s + \frac{\beta - 1}{\beta G_{EE^\dagger}(s)}\right) \beta^2 G_{EE^\dagger}^2(s)}$$

and

$$s G_{EE^\dagger}(s) = -1 - \sum_{k=1}^K \frac{a_k^2 \rho_k \left(s + \frac{\beta - 1}{\beta G_{EE^\dagger}(s)}\right) \beta G_{EE^\dagger}^2(s)}{\rho_k - a_k^2 \left(s + \frac{\beta - 1}{\beta G_{EE^\dagger}(s)}\right) \beta^2 G_{EE^\dagger}^2(s)}. \quad (40)$$

Now, we consider the matrix Y in (18) as a special case of E . This implies

$$K = 3 \quad (41)$$

$$\beta = \frac{R}{C} = \frac{1}{\kappa} \quad (42)$$

$$\rho_1 = \frac{T}{C} = \frac{\alpha}{\kappa} \quad (43)$$

$$a_1^2 = PTC \quad (44)$$

$$\rho_2 = \frac{LT}{C} = \frac{\alpha L}{\kappa} \quad (45)$$

$$a_2^2 = ILTC \quad (46)$$

$$\rho_3 \rightarrow \infty \quad (47)$$

$$a_3^2 = WC \quad (48)$$

and (19) is obtained. Note that the entries of $B_3 C_3$ become iid. as $\rho_3 \rightarrow \infty$.

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [2] —, "How much training is required for multiuser MIMO?" in *Fortieth Asilomar Conf. on Signals, Systems, & Computers*, Pacific Grove, CA, USA, Oct. 2006.
- [3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *To appear in IEEE Signal Processing Magazine*.
- [4] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.
- [5] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, San Francisco, CA, USA, Jun. 2011.

- [6] F. Fernandez, A. Ashikhmin, and T. Marzetta, "Interference reduction on cellular networks with large antenna arrays," in *Proc. of IEEE International Conference on Communications (ICC)*, Ottawa, Canada, Jun. 2012.
- [7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *Submitted to IEEE Journal on Selected Areas in Communications*, Mar. 2012, arXiv:1203.5924.
- [8] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2049–2069, Oct. 1998.
- [9] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012.
- [10] R. R. Müller and B. Çakmak, "Channel modelling of MU-MIMO systems by quaternionic free probability," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012.
- [11] R. R. Müller, "A random matrix model for communication via antenna arrays," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2002.
- [12] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*. Providence, RI: American Mathematical Society, 1992.
- [13] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*. Cambridge, U.K.: Cambridge University Press, 2006.