This is a repository copy of *Blind Pilot Decontamination*.

Article:

Blind Pilot Decontamination

Ralf R. Müller*, *Senior Member, IEEE* and Laura Cottatellucci and Mikko Vehkaperä

Abstract

A subspace projection to improve channel estimation in massive multi-antenna systems is proposed and analyzed. Together with power-controlled hand-off, it can mitigate the pilot contamination problem without the need for coordination among cells. The proposed method is blind in the sense that it does not require pilot data to find the appropriate subspace. It is based on the theory of large random matrices that predicts that the eigenvalue spectra of large sample covariance matrices can asymptotically decompose into disjoint bulks as the matrix size grows large. Random matrix and free probability theory are utilized to predict under which system parameters such a bulk decomposition takes place. Simulation results are provided to confirm that the proposed method outperforms conventional linear channel estimation if bulk separation occurs.

Index Terms

Multiple antennas, multiple-input multiple-output (MIMO) systems, massive MIMO, spread-spectrum, channel estimation, principal component analysis, random matrices, free probability.

I. INTRODUCTION

In [1], 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 considerable

attention recently, see e.g. [2] for a survey. 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.

Given perfect channel state information, the signals received at all antenna elements can be combined coherently. The array gain grows unboundedly with the number of antenna elements at the access point. Therefore, massive use of antennas elements can overcome both multiuser interference and thermal noise for any given number of users and any given powers of the interfering users.

Given a rich scattering environment, the access point is able to concentrate the radiated energy within an area around the designated receiver that is only a fraction of a squared wavelength [2]. This is in contrast to earlier methods of array processing in wireless communications based upon angles of arrival and departure which are limited to concentrate their energy into a certain direction rather than an area limited in both angular and radial domain.

Since power in the far-field can be amplified (in theory) beyond limits by means of array gains, the ultimate limit of a single-cell massive MIMO system is only constrained by the coherence time of the channel [3]. The number of terminals needs to stay below the coherence time to have sufficient degrees of freedom for channel estimation. In practice, the coherence time needs to be significantly larger than the number of terminals to also allow for data transmission in addition to pilot symbols for channel estimation.

In [3], however, a pessimistic conclusion about the performance of massive MIMO in *multi-cellular* systems was reached. Based on the explicit assumption of no coordination among cells and on the implicit assumption of linear channel estimation [3, Eq. (5)], it was concluded that the array gain can be achieved only for data detection, but not for channel estimation. The author argued that channel state information, though not required to be perfect, must have at least a certain quality in order to utilize unlimited array gains. As a result, pilot interference from neighboring cells would limit the ability to obtain sufficiently accurate channel estimates and be the new bottleneck of the system. This effect, commonly referred to as *pilot contamination* [4], is treated as a fundamental effect in many works, e.g. [3]–[8].

Recent works have indicated that pilot contamination may not be as fundamental as it was thought to be: Using Bayesian channel estimation, [9] found that pilot contamination can vanish under certain conditions on the channel covariance matrix if some cooperation among cells is allowed for. Using an eigenvalue decomposition of the sample covariance matrix of the received signal, [10] found that for a wide range of system parameters, the channel can be estimated with greater accuracy than with linear methods.

In early conference versions of this work [11], [12], we showed that pilot contamination is not a fundamental limitation for massive MIMO systems, if the coherence time is not smaller than the number of antennas at the base station and a power margin between users of interest and interfering users can be provided, e.g. by means of path loss. In that case the array gain can be utilized to have the accuracy of channel estimation growing unboundedly with the number of antennas at polynomial complexity. As in [10], the approach in [11], [12] starts with an eigenvalue decomposition of the sample covariance matrix (or equivalently a singular value decomposition of the received signal matrix). Unlike [10], it does not aim to subsequently estimate the channel matrix before performing data detection. It projects the received signal onto an (almost) interference-free subspace where communication is governed by a non-linear compound channel that can be estimated more easily.

The work in [11], [12] has received criticism for the assumption that the coherence time is assumed to be larger than the number of base station antennas. In this journal version of our work, we remove this assumption and also allow for coherence times shorter than the number of base station antennas. We find that the main conclusions of [11], [12] remain valid and show that our proposed method of channel estimation based on subspace projection still outperforms linear channel estimation, though by a smaller margin. In this journal version of our work, we also include proofs and derivations that were omitted in the conference proceedings due to space limitations. For sake of readability and convenience to the reader, we repeat parts of the material presented in [11], [12].

The paper is organized as follows: In Section II, we introduce the system model. In Section III, we introduce the algorithm for nonlinear channel estimation utilizing the array gain. In Section IV, we analyze the performance of the algorithm by means of random matrix theory in the limit of large number of base station antennas. In Section V, we investigate the performance of the algorithm for a finite number of antennas by simulative means. Finally, Section VI concludes the paper.

II. SYSTEM MODEL

Consider a wireless communication channel. In order to ease notation and for sake of conciseness, let the channel bandwidth be equal to 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. The coherence time of the channel measured in symbol intervals is thus given by [13]

$$C = \frac{3}{4\sqrt{\pi}f_0\tau}\frac{c}{v} \tag{1}$$

with f_0, τ, v, and c denoting carrier frequency, delay spread, mobile speed, and the speed of light, respectively. Considering extreme values for these parameters, e.g. bullet train speed of $v = 350$ km/h and very high delay spread $\tau = 5$ μs (which corresponds to an excess distance of 1.5 km) the coherence time at $f_0 = 2.6$ GHz is 99 symbols. For more typical speeds of mobile terminals and/or smaller cells, the coherence time can be one or several orders of magnitude larger.

In the following, we consider the uplink (reverse link) of a cellular massive MIMO system. Therefore, the number of receive antennas R is much larger than the number of transmit antennas T. The number of transmit antennas is limited by the richness of the propagation channel. Measurements in Manhattan [14, Fig. 5] show that the richness is limited to around 10 to 13 degrees of freedom. In more typical outdoor environments with fewer high buildings, the scattering richness will be even lower. It is therefore sensible to assume that the number of transmit antennas is small compared to both the number of receive antennas and the coherence time

$$R \gg T \ll C. \tag{2}$$

The number of receive antennas is constrained by their physical size. At 2.6 GHz, a uniform linear array spaced in half wavelengths reaches the length of $d = 6$ m for 104 elements. Spacing the elements on a two-dimensional grid, the number of receive antennas can be even larger. Obviously, the number of receive antennas can be larger, equal or smaller than the coherence time. We will therefore, introduce the normalized coherence time

$$\kappa = \frac{C}{R} \tag{3}$$

which is assumed to be a finite, non-zero constant throughout this paper.

Let the frequency-flat, block-fading propagation channel be described by the matrix equation

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{Z}, \tag{4}$$

where $\boldsymbol{X} \in \mathbb{C}^{T \times C}$ is the transmitted data (eventually multiplexed with pilot symbols), $\boldsymbol{H} \in \mathbb{C}^{R \times T}$ is the channel matrix of unknown propagation coefficients, $\boldsymbol{Y} \in \mathbb{C}^{R \times C}$ is the received signal, and $\boldsymbol{Z} \in \mathbb{C}^{R \times C}$ is the total impairment. Furthermore, we assume that channel, data, and impairment have zero mean, i.e. $\mathsf{E}\,\boldsymbol{X} = \mathsf{E}\,\boldsymbol{H} = \mathsf{E}\,\boldsymbol{Z} = \boldsymbol{0}$. The impairment includes both thermal noise and interference from other cells and is, in general, neither white nor Gaussian.

Note that (4), can also be understood as a code-division multiple-access (CDMA) system with the columns of $\boldsymbol{H}$ 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. [15]. Many of those algorithms can also be applied in massive MIMO systems. In the following section, we introduce an algorithm, which we consider particularly suited for cellular massive MIMO.

III. Proposed Algorithm

A. General Idea

Before going into the details of the proposed algorithm, we start with the idea behind the proposed procedure. Consider the channel model (4) for a single active transmit antenna, i.e. $T = 1$ and look for the matched filter $\boldsymbol{m}^\dagger$ such that the 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 by the power gain of the filter. Thus, the optimum filter is given by

$$\boldsymbol{m}^\circ = \operatorname*{argmax}_{\boldsymbol{m}} \frac{\boldsymbol{m}^\dagger \boldsymbol{J} \boldsymbol{m}}{\boldsymbol{m}^\dagger \boldsymbol{m}} \tag{5}$$

with

$$\boldsymbol{J} = \operatorname*{E}_{\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{H}} \left\{ \boldsymbol{Y} \boldsymbol{Y}^\dagger \right\}. \tag{6}$$

It is a well-known result of linear algebra that the vector $\boldsymbol{m}^\circ$ maximizing the right hand side of (5), commonly referred to as the Rayleigh quotient, is that eigenvector of $\boldsymbol{J}$ that corresponds to the largest eigenvalue of $\boldsymbol{J}$. Since we do not know the matrix $\boldsymbol{J}$, we have to cope with the approximate solution

$$\boldsymbol{m}^* = \operatorname*{argmax}_{\boldsymbol{m}} \frac{\boldsymbol{m}^\dagger \boldsymbol{Y} \boldsymbol{Y}^\dagger \boldsymbol{m}}{\boldsymbol{m}^\dagger \boldsymbol{m}}. \tag{7}$$

This approximation is tight for large number of antenna elements, i.e. we have the almost sure convergence of the inner product

$$|\langle \boldsymbol{m}^\circ; \boldsymbol{m}^* \rangle| \to ||\boldsymbol{m}^\circ|| \cdot ||\boldsymbol{m}^*|| \tag{8}$$

as $R \to \infty$, if the largest eigenvalue of the noise is negligible against the largest eigenvalue of the signal, i.e.

$$\lim_{R \to \infty} \frac{\displaystyle\max_{\boldsymbol{m}} \frac{\boldsymbol{m}^\dagger \boldsymbol{Z} \boldsymbol{Z}^\dagger \boldsymbol{m}}{\boldsymbol{m}^\dagger \boldsymbol{m}}}{\displaystyle\max_{\boldsymbol{m}} \frac{\boldsymbol{m}^\dagger \boldsymbol{H} \boldsymbol{X} \boldsymbol{X}^\dagger \boldsymbol{H}^\dagger \boldsymbol{m}}{\boldsymbol{m}^\dagger \boldsymbol{m}}} = 0. \tag{9}$$

Note that the limit $R \to \infty$ implies $C \to \infty$ due to (3). For finite number of antennas, there are better approximations for $\boldsymbol{J}$ than $\boldsymbol{Y} \boldsymbol{Y}^\dagger$, e.g. G-estimation [16]. However, such methods exceed the scope of the present work and are left for future research. G-estimation will further improve the performance of the proposed subspace projection method, in practice.

The limit condition (9) is not hard to fulfill. In fact, it holds true for independent constant variance entries in $\boldsymbol{Z}$, $\boldsymbol{H}$, and $\boldsymbol{X}$. To see this note that the largest eigenvalue of $\boldsymbol{Z}\boldsymbol{Z}^{\dagger}$ scales linearly with R, as the number of entries in $\boldsymbol{Z} \in \mathbb{C}^{R \times C}$ grows quadratically, but the number of non-zero eigenvalues grows linearly. At the same time the largest eigenvalue of $\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{H}^{\dagger}$ grows quadratically with R, as the number of entries in $\boldsymbol{H} \in \mathbb{C}^{R \times T}$ grows linearly, the number of entries in $\boldsymbol{X} \in \mathbb{C}^{T \times C}$ grows linearly, but the number of non-zero eigenvalues is T and thus constant.

B. Detailed Algorithm

Having found an 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

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\dagger} \tag{10}$$

with unitary matrices $\boldsymbol{U} \in \mathbb{C}^{R \times R}$ and $\boldsymbol{V} \in \mathbb{C}^{C \times C}$ and the $R \times C$ diagonal matrix $\boldsymbol{\Sigma}$ with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{R,C\}}$ sorted in non-increasing order. As shown in [10], the columns of $\boldsymbol{U}$ are highly correlated with the columns of $\boldsymbol{H}$. Based on this observation, [10] proposes two algorithms for improved nonlinear estimation of the channel matrix $\boldsymbol{H}$.

In the sequel, we pursue a strategy different from the one in [10]. We decompose the matrix of left singular vectors

$$\boldsymbol{U} = [\boldsymbol{S}|\boldsymbol{N}] \tag{11}$$

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

$$\tilde{\boldsymbol{Y}} = \boldsymbol{S}^{\dagger}\boldsymbol{Y}. \tag{12}$$

The null space basis $\boldsymbol{N}$ is not required in the sequel. In fact, there is no need to compute the full singular value decomposition (10). Only the basis of the signal subspace $\boldsymbol{S}$ is needed and there are efficient algorithms available to exclusively calculate $\boldsymbol{S}$.

Consider now the massive MIMO case, i.e. $R \gg T$: 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 $R \to \infty$. In other words: The considerations for the largest eigenvalue in (9) and its corresponding eigenvector in (8) are equally valid for the T largest eigenvalues and their corresponding eigenvectors, as long as T is finite.

Using the algorithm above, we can achieve an array gain even without the need for estimating the channel coefficients. In fact, channel estimation can be delayed until the received signal has been projected onto the signal subspace and the dominant part of the white noise has already been suppressed.

In order to save complexity it is sensible not to estimate the channel matrix $\boldsymbol{H}$, at all. Instead, we directly consider the subspace channel

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{H}}\boldsymbol{X} + \tilde{\boldsymbol{Z}} \tag{13}$$

and estimate the much smaller subspace channel matrix $\tilde{\boldsymbol{H}} \in \mathbb{C}^{T \times T}$ by standard methods of linear channel estimation based on pilot symbols. Hereby, we have transformed the problem of channel estimation for highly asymmetric massive MIMO systems into the well-explored problem of channel estimation for classical symmetric MIMO systems. Although the data dependent projection (12) implies that the noise $\tilde{\boldsymbol{Z}} = \boldsymbol{S}^{\dagger}\boldsymbol{Z} \in \mathbb{C}^{T \times C}$ is not independent from the data $\boldsymbol{X}$, neglecting this dependence is an admissible approximation that becomes exact due to (8), as the number of receive antennas R grows large.

In addition to white noise, there is co-channel interference from L neighboring cells. For sake of notational convenience, we assume that the number of transmit antennas is identical in all cells and equal to T. The interference from neighboring cells is anything but white. It is the more colored, the smaller the ratio

$$\alpha = \frac{T}{R} \tag{14}$$

which will be called *load* in the following. Any R-dimensional channel vector is orthogonal to any other channel vector in the limit $R \to \infty$ [3]. This holds regardless whether the two channel vectors correspond to transmitters in the same cell or in different cells. In the limit of zero load, i.e. $\alpha \to 0$, we have an even stronger result: the subspace spanned by the co-channel interference is orthogonal to the signal subspace.[1] That means that in the limit $R/T \to \infty$, the $(L+1)T$ largest singular values of the received signal matrix $\boldsymbol{Y}$ become identical to the Euclidean norms of the $(L+1)T$ channel vectors. We only need to identify which singular values correspond to channel vectors from inside the cell as opposed to channel vectors from transmitters in neighboring cells. Then, we can remove the interference from neighboring cells by subspace projection.

---

[1]Note that the pairwise orthogonality of channel vectors holds for $R \to \infty$, in general, and does not require $\alpha \to 0$. However, the orthogonality of subspaces requires $\alpha \to 0$ in addition to $R \to \infty$, as the accumulation of $T = \alpha R$ vanishing pairwise correlations is not vanishing, in general.

C. Identifying Signals of Interest

Note that for $R \to \infty$, the system has infinite diversity and 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 perfect received power control and a 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 in the limit $(R, \alpha) \to (\infty, 0)$, i.e. the number of receive antennas grows large while the number of transmit antennas does not.

For practical systems with small, but nonzero load, i.e. $0 < \alpha \ll 1$, a certain power margin is required between signals of interest and interfering signals. For most interfering users, such a power margin is created for free by shadowing and path loss. However, there might be few users close to cell boundaries who lack such a power margin. As a kind of countermeasure, a power margin has to be engineered for them. There are various ways to do so. In the sequel, we will exemplarily list two such potential methods.

One way to create an additional power margin is a smart choice of frequency or time re-use patterns. However, this requires coordination among cells. Another way to create an additional power margin is to equip each user with at least two transmit antennas. Then, the few users who suffer from insufficient power margin can form beams that favor one of the base stations or access points over others[2]. This will noticeable increase their power margins. The majority of users will not need to employ such methods and can use the two antennas for spatial multiplexing.

IV. Performance Analysis

We have demonstrated above, that the proposed algorithm works in principle in massive MIMO systems if the number of receive antennas is much larger than the product of transmit antennas and neighboring cells. In practical systems, the number of transmit and receive antennas is finite and the load α can be made very small but not arbitrarily small. The standard assumption of massive MIMO systems, i.e. $R \to \infty$, while T staying finite, gives overoptimistic results for finite systems of practical interest. In order to find the limits of the subspace projection method, we need to consider a more refined limit in the number of antennas. For that purpose, we utilized the asymptotic scale invariance of the eigenvalue spectra

---

[2]Note that such beam forming does not require channel state information. One can keep on forming random beams until a sufficient power margin is reached.

of large random matrices: For appropriate normalization of the entries, the spectra of random matrices are hardly affected (asymptotically invariant) if all matrix dimensions scale proportionally. Thus, a useful and insightful approach to understand the behavior of a real cellular massive MIMO network consists in assuming that both T and R grow large with a small, but fixed ratio α. This approach is tantamount to studying a system with $T = 5n$ transmit antennas, $R = 300n$ receive antennas and coherence time $C = 100n$ for $n \to \infty$ and assume that the result for a system with $T = 5$, $R = 300$ and $C = 100$ hardly differs.

We decompose the impairment process

$$\boldsymbol{Z} = \boldsymbol{W} + \boldsymbol{H}_\mathrm{I}\boldsymbol{X}_\mathrm{I} \tag{15}$$

into white noise $\boldsymbol{W}$ and interference from L neighboring cells where interfering data $\boldsymbol{X}_\mathrm{I} \in \mathbb{C}^{LT \times R}$ is transmitted in neighboring cells and received in the cell of interest through the channel $\boldsymbol{H}_\mathrm{I} \in \mathbb{C}^{R \times LT}$. Combining (4) and (15), we get

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{H}_\mathrm{I}\boldsymbol{X}_\mathrm{I} + \boldsymbol{W}. \tag{16}$$

Let the entries of the data signal $\boldsymbol{X}$ be iid with zero mean and variance P. Let the entries of the channel matrix $\boldsymbol{H}$ be also iid with zero mean, but have unit variance. Let the entries of the matrix of interfering signals $\boldsymbol{X}_\mathrm{I}$ be iid with zero mean and variance P and let the entries of the kth column of the matrix of interfering channels $\boldsymbol{H}_\mathrm{I}$ be independent with zero mean and variance I_k/P such that the ratio I_k/P accounts for the relative attenuation between out-of-cell user k and the intracell users. Let the empirical distribution of I_k converge to a limit distribution as $LT \to \infty$ which is denoted by $\mathrm{P}_I(\cdot)$. Furthermore, we assume that the elements of the noise $\boldsymbol{W}$ are independent and identically distributed (iid) with zero-mean and variance W. Let us denote the asymptotic eigenvalue distribution of $\boldsymbol{Y}\boldsymbol{Y}^\dagger$ as $\mathrm{P}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(x)$. In Appendix A, we show that this asymptotic eigenvalue distribution obeys

$$
\begin{aligned}
s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 = &- \frac{PTC\alpha \left(s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 - \kappa\right)\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s)}{\alpha\kappa - PTC\left(s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 - \kappa\right)\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s)} \\
&- \int \frac{xLTC\alpha \left(s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 - \kappa\right)\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s)\mathrm{d}\mathrm{P}_I(x)}{\alpha\kappa - xTC\left(s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 - \kappa\right)\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s)} \\
&- \frac{WC\left(s\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) + 1 - \kappa\right)\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s)}{\kappa}
\end{aligned}
\tag{17}
$$

with

$$\mathrm{G}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(s) = \int \frac{\mathrm{d}\mathrm{P}_{\boldsymbol{Y}\boldsymbol{Y}^\dagger}(x)}{x - s} \tag{18}$$

denoting its Stieltjes transform. By means of the Stieltjes inversion formula

$$\mathrm{p}(x) = \frac{1}{\pi} \lim_{y \to 0^+} \Im\mathrm{G}(x + \mathrm{j}y) \tag{19}$$
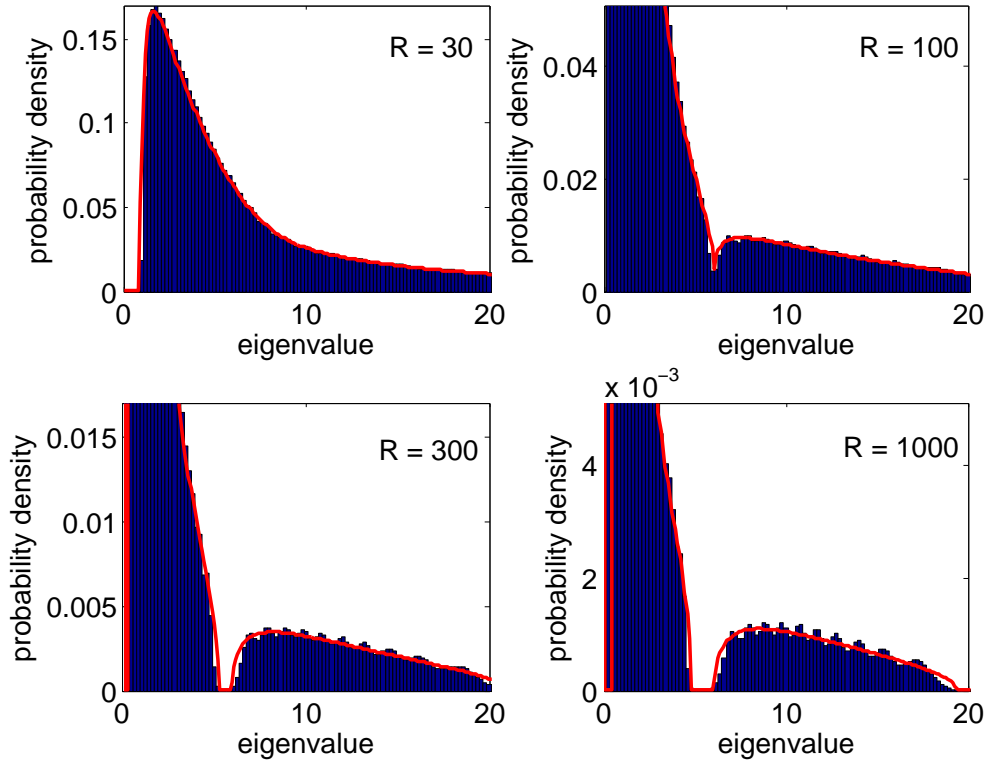
Fig. 1.   Asymptotic eigenvalue density of the matrix $\boldsymbol{Y}\boldsymbol{Y}^{\dagger}/R$ in solid red line for $\frac{\alpha}{\kappa} = \frac{1}{10}$, $P = -10$ dB, $I_k = \frac{P(k \bmod T)}{4T}$ $\forall k = 1 \ldots LT$, $L = 2$, $W = 0$ dB. The asymptotic eigenvalue distribution is compared to the empirical eigenvalue density for $T = 10$, $C = 100$, and various value of R given by the histograms in blue.

the asymptotic eigenvalue density is obtained.

In Figure 1, the solid lines in red shows the asymptotic eigenvalue distribution of $\boldsymbol{Y}\boldsymbol{Y}^{\dagger}/R$ obtained by (17)-(19). The histograms in blue show the empirical eigenvalue distributions of $\boldsymbol{Y}\boldsymbol{Y}^{\dagger}/R$ for $T = 10$, $C = 100$, and various values of R. We observe that for sufficiently large number of receive antennas R, the distribution decomposes into two disjunct bulks: A noise and interference bulk to the left and a bulk of the signal of interest to the right. If the bulks do not overlap, we can blindly separate the signals of interest from interference and noise as discussed in Section III-C.

The bulks are not disjunct in general, but only for certain values of the involved system parameters. It is therefore of utmost importance for practical design of blind pilot decontamination to know which system parameters do lead to bulk separation. The extremely good match between the asymptotic distribution and the empirical distribution for finite matrices corroborate the usefulness to study the support of the asymptotic eigenvalue distribution of $\boldsymbol{Y}\boldsymbol{Y}^{\dagger}$ and the asymptotic conditions of bulk separability. We remark

that, in principle, the common interference and noise bulk could also separate into two separate bulks for noise and interference. However, this happens only if the weakest interfering signal is sufficiently strong in comparison to the noise power.

A. Unilateral Approximation

The general result for the asymptotic eigenvalue distribution (17) is implicit and not very intuitive. In the following, we develop an approximate analysis for small, but not vanishing loads α. It is based upon the separate calculation of each bulk and subsequent rescaling of the bulks due to pairwise bulk-to-bulk repulsion. We will see that it leads to explicit and intuitive design guidelines.

In the large antenna limit $R = C/\kappa \to \infty$, the singular values of $\boldsymbol{W}/\sqrt{CW}$ follow the Marchenko-Pastur law, i.e.

$$\mathrm{p}_{\boldsymbol{W}}(x) = \frac{\sqrt{\frac{4}{\kappa} - (x - 1 - \frac{1}{\kappa})^2}}{\pi x} \tag{20}$$

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 $\tilde{\boldsymbol{Y}}$ is thus at most

$$TCW \left(1 + \frac{1}{\sqrt{\kappa}}\right)^2 . \tag{21}$$

The total power of the signal of interest at the receiver is $TRCP$ and the signal-to-noise ratio in $\tilde{\boldsymbol{Y}}$ is lower bounded by

$$\mathrm{SNR} \geq \frac{P}{W} \frac{R}{\left(1 + \frac{1}{\sqrt{\kappa}}\right)^2} \geq \frac{P}{W} \cdot \frac{\min\{R, C\}}{4} \tag{22}$$

For fixed normalized coherence time, the first lower bound on the signal-to-noise ratio scales linearly with the number of receive antennas R. For fixed absolute coherence time, second lower bound scales with the minimum of the coherence time and the number of receive antennas. Note, however, that the two lower bounds can be quite loose and the actual SNR might be considerably larger.

In addition to white noise, there is co-channel interference from neighboring cells. The co-channel interference is not white but, like the signal of interest, highly concentrated in certain subspaces. The empirical distribution of the squared singular values of the normalized signal of interest, i.e. $\boldsymbol{HX}/\sqrt{TR}$, is shown in [11] to converge, as $R \to \infty$, to a limit distribution which for $\alpha \ll 1$ is supported in the interval

$$\mathcal{P} = \left[\frac{\kappa P}{\alpha} - 2P\sqrt{\frac{\kappa^2 + \kappa}{\alpha}}; \frac{\kappa P}{\alpha} + 2P\sqrt{\frac{\kappa^2 + \kappa}{\alpha}}\right] . \tag{23}$$

The empirical distribution of the squared singular values of the normalized co-channel interference, i.e. $\boldsymbol{H}_{\mathrm{I}}\boldsymbol{X}_{\mathrm{I}}/\sqrt{TR}$, also converges to a limit distribution. For $\alpha \ll 1$, it is supported in the interval

$$\mathcal{I} = \left[\frac{\kappa I}{\alpha} - 2I\sqrt{L\,\frac{\kappa^2 + \kappa}{\alpha}}\,;\, \frac{\kappa I}{\alpha} + 2I\sqrt{L\,\frac{\kappa^2 + \kappa}{\alpha}} \right] \tag{24}$$

for $I_k = I \,\forall k$. We remark that the condition $I_k = I \,\forall k$ is unrealistic, in practice. However, the general case is not tractable by analytic means. We note, however, that setting all interference powers to the maximum interference power among the users is a worst case scenario.

When separately calculating the eigenvalue spectra of the signal-of-interest, the interference and the noise, the accuracy of the results suffers from the eigenvalues in different bulks repelling each other. In the following, we will correct for this effect up to first order. We decompose one bulk of eigenvalues into single eigenvalues. Then, we introduce correction factors that account for the scaling of one of the single eigenvalues due to the presence of one other bulk of eigenvalues. We will then approximate the influence of several other bulks, e.g. noise bulk and interference bulk, by multiplying the correction factors. This procedure is an approximation, since we neglect the fact that also the scaled bulk of eigenvalues repels the scaling bulk and that the two scaling bulks repel each other.[3]

The presence of additive noise scales the eigenvalues of both the signal of interest and the interference. As shown in Appendix B, the scale factors are given for $R \gg T$ by

$$n_{\mathrm{P}} = \left(1 + \frac{W}{PR}\right)\left(1 + \frac{W}{PC}\right) \tag{25}$$

and

$$n_{\mathrm{I}} = \left(1 + \frac{W}{IR}\right)\left(1 + \frac{W}{IC}\right), \tag{26}$$

respectively. Note that the two scale factors converge to 1 in the large system limit irrespective of the load α, if the noise power W does not scale with the system size.

The presence of interference scales the eigenvalues of the signal of interest and vice versa. As shown in Appendix B, the scale factors for non-overlapping bulks are given for $R \gg T$ by

$$i_{\mathrm{P}} = \left(1 + \frac{L\alpha/\kappa}{\frac{P}{I} - 1}\right)\left(1 + \frac{L\alpha}{\frac{P}{I} - 1}\right) \tag{27}$$

---

[3]To better understand this procedure consider a system with sun (noise bulk), earth (interference bulk) and moon (bulk of signal of interest) which mutually affect each other by gravity. We decompose the moon into single atoms. These atoms are too small to affect the path of the earth. So we can calculate the position of the moon atoms without accounting for the force the moon enacts onto the earth. Then, we apply the same procedure for the interaction between moon atoms and the sun and superimpose the results of the earth-moon and sun-moon interactions. The fact, that sun and earth influence each other is also ignored.

and

$$i_{\mathrm{I}} = \left(1 + \frac{\alpha/\kappa}{\frac{I}{P} - 1}\right)\left(1 + \frac{\alpha}{\frac{I}{P} - 1}\right), \tag{28}$$

respectively. Note, however, that these scale factors are only accurate if $P \gg I$. This somehow limits their usefulness, in practice.

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

$$n_{\mathrm{P}} i_{\mathrm{P}} \mathcal{P} \cap n_{\mathrm{I}} i_{\mathrm{I}} \mathcal{I} = \emptyset \tag{29}$$

or equivalently

$$\frac{P}{I} > \frac{n_{\mathrm{I}} i_{\mathrm{I}}}{n_{\mathrm{P}} i_{\mathrm{P}}} \cdot \frac{1 + 2\sqrt{\alpha L \left(1 + \frac{1}{\kappa}\right)}}{1 - 2\sqrt{\alpha \left(1 + \frac{1}{\kappa}\right)}}, \tag{30}$$

the singular value distribution of the sum of the signal of interest and the interference converges, as $R \to \infty$, to a limit distribution that is composed of two separate non-overlapping bulks [17]. Note that in the limit $\alpha \to 0$, 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.

B. Bilateral Approximation at High SNR

The previous approximation was intuitive, but its accuracy is limited. In this subsection, we use perturbation theory for a more precise approximation for small load $\alpha \ll 1$ and $I_k = I$, where we account for the mutual interaction between the interference bulk and the bulk of the signal of interest.

Let us denote by $\mathcal{P}_W$ and $\mathcal{I}_W$ the eigenvalue bulks corresponding to the signal subspace and the interference subspaces, respectively, when the white noise variance is W. Additionally, let us assume that $P > I$ as in systems of practical interest. Finally, let us define

$$r = \frac{\alpha}{PTC} = \frac{1}{PRC} \tag{31}$$

$$t = \frac{\alpha}{ITC} = \frac{1}{IRC} \tag{32}$$

$$\zeta = WC. \tag{33}$$

The following results are shown in Appendix C: In the high SNR regime, i.e. for $W = 0$, the inverse of the Stieltjes transform is well approximated by the rational function

$$s^{(1)}(G) = \frac{((L+1)(\kappa-2)\alpha - \kappa)G^2 + ((Lr+t)(\kappa-1)\alpha - \kappa(r+t))G - \kappa rt}{G((\kappa + 2(L+1)\alpha)G^2 + ((Lr+t)\alpha + \kappa(r+t))G + \kappa rt)}. \tag{34}$$

The extremes of the function $s^{(1)}(G)$ are the solutions G_1, G_2, G_3, G_4 to the quartic equation

$$(2(L+1)^2(\kappa - 2)\alpha^2 + (L+1)(\kappa - 4)\kappa\alpha - \kappa^2)G^4$$

$$+ (2((2(Lr+t))(L+1)(\kappa - 1)\alpha^2 + ((Lr+t)(\kappa - 1) - (2(L+1))(t+r))\alpha\kappa - (t+r)\kappa^2))G^3$$

$$+ ((Lr+t)^2(\kappa - 1)\alpha^2 + (t^2 + Lr^2)(\kappa - 2)\kappa\alpha - (6(L+1))rt\kappa\alpha - ((t+r)^2 + 2rt)\kappa^2)G^2$$

$$- 2rt\kappa((Lr+t)\alpha + (t+r)\kappa)G - \kappa^2 r^2 t^2 = 0. \quad (35)$$

If G_i, for $i = 1, 2, 3, 4$, are all real with $G_1 < G_2 < G_3 < G_4$ and $s^{(1)}(G_2) < s^{(1)}(G_3)$ then an approximation of $\mathcal{P}_0$ and $\mathcal{I}_0$ is given by

$$\mathcal{P}_0 \approx [s^{(1)}(G_3), s^{(1)}(G_4)] \quad (36)$$

$$\mathcal{I}_0 \approx [s^{(1)}(G_1), s^{(1)}(G_2)]$$

and the two intervals $[s^{(1)}(G_3), s^{(1)}(G_4)]$ and $[s^{(1)}(G_1), s^{(1)}(G_2)]$ are disjoint.

Different approximations of $\mathcal{P}_0$ and $\mathcal{I}_0$ can be obtained by approximating $s(G)$ by the function

$$s_0^{(2)}(G) = \begin{cases} \phi_0(G) + \rho_0(G), & G \in [G_-^\infty, G_+^\infty] \\ \phi_0(G) - \rho_0(G), & \text{elsewhere.} \end{cases} \quad (37)$$

where G_-^∞ and G_+^∞ are the instances of

$$G^\infty = \frac{\kappa(r+t) + \alpha(t+Lr)}{-2\kappa - 4\alpha - 4L\alpha} \pm \frac{\sqrt{\kappa^2(r-t)^2 + 2\alpha\kappa(Lr^2 + t^2 - 3rt - 3Ltr) + \alpha^2(t+Lr)^2}}{-2\kappa - 4\alpha - 4L\alpha} \quad (38)$$

with minus and plus sign, respectively,

$$\phi_0(G) = \frac{(2\alpha(L+1)(\kappa - 1) + \kappa(\kappa - 4))G^2 + \kappa(\alpha(t+Lr) + (\kappa - 2)(t+r))G + \kappa^2 rt}{2G^2((2\kappa + (L+1)\alpha)G + \kappa(t+r))}, \quad (39)$$

$$\rho_0(G) = \kappa \left[\kappa(\kappa - 4\alpha(L+1))G^4 + 2\kappa(\kappa(t+r) - 3\alpha(Lr+t))G^3 + ((t^2 + 4rt + r^2)\kappa^2 \right.$$

$$\left. - 2\alpha\kappa(Lr - t)(r - t) + \alpha^2(t+Lr)^2)G^2 \ 2\kappa rt(\kappa(t+r) + \alpha(t+Lr))G + \kappa^2 t^2 r^2 \right]^{1/2}$$

$$\times \frac{1}{2G^2((2\kappa + (L+1)\alpha)G + \kappa(t+r))}. \quad (40)$$

This approximation of the inverse Stieltjes transform is derived in Appendix C. The extremes of this function cannot be derived in close form. Then, we approximate them by the zeros of $\rho_0(G)$, $G_1^{(2)}$, $G_2^{(2)}$, $G_3^{(2)}$, and $G_4^{(2)}$. If $G_1^{(2)} < G_2^{(2)} < G_3^{(2)} < G_4^{(2)}$ and $s_0^{(2)}(G_2^{(2)}) < s_0^{(2)}(G_3^{(2)})$, we obtain the approximations

$$\mathcal{P}_0 \approx [s_0^{(2)}(G_3^{(2)}), s_0^{(2)}(G_4^{(2)})] = [\phi_0(G_3^{(2)}), \phi_0(G_4^{(2)})] \quad (41)$$

$$\mathcal{I}_0 \approx [s_0^{(2)}(G_1^{(2)}), s_0^{(2)}(G_2^{(2)})] = [\phi_0(G_1^{(2)}), \phi_0(G_2^{(2)})].$$
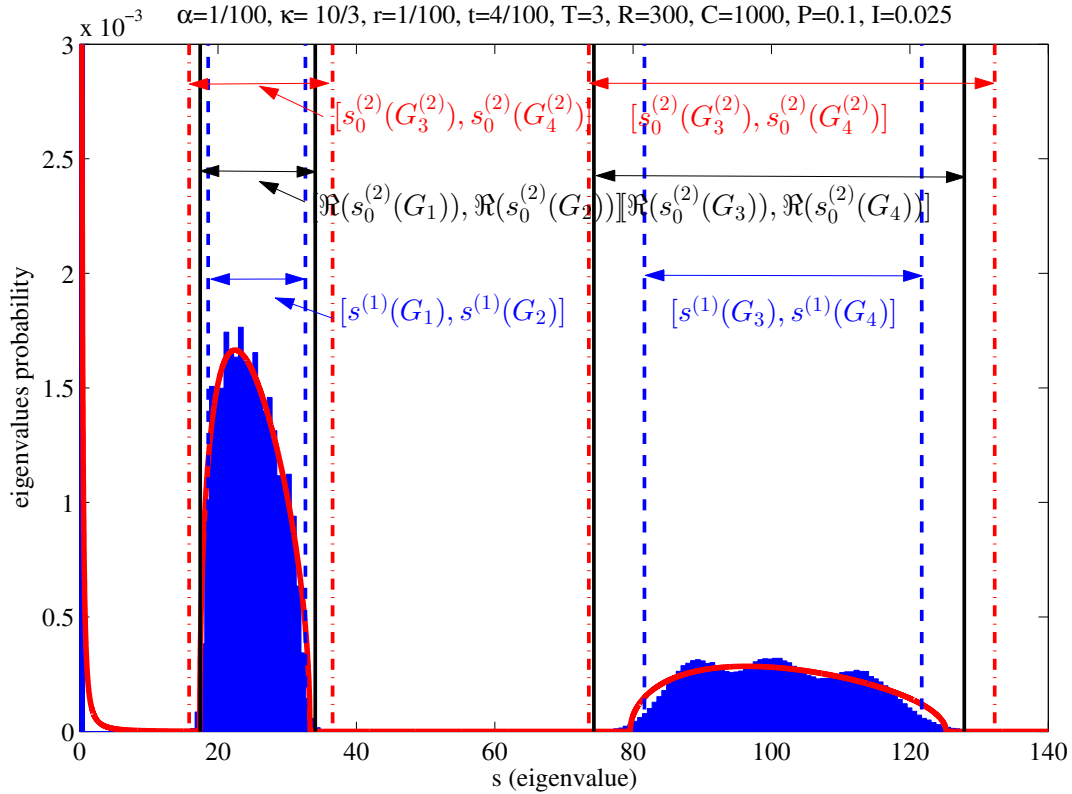
Fig. 2. Bulk-support approximation of the asymptotic eigenvalue density of the matrix $\boldsymbol{YY}^{\dagger}/R$ by (36) and (41) for $\alpha = \frac{1}{100}$, $\kappa = \frac{10}{3}$, $P = -10$ dB, $I_k = \frac{P}{4} \,\forall k$. A further approximation $[\Re(s_0^{(2)}(G_1)), \Re(s_0^{(2)}(G_2))] \bigcup [\Re(s_0^{(2)}(G_3)), \Re(s_0^{(2)}(G_4))]$ is also shown. The histogram in blue is the empirical eigenvalue density for $T = 3$, $R = 300$, and $C = 1000$ while the red line is the asymptotic eigenvalue density.

which are motivated in Appendix C. The approximated intervals in (36) and (41) obtained by application of perturbation theory are a very good approximation of $\mathcal{P}_0$ and $\mathcal{I}_0$ as shown in Figure 2. The approximation obtained by (41) contains the support of the asymptotic eigenvalue distribution.

As well known, the quartic equations to determine G_i and $G_i^{(2)}$, $i = 1, \ldots 4$, admit solutions in closed form. However, they are not insightful and handy because of their complexity. Thus, in the following, we propose looser approximations of the intervals $\mathcal{P}_0$ and $\mathcal{I}_0$ yielding handier conditions on bulk separation. Further approximations yield

$$\mathcal{P}_0 \subset \left[s_{\mathcal{P}}^{(2)}(G_{\mathcal{P}_{\ell}}), \, s_{\mathcal{P}}^{(2)}(G_{\mathcal{P}_u}) \right] \tag{42}$$

where

$$s_{\mathcal{P}}^{(2)}(x) = \frac{2\alpha\kappa(L+1) - 2\alpha(L+1) + 2\kappa(1-\kappa)}{2((1+L)\alpha - \kappa)x + 2\kappa(t-2r)} + \frac{\kappa(\kappa(t-5r) + \alpha(t+Lr) + 4r - 2t)x + \kappa^2 r(t-3r)}{2x^2(((1+L)\alpha - \kappa)x + \kappa(t-2r))} \tag{43}$$

and $G_{\mathcal{P}_\ell}$ and $G_{\mathcal{P}_u}$ equal the instances of

$$G_{\mathcal{P}}^{(2)} = -\kappa r(t-r)\frac{\kappa(t-r) + \alpha(t+(L-2)r) \pm 2\sqrt{\alpha\kappa(t-r)^2 - \alpha^2 r(t+(L-1)r)}}{(\alpha t + \alpha Lr - \kappa t + \kappa r)^2 + 4\alpha\kappa Lr(t-r)} \tag{44}$$

which is obtained by selecting the plus and minus sign, respectively. Similarly, for the bulk associated to the interference subspace,

$$\mathcal{I}_0 \subset \left[s_{\mathcal{I}}^{(2)}(G_{\mathcal{I}_\ell}), \, s_{\mathcal{I}}^{(2)}(G_{\mathcal{I}_u}) \right] \tag{45}$$

where

$$s_{\mathcal{I}}^{(2)}(x) = \frac{2\alpha\kappa(L+1) - 2\kappa(\kappa-1) - 2\alpha(L+1)}{2((\alpha(L+1) - \kappa)x - \kappa(2t-r))} + \frac{\kappa((4-5\kappa)t + (\kappa-2)r + \alpha(t+Lr))x + \kappa^2 t(r-3t)}{2x^2((\alpha(L+1) - \kappa)x - \kappa(2t-r))} \tag{46}$$

and $G_{\mathcal{I}_\ell}$ and $G_{\mathcal{I}_u}$ are obtained by selecting the instance of

$$G_{\mathcal{I}} = -\kappa t(t-r)\frac{\kappa(t-r) + \alpha((2L-1)t - Lr) \pm 2\sqrt{\alpha\kappa L(t-r)^2 + \alpha^2 Lt((L-1)t - Lr)}}{(\alpha t + \alpha Lr - \kappa t + \kappa r)^2 + 4\alpha\kappa Lr(t-r)} \tag{47}$$

with plus and minus sign, respectively. The derivation of the proposed approximations for $\mathcal{P}_0$ and $\mathcal{I}_0$ is detailed in Appendix C.

By enforcing $G_{\mathcal{I}_u} < G_{\mathcal{P}_\ell}$ we obtain a bound on the ratio $\frac{\alpha}{\kappa}$ as a decreasing function of the ratio $\beta = \frac{r}{t} = \frac{I}{P}$

$$\frac{\alpha}{\kappa} \leq \frac{(1-\beta)^2(L\beta^2 + 3(L+1)\beta + 1 - 2(1+\beta)\sqrt{3L\beta})}{(L\beta^2 - 1)(L\beta^2 + 6(L-1)\beta - 1) + (9L^2 - 2L + 9)\beta^2}. \tag{48}$$

Figure 3 shows the region of parameters $\frac{\alpha}{\kappa}$ and β where the bulks of the eigenvalues for the signals of interest and the one for interference do not overlap for various values of L. As expected, the separability region shrinks when the interference from adjacent cells increases, i.e., L increases.

C. Bilateral Approximation for General SNR

An approach similar to the one proposed for the high SNR regime in Appendix C can be applied for $W > 0$ to determine approximated supports of the bulks $\mathcal{P}_W$ and $\mathcal{I}_W$. We propose the conclusive results in the following while the derivations are detailed in Appendix D. Then,

$$\mathcal{P}_W \subset \left[\varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_\ell}), \, \varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_u}) \right] \tag{49}$$
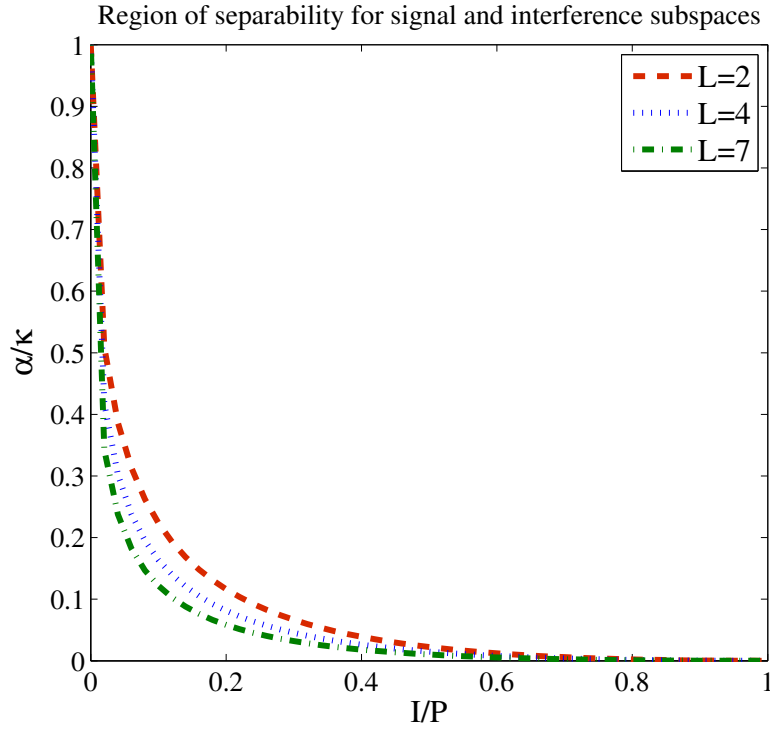
Fig. 3. Separability region obtained by (48) for $L = 2, 4$, and 7.

where

$$\varsigma_{\mathcal{P}}^{(2)}(x) = \frac{\alpha\kappa(L+1) - \alpha(L+1) + \kappa(1-\kappa) + \zeta(\kappa-1)(t-2r)}{((1+L)\alpha - \kappa + \zeta(t-2r))x + \kappa(t-2r)}$$
$$+ \frac{\kappa}{2x^2} \frac{(\kappa(t-5r) + \alpha(t+Lr) + 4r - 2t + \zeta r(t-3r))x + \kappa r(t-3r)}{((1+L)\alpha - \kappa + \zeta(t-2r))x + \kappa(t-2r)} \quad (50)$$

and $\Gamma_{\mathcal{P}_\ell}$ and $\Gamma_{\mathcal{P}_u}$ equal the instances of

$$\Gamma_{\mathcal{P}}^{(2)} = \frac{-\kappa r(t-r)\left[\zeta r(t-r) + \kappa(t-r) + \alpha(t+(L-2)r) \pm 2\sqrt{\alpha\kappa(t-r)^2 - \alpha^2 r(t+(L-1)r)}\right]}{(r(t-r)\zeta + (\alpha-\kappa)t + (\alpha L+\kappa)r)^2 + 4r\zeta((\alpha+\kappa)r^2 - (\alpha+2\kappa)tr + \kappa t^2) + 4\alpha\kappa Lr(t-r)} \quad (51)$$

which are obtained by selecting the plus and minus sign, respectively. Similarly, for the bulk associated to the interference subspace,

$$\mathcal{I}_W \subset \left[\varsigma_{\mathcal{I}}^{(2)}(\Gamma_{\mathcal{I}_\ell}), \varsigma_{\mathcal{I}}^{(2)}(\Gamma_{\mathcal{I}_u})\right] \quad (52)$$

where

$$\varsigma_{\mathcal{I}}^{(2)}(x) = \frac{\alpha\kappa(L+1) - \kappa(\kappa-1) - \alpha(L+1) - (\kappa-1)(2t-r)\zeta}{(\alpha(L+1) - \kappa - \zeta(2t-r))x - \kappa(2t-r)}$$
$$+ \frac{\kappa((4-5\kappa)t + (\kappa-2)r + \alpha(t+Lr) - \kappa\zeta t(3t-r))x + \kappa^2 t(r-3t)}{2x^2((\alpha(L+1) - \kappa - \zeta(2t-r))x - \kappa(2t-r))} \tag{53}$$

and $\Gamma_{\mathcal{I}_\ell}$ and $\Gamma_{\mathcal{I}_u}$ are obtained by selecting the instance of

$$\Gamma_{\mathcal{I}} = \frac{-\kappa t(t-r)\left[\kappa(t-r) + \alpha(2L-1)t - \alpha Lr + t(t-r)\zeta \pm 2\sqrt{\alpha\kappa L(t-r)^2 + \alpha^2 Lt((L-1)t - Lr)}\right]}{(\alpha t + L\alpha r - t\kappa + r\kappa + t(t-r)\zeta)^2 + 4(t-r)(t[(\kappa + \alpha L - \alpha)t - (\alpha L + \kappa)r]\zeta + \alpha\kappa Lr)} \tag{54}$$

with plus and minus sign, respectively.

Interestingly, the separability condition obtained by enforcing $\Gamma_{\mathcal{I}_u} < \Gamma_{\mathcal{P}_\ell}$ yields to condition (48) as in the case of absence of noise. This is not as surprising as it may look at first sight, as it was already observed from (25) and (26) that the noise does not affect the support in the large system limit.

The tightness of the proposed approximation is assessed by numerical simulations. In Figure 4 we consider the same communication system as in Figure 1 but additionally impaired by Gaussian noise with variance equal to 0dB. Besides the histogram of the eigenvalues for a finite system and the asymptotic eigenvalue pdf drawn in solid line, we show the intervals $\left[\varsigma_{\mathcal{I}}^{(2)}(\Gamma_{\mathcal{I}_\ell}), \varsigma_{\mathcal{I}}^{(2)}(\Gamma_{\mathcal{I}_u})\right]$ and $\left[\varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_\ell}), \varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_u})\right]$. The vertical lines indicate the approximation of the boundaries of the asymptotic pdf obtained by perturbation analysis. The approximations based on the second order Taylor expansion $\varsigma_{\mathcal{P}}^{(2)}(x)$ and $\varsigma_{\mathcal{I}}^{(2)}(x)$ include the actual asymptotic support.

V. NUMERICAL RESULTS

In this section, we provide simulation results for the uncoded bit error rate (BER) and compare the proposed SVD-based algorithm, with the conventional linear channel and data estimation scheme considered in [3]. For all cases we set $P/W = 0.1$ (SNR is -10 dB), that is, assume that the system operates in the low SNR region. An identical set of orthogonal pilot sequences of length T is adopted by all the access points to facilitate channel estimation. We consider first the effect of increasing the number of receive antennas while the rest of the parameters are fixed to $T = 5$, $L = 6$, and $C = 100$. As may be observed from Fig. 5, the proposed algorithm (SVD) widely outperforms the receiver based on linear channel estimation in [3] (conventional). Furthermore, it is evident that the proposed algorithm benefits from increased number of receive antennas, irrespective of whether the number of receive antennas is greater or smaller than the coherence time measured in symbol intervals.
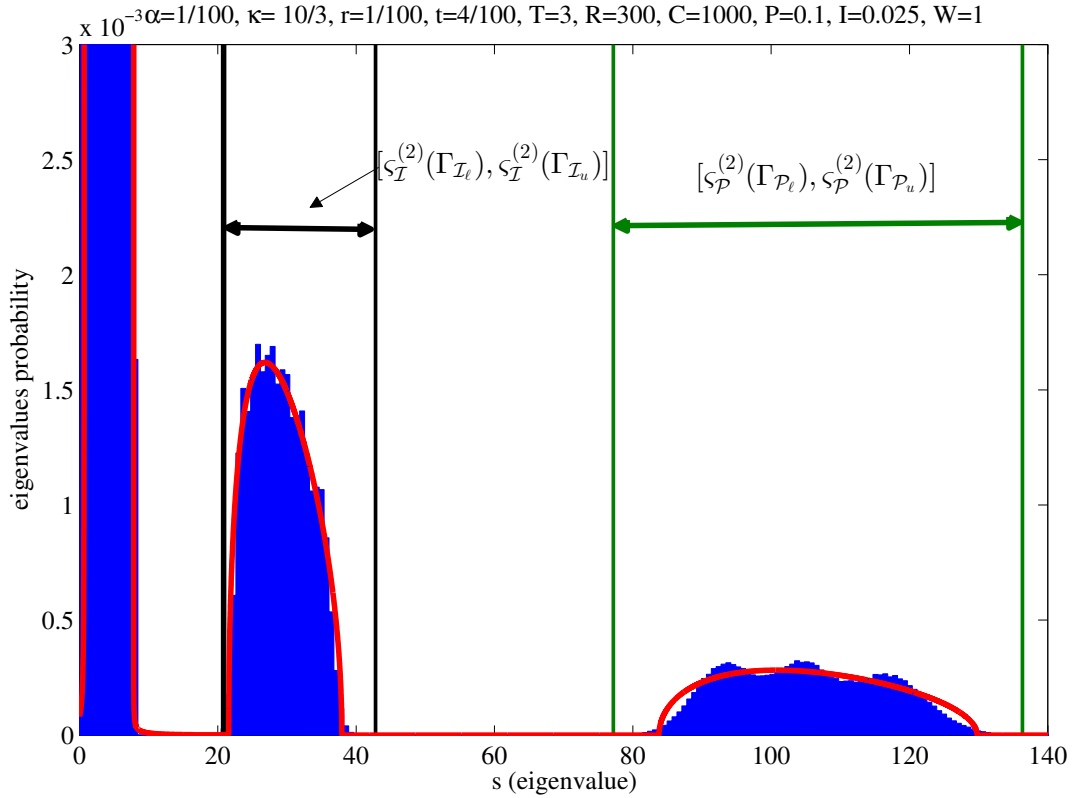
Fig. 4. Eigenvalue pdf of a network with $T = 3$, $L = 2$, $R = 300$, and $W = 0$ with approximation of its support boundaries

The effect of relative interference strength I/P and number of length-T pilot sequences τ is plotted in Fig. 6. For $\tau = 1$ the same orthogonal pilots are used for all access points. In the case $\tau = 5, 10$, random pilot sequences and zero-forcing channel estimation is employed. The RMT thresholds for the given parameters are $I/P = 0.61$ and $I/P = 0.78$ according to (30) and (48), respectively. The proposed algorithm achieves significant performance gains below the RMT thresholds when compared to linear channel estimation. For very strong interference, however, the conventional receiver outperforms the subspace approach. The reason is because we always select only the T strongest eigenvectors for projection, but for finite system sizes and close to the RMT threshold this is suboptimal and we lose a large amount of useful signal while projecting towards interference. This effect can be mitigated by selecting more than T eigenvectors for subspace projection when I/P is expected to be close to the threshold predicted by RMT.
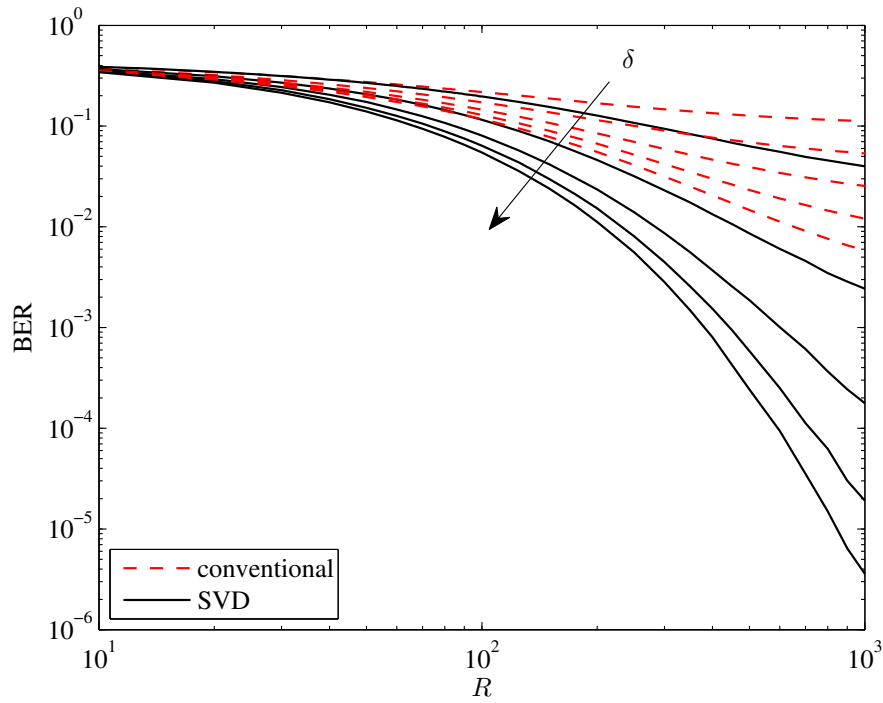
Fig. 5. BER vs. number of receive antennas with $T = 5$, $C = 100$, $L = 6$, $P/W = 0.1$ (SNR is -10 dB), and the interference distribution $I_k = \frac{P(k \bmod T)}{\delta T}, \forall k = 1 \ldots T$ for various values of the parameter δ incrementing from 2 to 6 in direction of arrow.
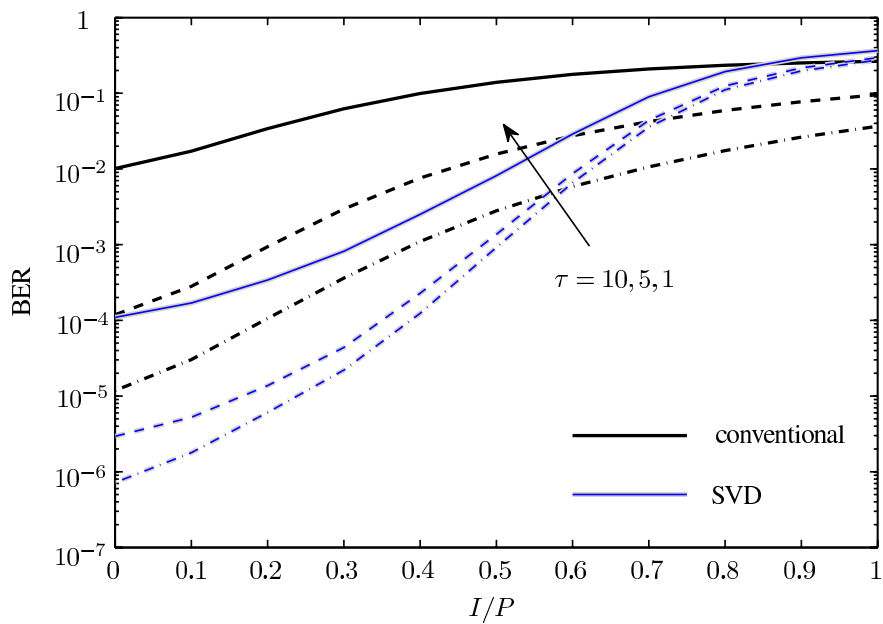

Fig. 6. BER vs. relative interference strength with $T = 3$, $R = 300$, $C = 1000$, $L = 2$, and $P/W = 0.1$ (SNR is -10 dB). The number of length-T pilot blocks is $\tau = 1, 5, 10$.

VI. Summary and Conclusions

We proposed a practical algorithm with polynomial complexity to mitigate pilot contamination in cellular systems with power controlled handoff. The dominant complexity of this algorithm is a singular value decomposition of the received signal block. The algorithm was analyzed by means of random matrix theory. It was found that the algorithm works well, if certain constraints relating the number of antennas to the coherence time and the signal-to-interference ratio are fulfilled. Simulations show that under that conditions, the algorithm significantly outperforms linear channel estimation.

This paper has focussed solely on the reverse link channel. For the forward link channel, one can exploit channel reciprocity in time-division duplex systems. Similar to the reverse link channel, knowledge of the full channel matrix is not required. Basic considerations of linear algebra show that it is sufficient to know the subspace which the channel vectors of interest span in order to solely require accurate channel estimates for the projected channel (13).

Appendix A

Eigenvalue Distribution

Consider the random matrix

$$\boldsymbol{D} = \sum_{k=-1}^{K} a_k \boldsymbol{B}_k \boldsymbol{C}_k \tag{55}$$

with $a_k \in \mathbb{R}$, $\boldsymbol{B}_k \in \mathbb{C}^{n \times m_k}$ and $\boldsymbol{C}_k \in \mathbb{C}^{m_k \times n}$ being random matrices with iid. zero-mean entries with variance $1/m_k$ and $1/n$, respectively. From [11, Eq. (35)], we have

$$\frac{1}{\mathrm{G}_{\boldsymbol{DD}^\dagger}(s)} = -s - \sum_{k=-1}^{K} \frac{a_k^2 \rho_k s \, \mathrm{G}_{\boldsymbol{DD}^\dagger}(s)}{\rho_k - a_k^2 s \, \mathrm{G}_{\boldsymbol{DD}^\dagger}^2(s)} \tag{56}$$

with $\mathrm{G}_{\boldsymbol{DD}^\dagger}(s)$ denoting the Stieltjes transform of the asymptotic eigenvalue distribution of $\boldsymbol{DD}^\dagger$.

Next we will distinguish two cases: $\beta \geq 1$ and $\beta \leq 1$. For $\beta \leq 1$, define the matrix $\boldsymbol{E} \in \mathbb{C}^{\beta n \times n}$ by the decomposition

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{E} \\ \boldsymbol{F} \end{bmatrix}. \tag{57}$$

From [18, Theorem 14.10], we have

$$\mathrm{R}_{\boldsymbol{EE}^\dagger}(w) = \mathrm{R}_{\boldsymbol{DD}^\dagger}(\beta w). \tag{58}$$

For $\beta \geq 1$, define the matrix $\boldsymbol{E} \in \mathbb{C}^{\beta n \times n}$ as

$$
\boldsymbol{E} = \begin{bmatrix} \boldsymbol{D} \\ \boldsymbol{G}_0 \boldsymbol{P} \\ \boldsymbol{G}_1 \\ \vdots \\ \boldsymbol{G}_{\lfloor \beta-1 \rfloor} \end{bmatrix}.
\tag{59}
$$

with the family $(\{\boldsymbol{D}\}, \{\boldsymbol{G}_0\}, \ldots, \{\boldsymbol{G}_{\lfloor \beta-1 \rfloor}\}, \{\boldsymbol{P}\})$ being asymptotically free, $\boldsymbol{G}_i \in \mathbb{C}^{n \times n}$ and $\mathrm{R}_{\boldsymbol{G}_i \boldsymbol{G}_i^\dagger}(w) = \mathrm{R}_{\boldsymbol{D}\boldsymbol{D}^\dagger}(w)$ for all i, and $\boldsymbol{P} \in \{0,1\}^{n \times (\beta - \lfloor \beta \rfloor)}$ be diagonal with full rank. From [18, Theorem 14.10], we have

$$
\mathrm{R}_{\boldsymbol{D}\boldsymbol{D}^\dagger}(w) = \mathrm{R}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(w/\beta).
\tag{60}
$$

Note that (60) is equivalent to (58). Thus, we can unify the two cases and continue with (58) for any β.

In the Stieltjes domain, (58) translates into

$$
\beta \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(s) = \mathrm{G}_{\boldsymbol{D}\boldsymbol{D}^\dagger}\left(s + \frac{\beta - 1}{\beta \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(s)} \right).
\tag{61}
$$

Together with (56), we find an expression similar to [11, Eq. (39)] which simplifies to

$$
s\mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(s) = -1 - \sum_{k=-1}^{K} \frac{a_k^2 \rho_k \left(s + \frac{\beta-1}{\beta \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(s)} \right) \beta \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}^2(s)}{\rho_k - a_k^2 \left(s + \frac{\beta-1}{\beta \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}(s)} \right) \beta^2 \mathrm{G}_{\boldsymbol{E}\boldsymbol{E}^\dagger}^2(s)}.
\tag{62}
$$

Now, we consider the matrix $\boldsymbol{Y}$ in (16) as a special case of $\boldsymbol{E}$. This implies

$$
K = LT
\tag{63}
$$

$$
\beta = \frac{R}{C} = \frac{1}{\kappa}
\tag{64}
$$

$$
\rho_{-1} = \frac{T}{C} = \frac{\alpha}{\kappa}
\tag{65}
$$

$$
a_{-1}^2 = PTC
\tag{66}
$$

$$
\rho_0 \to \infty
\tag{67}
$$

$$
a_0^2 = WC
\tag{68}
$$

$$
\rho_k = \frac{1}{C} \qquad \forall k > 0
\tag{69}
$$

$$
a_k^2 = I_k C \qquad \forall k > 0
\tag{70}
$$

and (17) is obtained in the limit $K \to \infty$. Note that the entries of $\boldsymbol{B}_0 \boldsymbol{C}_0$ become iid. as $\rho_0 \to \infty$.

Appendix B

Eigenvalue Repulsion

In order to find the support of the asymptotic eigenvalue distributions we follow [19, Eq. (48)]. There it is shown that the boundaries of the support of the asymptotic eigenvalue distribution are extrema of the inverse of the Stieltjes transform. For a particular example different from ours, the procedure is explained in greater detail in [20, Chapter 7].

Consider the random matrix

$$\boldsymbol{E} = \left(\sqrt{P}\boldsymbol{A}\boldsymbol{B} + \sqrt{I}\boldsymbol{C}\boldsymbol{D}\right)\left(\sqrt{P}\boldsymbol{A}\boldsymbol{B} + \sqrt{I}\boldsymbol{C}\boldsymbol{D}\right)^{\dagger} \tag{71}$$

where $\boldsymbol{A} \in \mathbb{C}^{R \times \alpha R}$, $\boldsymbol{B} \in \mathbb{C}^{\alpha R \times \kappa R}$, $\boldsymbol{C} \in \mathbb{C}^{R \times \beta R}$, and $\boldsymbol{D} \in \mathbb{C}^{\beta R \times \kappa R}$ with iid. zero-mean unit-variance entries. We have from (17) that the limiting Stieltjes transform of $\boldsymbol{E}\boldsymbol{E}^{\dagger}$ obeys

$$s_\alpha G + 1 + \frac{\alpha(s_\alpha G + 1 - \kappa)G}{r\kappa - (s_\alpha G + 1 - \kappa)G} + \frac{\beta(s_\alpha G + 1 - \kappa)G}{t\kappa - (s_\alpha G + 1 - \kappa)G} = 0 \tag{72}$$

with

$$r = \frac{\alpha}{PTC} = \frac{1}{PRC} \tag{73}$$

$$t = \frac{\alpha}{ITC} = \frac{1}{IRC}. \tag{74}$$

Solving for s_α leads to a cubic equation and is a tedious task. Nevertheless, Maple 16 can do it symbolically. For $\alpha = 0$, we get

$$s_0 = \frac{G\kappa - 2G + G\beta + t\kappa}{2G^2} - \frac{\sqrt{\beta^2 G^2 + 2\beta Gt\kappa - 2\beta G^2\kappa + \kappa^2(G+t)^2}}{2G^2}. \tag{75}$$

At the interval boundaries,

$$\frac{\partial s_\alpha}{\partial G} = \frac{Z(G)}{N(G)} \tag{76}$$

with obvious definition of the enumerator $Z(G)$ and the denominator $N(G)$, must vanish. For $\alpha \to 0$, one of the bulks will disappear. Thus, $N(G)$ and $Z(G)$ will have a common zero in the limit $\alpha \to 0$. This common zero corresponds to the position of the vanishing bulk. Instead of inspecting the zeros of $Z(G)$ when searching for the interval boundary of the vanishing bulk, we can also look at the zeros of $N(G)$[4]. We find that

$$\lim_{\alpha \to 0} N(G) = 0 \tag{77}$$

---

[4]This procedure is necessary since $Z(G)$ fills many pages even in the limit of $\alpha \to 0$ and finding its zeros is intractable. However, $\lim_{\alpha \to 0} N(G)$ only fills several lines and Maple can find its zeros in closed form.

has the following four solutions

$$G_1 = \frac{-\kappa t}{(\sqrt{\kappa} + \sqrt{\beta})^2} \tag{78}$$

$$G_2 = \frac{-\kappa t}{(\sqrt{\kappa} - \sqrt{\beta})^2} \tag{79}$$

$$G_3 = 0 \tag{80}$$

$$G_4 = \frac{r\kappa(t - r)}{\kappa(r - t) - \beta r}. \tag{81}$$

Obviously, G_4 is the desired zero, since the other zeros do not depend on r. Plugging into (75) gives

$$s_0(G_4) = \frac{(t - r + \frac{\beta r}{\kappa})(t - r + \beta r)}{r(t - r)^2}. \tag{82}$$

Thus, the presence of interference scales the signal of interest by a factor of

$$\frac{s_0(G_4)}{s_0(G_4)\big|_{\beta=0}} = \left(1 + \frac{\beta/\kappa}{t/r - 1}\right)\left(1 + \frac{\beta}{t/r - 1}\right). \tag{83}$$

The scale factor of the interference is obtained by exchanging the role of signal and interference.

In order to obtain the scale factor for the white noise, we note that for infinite load the interference becomes white. Thus, we take the limit $t, \beta \to \infty$ with $\zeta = \frac{\beta}{t}$ in (83) and obtain

$$s_0^\infty(G_4) = \left(\zeta + \frac{1}{r}\right)\left(1 + \frac{r\zeta}{\kappa}\right). \tag{84}$$

Without noise, i.e. $\zeta = 0$, the signal of interest would be positioned at $1/r$. Thus, the presence of noise scales the signals by a factor of

$$\frac{s_0^\infty(G_4)}{s_0^\infty(G_4)\big|_{\zeta=0}} = (1 + r\zeta)\left(1 + \frac{r\zeta}{\kappa}\right). \tag{85}$$

APPENDIX C

THE NOISELESS SYSTEM

In this section we analyze the behaviour of the noiseless system when the number of interfering signals and signals of interest are proportional and very small compared to the number of receive antennas but not vanishing, i.e. $\alpha \to 0$. We still consider the random matrix in (71) but both the dimensions of the interference and signal subspace grow proportionally, i.e., $\beta = \alpha L$. Under these assumptions, (72) can be written as

$$\frac{\alpha G(sG + 1 - \kappa)(\kappa(t + Lr) - (L + 1)(sG + 1 - \kappa)G)}{(r\kappa - (sG + 1 - \kappa)G)(t\kappa - (sG + 1 - \kappa)G)} +$$
$$\frac{(sG + 1)(t\kappa - (sG + 1 - \kappa)G)(r\kappa - (sG + 1 - \kappa)G)}{(r\kappa - (sG + 1 - \kappa)G)(t\kappa - (sG + 1 - \kappa)G)} = 0. \tag{86}$$

By simple inspection, we observe that the numerator $N(s)$ of the l.h.s. in (86) is a function obtained by perturbation of a cubic function in s

$$N_0(s) = (sG + 1)(t\kappa - (sG + 1 - \kappa)G)(r\kappa - (sG + 1 - \kappa)G) \tag{87}$$

by a quadratic function in s proportional to α

$$N_p(s) = G(sG + 1 - \kappa)(\kappa(t + Lr) - (L + 1)(sG + 1 - \kappa)G). \tag{88}$$

Then, for small α the zeros of the original numerator $N(s) = N_0(s) + \alpha N_p(0)$ can be computed as a perturbed version of the zeros in $N_0(s)$ given by

$$s_{0,0} = -\frac{1}{G} \tag{89}$$

$$s_{0,\mathcal{P}} = -\frac{(1 - \kappa)G - \kappa r}{G^2} \tag{90}$$

$$s_{0,\mathcal{I}} = -\frac{(1 - \kappa)G - \kappa t}{G^2}. \tag{91}$$

Let us observe that (89) corresponds to the Stieltjes transform of a pdf $p(x) = \delta(x)$, i.e. the eigenvalue distribution of a matrix with all zero eigenvalues and we are interested in its perturbed version by the signal and interference subspaces. Then, we focus on the perturbation of this function to determine the inverse Stieltjes transform. This initial observation will avoid further discussions on the selection of the multiple zeros of $N(s)$. Then, a first order Taylor expansion of $N(s)$ in $s_{0,0}$

$$N(s) \approx N_p(s_{0,0}) + \left.\frac{\partial N(s)}{\partial s}\right|_{s=s_{0,0}} (s - s_{0,0}) \tag{92}$$

yields a linear equation in s to determine the approximation of the inverse Stieltjes transform $s^{(1)}(G)$

$$s^{(1)}(G) = s_{0,0} + \frac{N_p(s_{0,0})}{\left.\frac{\partial N(s)}{\partial s}\right|_{s=s_{0,0}}}$$

presented in (34). Note that $s^{(1)}(G)$ maintains the pole in $G = 0$ as the Stieltjes transform of $p(x) = \delta(x)$ but also presents two additional poles in (38) as effect of the perturbation. In Figure 7 we show the exact inverse Stieltjes transform in solid blue lines and compare it with $s^{(1)}(x)$, the approximation via perturbation theory, and $s = -\frac{1}{G}$. In Figure (7), the gaps of the solid blue lines correspond to regions where $s(G)$ assumes complex conjugate values for real values of G. The extremes of the function $s(G)$ determine the support $\mathcal{P}_0 \cup \mathcal{I}_0$ of the asymptotic eigenvalue distribution of $YY^\dagger$ while the extremes of $s^{(1)}(G)$ are related to the estimation $[s^{(1)}(G_1), s^{(1)}(G_2)] \cup [s^{(1)}(G_3), s^{(1)}(G_4)]$. The presence of poles in $s^{(1)}(G)$ is an artefact of the first order Taylor expansion of the polynomial $N(s)$ and corresponds to the region where the $N(s)$ has two complex conjugate solutions.
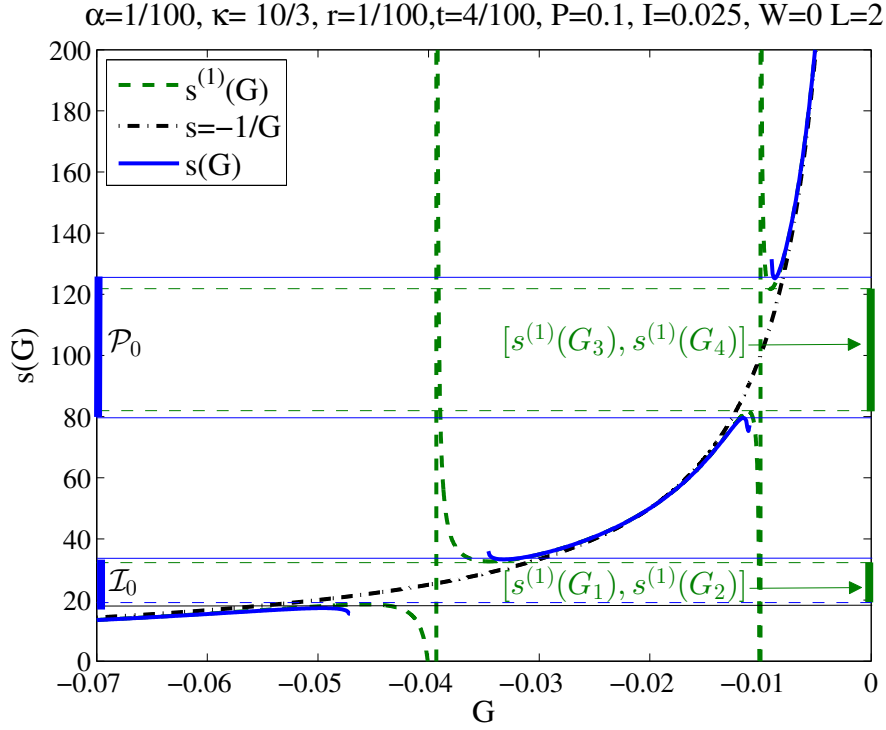
Fig. 7. Analysis of the inverse Stieltjes transform $s(G)$, its version without perturbation, i.e. $s = -\frac{1}{G}$, and its approximation via perturbation theory $s^{(1)}(G)$. Eigenvalue pdf support of a noiseless system with $L = 2, \alpha = \frac{1}{100}, \kappa = \frac{10}{4}, P = 0.1$ and $I = \frac{P}{4}$ and its approximation by the estimation $[s^{(1)}(G_1), s^{(1)}(G_2)] \bigcup [s^{(1)}(G_3), s^{(1)}(G_4)]$.

In order to improve the approximation of the zeros of $N(s)$ in the intervals where they are complex, we consider a second order Taylor expansion of $N(s)$ around $s_{0,0}$ and we obtain the quadratic function in s :

$$N_0^{(2)}(s) = \left((-2\kappa - \alpha(L+1))G^4 - \kappa(t+r)G^3\right)s^2 + \left((\kappa^2 + \kappa(2(1+L)\alpha - 4) - 2(1+L)\alpha)G^3\right.$$

$$+ ((r+t)\kappa^2 + ((t+Lr)\alpha - 2t - 2r)\kappa)G^2 + \kappa^2 rtG)s + (\kappa^2(1 - \alpha(L+1)) + 2\kappa(\alpha(L+1) - 1)$$

$$- \alpha(L+1))G^2 + (\kappa^2(r+t - (Lr+t)\alpha) + (\alpha(t+Lr) - t - r)\kappa)G + \kappa^2 rt \qquad (93)$$

which is a polynomial in s with two zeros

$$\widehat{s}_0^{(2)}(G) = \phi_0 \pm \rho_0(G)$$

where $\rho_0(G)$ and $\phi_0(G)$ are defined in (40) and (39), respectively. The inverse of the Stieltjes transform, selected as perturbation of $s = -\frac{1}{G}$, is $s_0^{(2)}(G)$ as defined in (37). Note that $\phi_0(G) - \rho_0(G)$ cannot be the desired inverse in the interval $[G_-^\infty, G_+^\infty]$ since it presents a pole in $G = -\frac{(t+r)\kappa}{2\kappa + \alpha(L+1)} \in [G_-^\infty, G_+^\infty]$
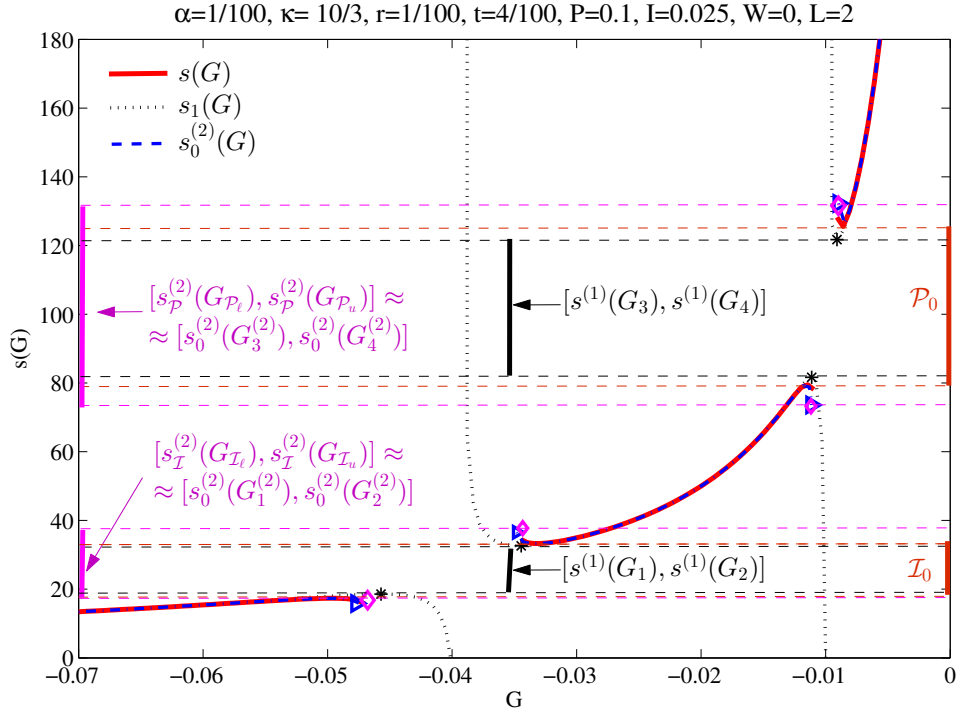
Fig. 8. Comparison of the inverse Stieltjes transform $s(G)$ with its approximations $s^{(1)}(G)$ and $s_0^{(2)}(G)$ for the case of a noiseless system with $L = 2, \alpha = \frac{1}{100}, \kappa = \frac{10}{4}, P = 0.1$ and $I = \frac{P}{4}$. The star markers show the extremes of the function $s^{(1)}(G)$. The triangle markers show the points where $s_0^{(2)}(G)$ becomes complex for real values of G. The diamond markers show the points where $s_{\mathcal{P}}^{(2)}(G)$ and $s_{\mathcal{I}}^{(2)}(G)$ become complex for real values of G, i.e. in ascending order $G_{\mathcal{I}_\ell}$, $G_{\mathcal{I}_u}$, $G_{\mathcal{P}_\ell}$, and $G_{\mathcal{P}_u}$.

while $\phi_0(G) + \rho_0(G)$ does not. However, $\phi_0(G) + \rho_0(G)$ is not the desired inverse outside the interval $[G_-^\infty, G_+^\infty]$ since it behaves like G^{-2} in a surrounding of $G = 0$ and like G^{-1} for $G \to \pm\infty$.

In contrast to the analogous problem with the first order Taylor approximation $s^{(1)}(G)$, the computation of the extremes of the function $s_0^{(2)}(G)$ do not have a closed form solution since their computation implies the solution of two polynomial equations of degree seven. In order to acquire deeper insight on the problem, let us observe the behaviour of $s_0^{(2)}(G)$ shown in Figure 8. The match between $s(G)$ and $s_0^{(2)}(G)$ is nearly perfect in the surroundings of the extremes. Figure 8 suggests to approximate the extremes of $s_0^{(2)}(G)$ by the points where $s_0^{(2)}(G)$ becomes complex, i.e. the zeros of $\rho_0(G)$. This approximation implies again the solution of a polynomial equation of degree four yielding the zeros $G_1^{(2)}$, $G_2^{(2)}$, $G_3^{(2)}$, and $G_4^{(2)}$ with $G_1^{(2)} \leq G_2^{(2)} \leq G_3^{(2)} \leq G_4^{(2)}$. Although the zeros $G_1^{(2)}$, $G_2^{(2)}$, $G_3^{(2)}$, and $G_4^{(2)}$ can be expressed in closed form, their expression is too cumbersome to be insightful. In order to obtain

more practical and useful results we consider the second order Taylor expansion of $N(s)$ in $s_{0,\mathcal{P}}$ and $s_{0,\mathcal{I}}$, which yields

$$
\begin{aligned}
N(s) \approx &\left((\kappa - \alpha(L+1))G^4 + \kappa(2r-t)G^3 \right) s^2 + \Big((2\kappa(\alpha(L+1)+1-\kappa) - 2\alpha(L+1))G^3 \\
&+ ((t-5r)\kappa^2 + \kappa(4r-2t+\alpha t+\alpha Lr))G^2 + \kappa^2 r(t-3r)G\Big)s + \big(\kappa^2(-2-L\alpha-\alpha+\kappa) \\
&- \alpha(L+1) + \kappa(2\alpha(L+1)+\kappa))G^2 + (\kappa^2(3\kappa r - \alpha(Lr+t)+t-5r) + \kappa(2-t) \\
&+ \alpha\kappa(t+Lr))G + 3\kappa^2 r(\kappa r - r + \kappa^2 t + \frac{\kappa r^2}{G})
\end{aligned}
\tag{94}
$$

$$
\begin{aligned}
N(s) \approx &\left((\kappa - \alpha(L+1))G^4 + \kappa(2t-r)G^3 \right) s^2 + \Big((-2\kappa^2 + 2(1+(1+L)\alpha)\kappa - 2\alpha(L+1))G^3 \\
&+ ((r-5t)\kappa^2 + ((t+Lr)\alpha - 2r + 4t)\kappa)G^2 + \kappa^2(r-3t)tG\Big)s + \big(\kappa^3 - (2+(L+1)\alpha)\kappa^2 \\
&+ (1+2(1+L)\alpha)\kappa - (L+1)\alpha)G^2 + (3\kappa^3 t + (r-(Lr+t)\alpha - 5t)\kappa^2 + ((t+Lr)\alpha \\
&- r + 2t)\kappa)G + 3\kappa^3 t^2 + t(r-3t)\kappa^2 + \frac{t^3\kappa^3}{G}.
\end{aligned}
\tag{95}
$$

The zeros of (94) and (95) are relatively good approximations of the actual inverse Stieltjes transform $s(G)$ in the surrounding of the poles (38). Let us denote them $\widehat{s}_{\mathcal{P}}^{(2)}(G)$ and $\widehat{s}_{\mathcal{I}}^{(2)}(G)$, respectively. By using again as approximation for the extremes of $\widehat{s}_{\mathcal{P}}^{(2)}(G)$ and $\widehat{s}_{\mathcal{I}}^{(2)}(G)$, the values of G where $\widehat{s}_{\mathcal{P}}^{(2)}(G)$ and $\widehat{s}_{\mathcal{I}}^{(2)}(G)$ become complex, i.e. the points where the discriminants of (94) and (95) vanish, we can obtain simpler approximations of the extremes. In fact, the discriminants of (94) and (95) are again quartic polynomials in G but with two zeros in $G = 0$. The other two zeros can be easily computed and are given by (44) for $\widehat{s}_{\mathcal{P}}^{(2)}(G)$ and by (47) for $\widehat{s}_{\mathcal{I}}^{(2)}(G)$. Then, observing that the irrational components of $\widehat{s}_{\mathcal{P}}^{(2)}(G)$ and $\widehat{s}_{\mathcal{I}}^{(2)}(G)$ vanish in $G = G_{\mathcal{P}}^{(2)}$ and $G = G_{\mathcal{I}}^{(2)}$, respectively, $\widehat{s}_{\mathcal{P}}^{(2)}(G_{\mathcal{P}}^{(2)}) = s_{\mathcal{P}}^{(2)}(G_{\mathcal{P}}^{(2)})$ and $\widehat{s}_{\mathcal{I}}^{(2)}(G_{\mathcal{I}}^{(2)}) = s_{\mathcal{I}}^{(2)}(G_{\mathcal{I}}^{(2)})$ with $s_{\mathcal{P}}^{(2)}(G)$ and $s_{\mathcal{I}}^{(2)}(G)$ defined in (43) and (46), respectively. The observation that the instances of $G_{\mathcal{P}}$ and $G_{\mathcal{I}}$ with sign plus are not greater than the corresponding instances with sign minus, i.e. $G_{\mathcal{P}_\ell}^{(2)} \leq G_{\mathcal{P}_u}^{(2)}$ and $G_{\mathcal{I}_\ell}^{(2)} \leq G_{\mathcal{I}_u}^{(2)}$, yields the approximations (42) and (45).

By appealing to the previous results, in the following we derive condition (48) for bulk separability.

Under the assumptions of physical interest that $L \in \mathbb{N}^+$ and $t \geq r \geq 0$, $G_{\mathcal{P}}^{(2)}$ and $G_{\mathcal{I}}^{(2)}$ are all negative real zeros[5] if

$$
\frac{t}{r} \geq \max\left(1 + \frac{\alpha}{2\kappa} + \frac{\alpha}{2\kappa}\sqrt{1 + \frac{4L\kappa}{\alpha}}, \frac{1 + \frac{L\alpha}{2\kappa} + \frac{L\alpha}{2\kappa}\sqrt{1 + \frac{4\kappa}{L^2\alpha}}}{1 + (L-1)\frac{\alpha}{\kappa}} \right) = 1 + \frac{\alpha}{2\kappa} + \frac{\alpha}{2\kappa}\sqrt{1 + \frac{4L\kappa}{\alpha}}
\tag{96}
$$

or equivalently

$$
0 \leq \frac{\alpha}{\kappa} \leq \frac{(t-r)^2}{r(t+(L-1)r)}.
\tag{97}
$$

---

[5]These conditions are obtained by enforcing that the arguments of the square roots in (94) and (95) are nonnegative.

By simple inspection, it is easy to verify that $G_{\mathcal{I}_\ell}^{(2)} \leq G_{\mathcal{P}_u}^{(2)}$ under the above mentioned conditions of physical interest. However, it is interesting to determine under which conditions the two intervals $[G_{\mathcal{P}_\ell}^{(2)}, G_{\mathcal{P}_u}^{(2)}]$ and $[G_{\mathcal{I}_\ell}^{(2)}, G_{\mathcal{I}_u}^{(2)}]$ do not intersect, i.e. when $G_{\mathcal{I}_u}^{(2)} \leq G_{\mathcal{P}_\ell}^{(2)}$. It can be verified, for example using Maple, that this last condition is satisfied if $\frac{\alpha}{\beta}$ and $\beta = \frac{r}{t}$ satisfy (48). Additionally, condition (48) implies also (97).

APPENDIX D

THE NOISY SYSTEM

The analysis of the system with noise follows along lines similar to the ones adopted in the previous section. The fixed point equation for the Stieltjes transform of the eigenvalue pdf in (17) can be rewritten as

$$sG + 1 + \frac{\zeta(sG + 1 - \kappa)G}{\kappa} + \frac{\alpha(sG + 1 - \kappa)G}{r\kappa - (sG + 1 - \kappa)G} + \frac{\alpha L(sG + 1 - \kappa)G}{t\kappa - (sG + 1 - \kappa)G} = 0 \qquad (98)$$

with r, t and ζ defined in (31), (32), and (33).

The inverse function $s(G)$ can be obtained as a zero of the numerator of (98). As in the previous section, this is a cubic function in s obtained as perturbation of the cubic function

$$N_0(s) = \big(G(\kappa + \zeta G)s + \zeta(1 - \kappa)G + \kappa\big)\big(r\kappa - (sG + 1 - \kappa)G\big)\big(t\kappa - (sG + 1 - \kappa)G\big) \qquad (99)$$

by a quadratic function

$$N_p(s) = \kappa G(sG + 1 - \kappa)(\kappa(t + Lr) - (L + 1)(sG + 1 - \kappa)G). \qquad (100)$$

Simple inspection of (99) and (100) shows that the introduction of noise has the only effect of modifying $s_{0,0}$ in (89) into

$$\widetilde{s}_{0,0} = -\frac{\zeta(1 - \kappa)G + \kappa}{(\kappa + \zeta G)G} \qquad (101)$$

while, up to a scaling factor κ, it leaves unchanged the perturbation $N_p(s)$. The first order Taylor expansion of $N(s)$ in $\widetilde{s}_{0,0}$ yields to an approximation of the inverse Stieltjes transform whose extremes computation requires the solution of a polynomial of degree six and it is not feasible in closed form. Thus, we do not discuss further this case. On the contrary, for the second order expansion, all the results obtained for the noiseless system can be extended. The second order expansion of $N(s)$ in $\widetilde{s}_{0,0}$ yields a polynomial whose discriminant is again a quartic equation in G. Similarly to the noiseless case, to obtain approximations of practical use we consider the second order expansions of $N(s)$ in $s_{0,\mathcal{P}}$ and $s_{0,\mathcal{I}}$ and we approximate the extremes of the inverse Stieltjes transform by the zeros of the corresponding discriminants. The expansion

in $s_{0,\mathcal{P}}$ yields the approximations $\Gamma_{\mathcal{P}_\ell}^{(2)}$ and $\Gamma_{\mathcal{P}_u}^{(2)}$ in (51) for the extremes of the inverse Stieltjes transform and the approximation of the inverse Stieltjes transform boils down to $\varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_u}^{(2)})$ and $\varsigma_{\mathcal{P}}^{(2)}(\Gamma_{\mathcal{P}_\ell}^{(2)})$, with $\varsigma_{\mathcal{P}}^{(2)}(G)$ defined in (50), when evaluated in $\Gamma_{\mathcal{P}_\ell}^{(2)}$ and $\Gamma_{\mathcal{P}_u}^{(2)}$. Similar considerations hold for approximation based on the second order expansion in $s_{0,\mathcal{I}}$.

References

[1] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Fortieth Asilomar Conf. on Signals, Systems, & Computers*, Pacific Grove, CA, USA, Oct. 2006.

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

[3] 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.

[4] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "The multicell multiuser MIMO uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2350–2361, Jun. 2013.

[6] 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.

[7] 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.

[8] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Cellular systems with many antennas: Large system analysis under pilot contamination," in *Proc. of 50th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Oct. 2012.

[9] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[10] 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.

[11] R. R. Müller, M. Vehkaperä, and L. Cottatellucci, "Blind pilot decontamination," in *Proc. of ITG Workshop on Smart Antennas*, Stuttgart, Germany, Mar. 2013.

[12] L. Cottatellucci, R. R. Müller, and M. Vehkaperä, "Analysis of pilot decontamination based on power control," in *Proc. of IEEE Vehicular Technology Conference (VTC)*, Dresden, Germany, Jun. 2013.

[13] T. S. Rappaport, *Wireless Communications*. New Jersey: Prentice–Hall, 1996.

[14] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in manhattan," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 321–331, Apr. 2003.

[15] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2049–2069, Oct. 1998.

[16] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using theor sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.

[17] 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.

[18] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*. Cambridge, U.K.: Cambridge University Press, 2006.

[19] A. Zee, "Law of addition in random matrix theory," *Nuclear Physics*, pp. 726–744, 1996.

[20] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.