Abstract—Linear precoding has been widely studied in the context of Massive multiple-input-multiple-output (MIMO) together with two common power normalization techniques, namely, matrix normalization (MN) and vector normalization (VN). Despite this, their effect on the performance of Massive MIMO systems has not been thoroughly studied yet. The aim of this paper is to fulfill this gap by using large system analysis. Considering a system model that accounts for channel estimation, pilot contamination, arbitrary pathloss, and per-user channel correlation, we compute tight approximations for the signal-to-interference-plus-noise ratio and the rate of each user equipment in the system while employing maximum ratio transmission (MRT), zero forcing (ZF), and regularized ZF precoding under both MN and VN techniques. Such approximations are used to analytically reveal how the choice of power normalization affects the performance of MRT and ZF under uncorrelated fading channels. It turns out that ZF with VN resembles a sum rate maximizer while it provides a notion of fairness under MN. Numerical results are used to validate the accuracy of the asymptotic analysis and to show that in Massive MIMO, non-coherent interference and noise, rather than pilot contamination, are often the major limiting factors of the considered precoding schemes.

Index Terms—Massive MIMO, linear precoding, power normalization techniques, large system analysis, pilot contamination.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is a multi-user MIMO system that employs a large number of antennas at the base stations (BSs) to serve a relatively smaller number of user equipments (UEs) [1]–[4]. This large number of antennas enables each BS to focus the radiated energy into a specific location in space or to intercept the power of transmitted electromagnetic waves more efficiently. Therefore, Massive MIMO has higher spectral efficiency and energy efficiency compared to classical multiuser MIMO systems [3], [5]–[7]. Due to the quasi-orthogonal nature of channels in Massive MIMO, linear precoding and detection schemes perform close-to-optimal [5], [6], [8]. If the channel reciprocity is exploited, the overhead of the channel state information (CSI) acquisition is independent of the number of BS antennas [9]. Moreover, it has been recently shown that the capacity of Massive MIMO increases without bound as the number of antennas increases, even under pilot contamination [10], [11]. These remarkable features candidate Massive MIMO as one of the most promising technologies for next generation of cellular networks [8], [12], [13].

Linear precoding has a central role in Massive MIMO and has been extensively studied in the past few years [5], [14]–[22]. The spectral efficiency and energy efficiency of maximum ratio transmission (MRT) and zero forcing (ZF) precoding in single-cell Massive MIMO systems are investigated in [14]. In [15], a multicell linear precoding is proposed to mitigate the effect of pilot contamination. Multicell processing is also considered in [10], [11], [21]. The performance of MRT, ZF, and regularized ZF (RZF) precoding in single-cell large-scale MIMO systems is studied in [16], considering a per-user channel correlation model. A seminal treatment of MRT and RZF precoding schemes in multicell Massive MIMO systems is presented in [5], followed by [17] where downlink training and linear pilot contamination precoding are also considered. In [18], closed-form approximations for the achievable downlink rates of MRT and ZF precoding schemes are presented for multicell Massive MIMO systems. A linear truncated polynomial expansion based precoding is proposed in [19], which reduces the complexity of RZF precoding. The effect of phase noise on the signal-to-interference-plus-noise (SINR) of MRT, ZF, and RZF precoding schemes is studied in [20].

In order to utilize linear precoding, the power should be adjusted to meet the power constraint at the BS. This can be done either by optimized power allocation among the downlink data streams [21], [23]–[25], or simply by uniform power allocation among downlink data streams jointly with precoder power normalization [5], [14], [17]–[19]. Although the latter approach may provide a weaker performance compared to the former, it is the most used in the Massive MIMO literature [5], [9], [14], [17]–[19]. The reason for this is that power allocation presents the following major issues: (i) finding a global solution is a challenging task [24], [26], [27]; (ii) a certain level of coordination or cooperation among cells is required; and (iii) it should be performed very frequently, even for static users, as scheduling may change rapidly in practice.

The two commonly used power normalization techniques
in Massive MIMO are matrix normalization (MN) and vector normalization (VN) [26], [27]. In MN, the precoding matrix of each BS is adjusted by multiplying it with a scalar such that the power constraint at the BS is met [5], [9], [14], [17], [19]. On the other hand, with VN the precoding matrix is normalized such that equal amount of power is allocated to each UE while satisfying the power constraint [18], [26], [27]. Note that these two methods yield the same performance with optimal power allocation, but not with practical suboptimal power allocation [27], [28].

Although linear precoding has been largely studied in Massive MIMO, a detailed treatment of the impact of power normalization does not exist in the literature. The first attempt in this direction was carried out in [26] and extended in [27] wherein the authors study the impact of MN and VN on MRT and ZF precoding schemes. However, both [26], [27] do not grasp the essence of a practical Massive MIMO system since: (i) a single-cell network composed of three radio units is considered; (ii) perfect CSI is assumed and thus CSI acquisition or pilot contamination are not accounted for; and (iii) large-scale attenuation is neglected, though it has a fundamental impact on power normalization, as detailed later.

The goal of this paper is to study the effect of MN and VN on the performance of MRT, ZF, and RZF in Massive MIMO, in the simple and practical case of uniform power allocation. Particularly, the following contributions are provided.

- We extend the analysis in [26], [27] to a multicell Massive MIMO system, which accounts for channel estimation, pilot contamination, an arbitrary pathloss model, and per-user channel correlation. Asymptotically tight approximations of the signal-to-interference-plus-noise ratio (SINR) and rate of each UE are given and validated by numerical results for MRT, ZF, and RZF with VN and MN.

- Explicit asymptotic approximations for the SINR and rate of each UE are given for a Rayleigh fading channel model. These results are used: (i) to elaborate on how the two different normalization techniques affect the signal, noise, and interference powers as well as the pilot contamination experienced by each UE in the system; (ii) to prove that large-scale fading has a fundamental role on the performance provided by the two normalization techniques while both perform the same if neglected; (iii) to show that ZF conveys a notion of sum rate maximization with VN and of fairness with MN.

- The asymptotic approximations of SINRs are used together with numerical results to study the main limiting factors of the investigated schemes. Particularly, we reveal that in Massive MIMO, non-coherent interference and noise, rather than pilot contamination, are often the major limiting factors for all schemes.

The remainder of this paper is organized as follows. Section II introduces the network model, the channel estimation scheme, the precoding and power normalization methods, as well as the downlink achievable rates. The large system analysis is provided in Section III. The effect of power normalization techniques is elaborated in Section IV for uncorrelated fading channels. The provided asymptotic approximations are verified by means of numerical results in Section V. Some conclusions are drawn in Section VI.

Notations: The following notation is used throughout the paper. Scalars are denoted by lower case letters whereas boldface lower (upper) case letters are used for vectors (matrices). We denote by \( I_N \) the identity matrix of size \( N \) and represent the element on the \( i \)th row and \( k \)th column of \( \mathbf{A} \) as \( [\mathbf{A}]_{i,k} \). The symbol \( \mathcal{CN}(\cdot,\cdot) \) denotes the circularly symmetric complex Gaussian distribution. The trace, transpose, conjugate transpose, real part, and expectation operators are denoted by \( \text{tr}(\cdot), (\cdot)^T, (\cdot)^H, \text{Re}(\cdot), \) and \( \mathbb{E} [\cdot] \), respectively. The notation \( \stackrel{\triangle}{=} \) represent almost sure convergence.

II. COMMUNICATION SCHEME

Next, we introduce the system model, the channel estimation method, the precoding and power normalization techniques, and compute the downlink achievable rates.

A. System Model

We consider the downlink of a Massive MIMO system composed of \( L \) cells, where the set of all cells is denoted by \( \mathcal{L} \). The BS of each cell has \( N \) antennas and serves \( K \) single-antenna UEs in the same time-frequency resource. The set of UEs belonging to cell \( l \) is denoted by \( K_l \). We assume transmissions over flat-fading channels. We employ a double index notation to refer to each UE as e.g., “user \( k \) in cell \( l \)”. Under this convention, let \( \mathbf{h}_{jlk} \in \mathbb{C}^N \) be the channel between BS \( j \) and UE \( k \) in cell \( l \) within a block and assume that

\[
\mathbf{h}_{jlk} = \Theta_{jlk}^{1/2} \mathbf{z}_{jlk}
\]

where \( \mathbf{z}_{jlk} \sim \mathcal{CN}(0, \mathbf{I}_N) \) and \( \Theta_{jlk} \in \mathbb{C}^{N \times N} \) accounts for the corresponding channel correlation matrix. Note that (1) enables us to assign a unique correlation matrix between each user-BS pair and it includes many channel models in the literature as special cases [16].

B. Channel Estimation

We assume that BSs and UEs are perfectly synchronized and operate according to a time-division duplex (TDD) protocol. Then, the channels can be found by an uplink training phase and used in the downlink by exploiting channel reciprocity. Using orthogonal pilots in each cell while reusing them in all other cells, after correlating the received training signal with the pilot sequence of UE \( k \), the observed channel of user \( k \) in cell \( j \) is

\[
y_{jk}^t = \mathbf{h}_{jlk} + \sum_{l=1,l\neq j}^{L} \mathbf{h}_{jlk} + \frac{1}{\sqrt{\rho_{tr}}} \mathbf{n}_{jk}
\]

where \( \mathbf{n}_{jk} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N) \) with \( \sigma^2 \) being the noise spectral density and \( \rho_{tr} \) is proportional to the training SNR. Applying the MMSE estimation, the estimated channel can be computed as follows [5]

\[
\hat{\mathbf{h}}_{jlk} = \Theta_{jlk} \mathbf{Q}_{jlk} y_{jk}^t
\]
where $\hat{h}_{jjk} \sim \mathcal{CN}(0, \Phi_{jjk})$. Also, $Q_{jk}$ and $\Phi_{jlk}$ are given by

$$Q_{jk} = \left( \frac{L}{\sum_{l=1}^{L} \Theta_{jlk} + \sigma^2/\rho_{tr}} I_N \right)^{-1} \quad \forall j, k \tag{4}$$

$$\Phi_{jlk} = \Theta_{jlk} Q_{jk} \Theta_{jlk} \quad \forall j, l, k. \tag{5}$$

Note that due to the orthogonality principle of MMSE, the estimation error $\hat{h}_{jjk} = h_{jjk} - \hat{h}_{jjk}$ is independent of $h_{jjk}$ and such that $\hat{h}_{jjk} \sim \mathcal{CN}(0, \Phi_{jjk})$. For notational simplicity, we denote $\hat{H}_{jj} = [\hat{h}_{jj1}, \ldots, \hat{h}_{jjK}]$ as the matrix collecting the estimated channels of cell $j$.

C. Precoding and Power Normalization Techniques

As mentioned earlier, we consider MRT, ZF, and RZF with VN and MN [26, 27]. Denoting by $G_j = [g_{j1}, \ldots, g_{jK}] \in \mathbb{C}^{N \times K}$ the precoding matrix of BS $j$, where $g_{jk} \in \mathbb{C}^{N}$ is the precoding vector of UE $k$ in cell $j$, we have

$$G_j = F_j D_j^{1/2} \tag{6}$$

where $F_j = [f_{j1}, \ldots, f_{jK}] \in \mathbb{C}^{N \times K}$ determines the precoding scheme and $D_j \in \mathbb{C}^{K \times K}$ characterizes the power allocation strategy. Therefore, $F_j$ takes one of the following forms:

$$F_j = \begin{cases} 
  \hat{H}_{jj} & \text{MRT} \tag{7} \\
  \hat{H}_{jj} \left( \hat{H}_{jj}^H \hat{H}_{jj} \right)^{-1} & \text{ZF} \tag{8} \\
  \left( \hat{H}_{jj} \hat{H}_{jj}^H + Z_j + N\alpha_j I_N \right)^{-1} \hat{H}_{jj} & \text{RZF} \tag{9}
\end{cases}$$

where $\alpha_j > 0$ is the regularization parameter and $Z_j \in \mathbb{C}^{N \times N}$ is an arbitrary Hermitian nonnegative definite matrix that can be used to leverage the system performance [5].

As mentioned in the introduction, finding the optimal values for the entries of $D_j$ is challenging in practice [24]. This is why VN or MN are usually employed [27]. In this case, $D_j$ is diagonal with entries chosen so as to satisfy the following average power constraint $\mathbb{E}[\text{tr} G_j G_j^H] = K \forall j$. If VN is used, then the $k$th diagonal element of $D_j$ is computed as

$$[D_j]_{k,k} = d_{jk} = \frac{1}{\mathbb{E}[\text{tr} F_j^H F_j]} \tag{10}$$

On the other hand, if MN is employed, then $D_j = \eta_j I_K$ with

$$\eta_j = \frac{K}{\mathbb{E}[\text{tr} F_j F_j^H]} \tag{11}$$

D. Downlink Achievable Rate

The received signal of user $k$ in cell $j$ can be written as

$$y_{jk} = h_{jjk}^H g_{jk} s_{jk} + \sum_{i=1, i \neq k}^{K} h_{jjk}^H g_{ji} s_{ji} + \sum_{l=1, l \neq j}^{L} \sum_{i=1}^{K} h_{ljk}^H g_{li} s_{li} + n_{jk} \tag{12}$$

with $s_{li} \in \mathbb{C}$ being the signal intended to UE $i$ in cell $l$, assumed independent across $(l, i)$ pairs, of zero mean and unit variance, and $n_{jk} \sim \mathcal{CN}(0, \sigma^2/\rho_{dl})$ where $\rho_{dl}$ is proportional to the downlink signal power.

As in [1], [5], [6], [15] (among many others), we assume that there are no downlink pilots such that the UEs do not have knowledge of the current channels but can only learn the average channel gain $\mathbb{E}[h_{jjk} g_{jk}]$ and the total interference power. Note this is the common approach in Massive MIMO due to the channel hardening [29]. Using the same technique as in [30], an ergodic achievable information rate for UE $k$ in cell $j$ is obtained as $r_{jk} = \log_2(1 + \gamma_{jk})$ where $\gamma_{jk}$ is given by

$$\gamma_{jk} = \frac{\mathbb{E}[h_{jjk}^H g_{jk}^H g_{jk}^H]}{\sigma^2 + \sum_{l=1}^{L} \sum_{i=1}^{K} \mathbb{E}[h_{ljk}^H g_{li}^H g_{li}^H] - \mathbb{E}[h_{jjk}^H g_{jk}^H g_{jk}^H]} \tag{13}$$

where the expectation is taken with respect to the channel realizations. The above result holds true for any precoding scheme and is obtained by treating the interference (from the same and other cells) and channel uncertainty as worst-case Gaussian noise. By using VN and MN, i.e. (10) and (11), the SINR takes respectively the form in (16) and (17), given on the top of next page.

As for all precoding schemes, $\gamma_{jk}^M$ and $\gamma_{jk}^V$ depend on the statistical distribution of $\{h_{jk}\}$ and $\{h_{jk}\}$. This makes it hard to compute both in closed-form. To overcome this issue, a large system analysis is provided next to find tight asymptotic approximations (hereafter called deterministic equivalents) for $\gamma_{jk}^M$ and $\gamma_{jk}^V$ and their associated achievable rates.

III. LARGE SYSTEM ANALYSIS

We consider a regime in which $N$ and $K$ grow large with a non-trivial ratio $N/K$, where $1 < \liminf N/K \leq \limsup N/K < \infty$. We will represent it as $N \rightarrow \infty$. Under this assumption, we provide asymptotic approximations, also called deterministic equivalents (DEs), for $\gamma_{jk}$ with MRT, ZF, and RZF and either MN or VN. The DE is represented by $\bar{\gamma}_{jk}$, and it is such that $\gamma_{jk} \rightarrow \bar{\gamma}_{jk} \text{ as } N \rightarrow \infty$. By applying the continuous mapping theorem [31], the almost sure convergence of the results illustrated below implies that $r_{jk} - \bar{\gamma}_{jk} \xrightarrow{N \rightarrow \infty} 0$ with $\bar{\gamma}_{jk} = \log_2(1 + \bar{\gamma}_{jk})$, where $\bar{\gamma}_{jk}$ denotes one of the asymptotic approximations computed below.

As limiting cases are considered, the following conditions (widely used in the literature [5, 16, 32, 33]) are needed.

$$A_1 : \limsup \mathbb{E}[\frac{1}{N} \text{tr} (\Theta_{jk})] < \infty \quad \text{and} \quad \liminf \frac{1}{N} \text{tr} (\Theta_{jk}) > 0$$

$$A_2 : 3\epsilon > 0 : \lambda_{\min} \left( \frac{1}{N} H_{il}^H H_{il} \right) > \epsilon$$

$$A_3 : \limsup \left( \frac{1}{N} Z_{ii} \right) < \infty$$

$$A_4 : \text{rank}(\hat{H}_{il}) \geq K$$

A. Large System Results for Vector Normalization

In this subsection, we derive DEs for $\gamma_{jk}^V$, when any of MRT, ZF, and RZF precoding schemes is used.
\[ \gamma_{jk}^{\text{VN}} = \frac{d_{jk} \left| \mathbb{E}[h_{j}^H k f_{jk}] \right|^2}{\sigma_{\eta}^2 + d_{jk} \text{var}(h_{j}^H k f_{jk}) + \sum_{l=1}^{L} \sum_{i=1, i \neq j}^{K} d_{li} \mathbb{E}[|h_{j}^H k f_{li}|^2] + \sum_{l=1, l \neq j}^{L} d_{lk} \mathbb{E}[|h_{j}^H k f_{lk}|^2]} \]  

(16)

\[ \gamma_{jk}^{\text{MN}} = \frac{\eta_j \left| \mathbb{E}[h_{j}^H k f_{jk}] \right|^2}{\sigma_{\eta}^2 + \eta_j \text{var}(h_{j}^H k f_{jk}) + \sum_{l=1}^{L} \sum_{i=1, i \neq j}^{K} \eta_i \mathbb{E}[|h_{j}^H k f_{li}|^2] + \sum_{l=1, l \neq j}^{L} \eta_l \mathbb{E}[|h_{j}^H k f_{lk}|^2]} . \]  

(17)

**Theorem 1.** Let A1 hold true. If MRT with VN is used, then \( \gamma_{jk}^{\text{VN}} \to \gamma_{jk}^{(\text{MRT} - \text{VN})} \) almost surely with \( N \to \infty \) where

\[ \gamma_{jk}^{(\text{MRT} - \text{VN})} = d_{jk}^2 \left( \frac{1}{N} \mathbb{E}[\Phi_{ljk}] \right)^2 \]  

(18)

where

\[ d_{jk}^2 = \frac{1}{N} \mathbb{E}[\Phi_{ljk}] \]  

(19)

\[ z_{ljk} = \frac{1}{N} \mathbb{E}[\Phi_{ljk}] . \]  

(20)

**Proof.** The proof is provided in Appendix A.

**Theorem 2.** Let A1 and A3 hold true. If RZF with VN is used, then \( \gamma_{jk}^{\text{VN}} \to \gamma_{jk}^{(\text{RZF} - \text{VN})} \) almost surely with

\[ \gamma_{jk}^{(\text{RZF} - \text{VN})} = \frac{u_{jk}^2}{N} \]  

(21)

with

\[ d_{li}^0 = \frac{1}{N} \mathbb{E}[\Phi_{ljk} T_{ljk}^\dagger] \]  

(22)

\[ u_{jk} = \frac{1}{N} \mathbb{E}[\Phi_{ljk}] \]  

(23)

\[ u_{ljk} = \frac{1}{N} \mathbb{E}[\Phi_{ljk} T_{ljk}] \]  

(24)

\[ \epsilon_{ljk} = \frac{1}{N} \mathbb{E}[\Phi_{ljk} T_{ljk}^\dagger \Phi_{ljk}] \]  

(25)

and \( S_l = \frac{Z_l}{N} \). Also, \( T_l, T_{ljk}^\dagger \), and \( T_{ljk} \Phi_{ljk} \) are given in Theorems 7 and 8 in Appendix E.

**Proof.** The proof is provided in Appendix B.

**Theorem 3.** Let A1, A2 and A4 hold true. If ZF with VN is employed, then \( \gamma_{jk}^{\text{VN}} \to \gamma_{jk}^{(\text{ZF} - \text{VN})} \) almost surely with

\[ \gamma_{jk}^{(\text{ZF} - \text{VN})} = \frac{u_{jk}^2}{N} \]  

(26)

**Proof.** The proof is provided in Appendix B.

Notice that the computation of the DEs with ZF precoding (either VN or MN) for the considered multiccil Massive MIMO system is more involved than with MRT or RZF precoding schemes. This is mainly due to the fact that it is not straightforward to start with ZF precoder and then compute the DEs by applying common techniques, e.g., matrix inversion lemma. Therefore, in proving Theorem 3 (and also Theorem 6) we start with the DE of RZF and then use a bounding and limiting technique to compute the DE for ZF.

**B. Large System Results for Matrix Normalization**

Next, the DEs of \( \gamma_{jk}^{\text{MN}} \) are given for MRT, ZF, and RZF. Note that the DEs of \( \gamma_{jk}^{\text{MN}} \) for MRT and RZF are obtained from [5].
Theorem 4. [5, Theorem 4] Let A1 hold true. If MRT with MN is used, then \( \gamma_{MN}^{(MRT-MN)} - \gamma_{jk}^{(MRT-MN)} \xrightarrow{N \to \infty} 0 \) almost surely with
\[
\gamma_{jk}^{(MRT-MN)} = \frac{\lambda_j (\frac{1}{N} \text{tr}\Phi_{jk})^2}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} \lambda_l z_{i,jk} + \sum_{l=1,l\neq j}^{L} \lambda_l |\frac{1}{N} \text{tr}\Phi_{ljk}|^2}
\]
where \( z_{i,jk} \) is given in (20) and
\[
\lambda_j = \left( \frac{1}{K} \sum_{k=1}^{K} \frac{1}{N} \text{tr}\Phi_{jk} \right)^{-1}.
\]

Theorem 5. [5, Theorem 6] Let A1 and A3 hold true. If RZF with MN is used, then \( \gamma_{MN}^{(RZF-MN)} - \gamma_{jk}^{(RZF-MN)} \xrightarrow{N \to \infty} 0 \) almost surely with
\[
\gamma_{jk}^{(RZF-MN)} = \frac{\lambda_j u_{jk}^2}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} \lambda_l \epsilon_{i,jk} + \sum_{l=1,l\neq j}^{L} \lambda_l |\frac{1}{N} \text{tr}\Phi_{ljk}|^2}
\]
with
\[
\lambda_j = \left( \frac{K}{N} \frac{1}{\text{tr}\Phi_{jk} - \frac{1}{N} \text{tr}(Z_i, I_N)T_i, I_N) } \right)^{-1}
\]
where \( S_i = Z_i, N \) and \( T_i, I_N \) are given by Theorem 7 and Theorem 8. Also \( u_{li}, u_{ljk}, \) and \( \epsilon_{i,jk} \) are defined in Theorem 2.

Theorem 6. Let A1, A2 and A4 hold true. If ZF with MN is used, then \( \gamma_{MN}^{(ZF-MN)} - \gamma_{jk}^{(ZF-MN)} \xrightarrow{N \to \infty} 0 \) almost surely with
\[
\gamma_{jk}^{(ZF-MN)} = \frac{\lambda_j}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} \lambda_l u_{i,jk}^2 + \sum_{l=1,l\neq j}^{L} \lambda_l |\frac{1}{N} \text{tr}\Phi_{ljk}|^2}
\]
with \( \lambda_j = \left( \frac{1}{K} \sum_{i=1}^{K} \frac{1}{u_{i,jk}} \right)^{-1} \) where \( u_{li}, u_{ljk}, \) and \( \epsilon_{i,jk} \) are given in Theorem 3.

Proof sketch. The proof follows the same procedure as the proof of Theorem 3 presented in Appendix C. Start with the triangle equality and bound \( \gamma_{jk}^{(ZF-MN)} - \gamma_{jk}^{(RZF-MN)} \). Then find the DE of \( \gamma_{jk}^{(ZF-MN)} \) by letting \( \alpha \to 0 \) in \( \gamma_{jk}^{(RZF-MN)} \).

IV. EFFECT OF POWER NORMALIZATION TECHNIQUES

In this section, we use the asymptotic approximations provided above to gain novel insights into the interplay between the different system parameters and the power normalization techniques in Massive MIMO. To this end, we consider a special case of the general channel model of (1) in which \( \Theta_{jk}^N = d_{jk} I_N \) such that
\[
h_{jk} = \sqrt{d_{jk} z_{jk}}
\]
where \( z_{jk} \sim CN(0, I_N) \) and \( d_{jk} \) accounts for an arbitrary large-scale fading coefficient including pathloss and shadowing. Note this corresponds to an uncorrelated fading channel model, which is a quite popular model in Massive MIMO that allows us to capture the essence of the technology [1], [6]. Under the above circumstances, we have that:

Corollary 1. Let \( \lambda_j = d \frac{1}{\sum_{l=1}^{K} \frac{1}{a_{jk}^l}} \) and \( d = 1 - \frac{K}{N} \). If the channel is modelled as in (40), then
\[
\gamma_{jk}^{(ZF-VN)} = \frac{d_{jk}^2}{\sigma^2_{\nu_{jk}}} \left( \nu_{jk} + \sum_{l=1,l \neq j}^{L} \frac{d_{ij}^2}{\alpha_{lk}} \right),
\]
\[
\gamma_{jk}^{(ZF-MN)} = \frac{\lambda_j}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} \lambda_l u_{i,jk}^2 + \sum_{l=1,l\neq j}^{L} \lambda_l |\frac{1}{N} \text{tr}\Phi_{ljk}|^2}
\]
with \( \alpha_{lk} = \sum_{d=1}^{L} d_{lk} + \frac{\sigma^2_{\nu_{jk}}}{\sigma^2_{\nu_{jk}}} \).

Proof. See Appendix D.

Corollary 2. Let \( \theta_l = \left( \frac{1}{K} \sum_{i=1}^{K} \frac{d_{ij}^2}{\alpha_{ij}^l} \right)^{-1} \). If the channel is modelled as in (40) and MRT is used, then
\[
\gamma_{jk}^{(MRT-VN)} = \frac{d_{jk}^2}{\sigma^2_{\nu_{jk}}} \left( \nu_{jk} + \sum_{l=1,l \neq j}^{L} \frac{d_{ij}^2}{\alpha_{lk}} \right),
\]
\[
\gamma_{jk}^{(MRT-MN)} = \frac{\theta_l (d_{lk} d_{jk})^2}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} d_{ij} + \sum_{l=1,l\neq j}^{L} \theta_l (d_{lk} d_{jk})^2}
\]
with
\[
\theta_l = \frac{\sigma^2}{\sum_{d=1}^{L} \frac{\sigma^2}{N \rho_{dl}} + \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} d_{ij} + \sum_{l=1,l\neq j}^{L} \theta_l (d_{lk} d_{jk})^2}
\]
Proof. The proof follows a similar procedure as that of Corollary 1.

The results of Corollaries 1 and 2 are instrumental in obtaining the following insights into MRT and ZF with either MN or VN.

**Remark 1** (Effect of VN and MN). The terms $\nu_{jk}$ and $\vartheta_{jk}$ in (43) and (46) are the same for both VN and MN. This means that both normalization techniques have exactly the same effect on the resulting noise and interference terms experienced by each UE in the system. On the other hand, they affect differently the signal and pilot contamination powers. The expressions (41)-(46) explicitly state the relation between the SINR contributions (signal, interference, noise, and pilot contamination), the propagation environment, and the two normalization techniques for ZF and MRT precoding schemes.

**Remark 2** (On the mutual effect of UEs). If VN is employed, then the signal power and the pilot contamination of UE $k$ in cell $j$, for both MRT and ZF precoding, depends only on the coefficients $d_{ljk}$ $\forall l, n \in L$ through $\alpha_{lk}$. This means that they are both affected only by the large-scale gains of the UEs in the network using the same pilot. On the other hand, under MN both terms depend on the coefficients $\lambda_i \forall l \in L$ (or $\vartheta_i$ for MRT) and thus are influenced by all the UEs in the network, even though they make use of different pilot sequences.

**Remark 3** (Large-scale fading and power normalization). Assume that the large-scale fading is neglected such that it is the same for every UE in the network, i.e., $d_{ljk} = d \forall l, j, k$. Then, the expressions in (41) and (42) for ZF and those in (44) and (45) for MRT become equal. This means that the large-scale fading has a fundamental impact on VN and MN and cannot be ignored.

Consider now, for further simplicity, a single-cell setup, i.e., $L = 1$. Dropping the cell index, $\alpha_{lk}$ reduces to $\alpha_k = d_k + \sigma^2 / \rho_{\text{tr}}$. Also assume that the UEs operate in the high training SNR regime such that $\rho_{\text{tr}} \gg 1$. Under these conditions, we have that:

**Lemma 1.** If $L = 1$ and $\rho_{\text{tr}} \gg 1$, then for ZF precoding, VN outperforms MN in terms of sum rate and the sum rate gap $\Delta r \geq 0$ is given by

$$\Delta r = \sum_{k=1}^{K} \log \left(1 + \frac{1}{\mathcal{N} \rho_{\text{tr}} \bar{d}_k} \right) - K \log \left(1 + \frac{1}{\mathcal{N} \rho_{\text{tr}} \bar{d}} \right).$$

Proof. From Corollary 1, setting $L = 1$ and assuming $\rho_{\text{tr}} \gg 1$ we obtain that $\alpha_k \simeq d_k$ and $\nu_k \simeq \sigma^2 / \mathcal{N} \rho_{\text{tr}}$. Then, the result follows by applying the Jensen’s inequality (by the convexity of log $(1 + 1/x)$).

Notice that Lemma 1 extends the results of [26] and [27] to a system that accounts for CSI acquisition and arbitrary pathloss and UEs’ distribution. Also, observe that (41) and (42) simplify as:

$$\gamma_{jk}^{(\text{ZF-VN})} = \frac{(N - K) \rho_{\text{dl}}}{\sigma^2} d_k$$

$$\gamma_{jk}^{(\text{ZF-MN})} = \frac{(N - K) \rho_{\text{dl}}}{\mathcal{N} \sum_{i=1}^{K} \frac{1}{d_i}}$$

from which it follows that VN provides higher SINR to the UEs that are closer to the BS and lower SINR for those that are far away from the BS (which resembles opportunistic resource allocation). On the other hand, MN provides a uniform quality of experience to all UEs. This proves evidence of the fact that ZF with VN resembles a sum rate maximizer. On the other hand, it provides a notion of fairness under MN. Notice that fairness means similar SINR (quality of experience) and it should not be confused with equal power allocation. The above results and observations will be validated below in Section V by means of numerical results. Also, the DEs provided in Corollaries 1 and 2 will be used to investigate the main limiting factors of Massive MIMO.

V. NUMERICAL RESULTS

Monte-Carlo simulations are now used to validate the asymptotic analysis for different values of $N$ and $K$. We consider a multicell network composed of $L = 7$ cells, one in the center and six around. Each cell radius is 1000 meters. A 20 MHz channel is considered and the thermal noise power is assumed to be $-174$ dBm/Hz. The UEs are randomly and uniformly distributed within each cell excluding a circle of radius 100 meters. The channel is modeled as in [36]. In particular, we assume that the matrices $\Theta_i^{1/2}$ are given by

$$\Theta_i^{1/2} = \sqrt{d_{ljk} A}$$

where $A = [a(\theta_1), \ldots, a(\theta_N)] \in \mathbb{C}^N$ with $a(\theta_i)$ given by

$$a(\theta_i) = \frac{1}{\sqrt{N}} [1, e^{-i2\pi \omega \sin(\theta_i)}, \ldots, e^{-i2\pi \omega (N - 1) \sin(\theta_i)}]^T$$

where $\omega = 0.3$ is the antenna spacing and $\theta_i = -\pi / 2 + (i - 1) \pi / N$. Also, $d_{ljk}$ is the large-scale attenuation, which is modeled as $d_{ljk} = x_{ljk}^{-2}$ where $x_{ljk}$ denotes the distance of UE $k$ in cell $j$ from BS $l$ and $\beta = 3.7$ is the path-loss exponent. We let $\rho_{\text{tr}} = 6$ dB and $\rho_{\text{dl}} = 10$ dB, which corresponds to a practical setting [5]. The results are obtained for 100 different channel and UE distributions realizations.

Figs. 1 and 2 validate the accuracy of the DEs provided in Theorems 1, 2, 3, and 6. In particular, both figures report the ergodic achievable sum rate of the center cell versus $N$ for $K = 8$ and 16, respectively. The solid lines correspond to the asymptotic sum rate whereas the markers are achieved through Monte Carlo simulation. As it is depicted, the asymptotic approximation match perfectly with numerical results. Notice that Figs. 1 and 2 (and also Table 1) extend the results in [26] and [27] in the sense that account for CSI acquisition, pilot contamination, arbitrary pathloss and UEs’ distribution.

In Lemma 1, it is shown that ZF under VN conveys a notion of sum rate maximization, while ZF with MN resembles a fairness provisioning precoder. Now, we use Table I to
and 5 are now used to investigate a common belief in the Massive MIMO literature, that is: under uncorrelated fading when $N \to \infty$ the noise and interference contributions vanish asymptotically and pilot contamination becomes the unique bottleneck of the system performance. This follows also from the results in Corollaries 1 and 2 by letting $N$ grow large with $K$ kept fixed. However, in [29] it is shown that it is desirable for Massive MIMO systems to work in a regime where $N/K \leq 10$. Therefore, it is interesting to see what is the major impairment for Massive MIMO under this practical regime: (i) is it pilot contamination (or coherent interference)?; (ii) is it the noise and interference (or more exactly the non-coherent interference)?; (iii) how is the answer related to the choice of the power normalization technique and precoding scheme?

To answer these questions, we employ the so-called pilot contamination-to-interference-plus-noise ratio (PCINR) metric, which is computed by using the DEs provided in Corollaries 1 and 2. Fig. 3 plots the PCINR as a function of $N/K$, i.e., the number of degrees of freedom per-user in the system. Although, the optimal operating regime for maximal spectral efficiency is for $N/K < 10$ [29], we consider $N/K$ up to 20 to cover a wider range of Massive MIMO configurations. Moreover, as the interference increases by having more UEs in the system, we consider three different scenarios with $K = 5$, $K = 10$, and $K = 15$.

Fig. 3 is divided into 3 regions based on the significance of the PCINR term such that, as we move away from region 1 towards region 3, the importance of pilot contamination increases while that of the interference plus noise reduces. Region 1 is where the noise and interference are the dominant limiting factors and pilot contamination has a negligible effect—less than 10% of the noise and interference. As it is depicted, MRT with MN operates within this regime, therefore pilot contamination is never a bottleneck for this scheme, which is mainly limited by noise and interference. Notice that by adding more UEs in the system, the PCINR reduces and pilot contamination becomes even less important. Hence, when MRT with MN is studied in Massive MIMO the effect of pilot contamination can be safely neglected.

Region 2 represents the regime where the noise and inter-

---

**TABLE I: SINR of each UE under ZF with VN and MN.**

| No. Ant. | MN | MC | Ef. % | VN | MC | Ef. % |
|----------|----|----|------|----|----|------|
| N = 40   |    |    |      |    |    |      |
| 1        | 2.25 | 2.19 | 2.6 | 1.85 | 1.84 | 0.5 |
| 2        | 4.89 | 4.84 | 1.0 | 147 | 149 | 1.3 |
| 3        | 3.34 | 3.29 | 1.5 | 3.61 | 3.53 | 2.2 |
| 4        | 5.14 | 5.12 | 0.3 | 37.5 | 37.6 | 0.2 |
| 5        | 4.09 | 4.02 | 1.7 | 1.97 | 1.96 | 0.5 |
| 6        | 4.85 | 4.84 | 2.5 | 85 | 87 | 2.3 |
| 7        | 3.30 | 3.33 | 0.9 | 2.14 | 2.2 | 2.8 |
| 8        | 3.32 | 3.30 | 0.5 | 2.52 | 2.49 | 1.2 |
| N = 80   |    |    |      |    |    |      |
| 1        | 3.20 | 3.15 | 1.5 | 31 | 2.98 | 3.8 |
| 2        | 10.57 | 10.40 | 1.6 | 316 | 310 | 1.9 |
| 3        | 5.94 | 6.03 | 1.5 | 6.40 | 6.36 | 0.6 |
| 4        | 9.33 | 9.47 | 1.5 | 72.2 | 72.2 | 0 |
| 5        | 8.41 | 8.62 | 2.5 | 3.87 | 3.90 | 0.7 |
| 6        | 9.85 | 9.12 | 4.1 | 182 | 185 | 1.6 |
| 7        | 5.02 | 4.90 | 2 | 3.50 | 3.43 | 2 |
| 8        | 5.89 | 5.74 | 2.5 | 4.28 | 4.21 | 1.6 |
ference are the main limiting factors, but pilot contamination is not negligible any more. It is interesting to observe that for the other schemes (other than MRT-MN), Massive MIMO often operates within this region. This shows that, although pilot contamination is a major challenge in Massive MIMO, the interference and noise have still the leading role in limiting the system performance.

Finally, region 3 presents the superiority of pilot contamination effect. If \( K = 10 \), then Fig. 3b shows the superiority of interference and noise over pilot contamination for ZF-MN and RZF-MN (ZF-VN and RZF-VN) up to \( N = 233 \) antennas at the BS. With MRT-VN, the system requires more than \( N = 510 \) to experience the superiority of pilot contamination over interference and noise. This increases to \( N = 2650 \) with MRT-MN. From Fig. 3, we see also that, for a given value of \( N/K \), the value of PCINR for the considered schemes can be ordered as: ZF-MN = RZF-MN ≥ ZF-VN = RZF-VN ≥ MRT-VN ≥ MRT-MN. Based on the above discussion, it is clear that the choice of precoding scheme and normalization technique change the importance of pilot contamination, interference, and noise dramatically and it should be considered carefully when designing Massive MIMO systems.

VI. CONCLUSIONS

Linear precoding schemes, such as MRT and ZF, have a fundamental role in Massive MIMO. Although these precoding schemes can be employed with optimized power control policies, they are usually implemented by simple matrix or vector power normalization techniques. This is due to the complexity of attaining optimal power control policies [24], as it requires coordination and cooperation among cells and computationally demanding algorithms. On the other hand, the two above precoding power normalization techniques are simple and efficient [2], [3], [5], [6].
This work made use of large system analysis to compute tight asymptotic approximations of the SINR experienced by each UE in the system when using MN or VN. These results can be used to evaluate the performance of practical Massive MIMO systems without the need for time-consuming Monte Carlo simulations. Under uncorrelated fading channels, we analytically showed that MN and VN treat the noise and interference in the same manner, but have different effects on pilot contamination and received signal power. We also revealed the key role played by large-scale fading, positions of UEs, and pilot assignment into power normalization. We explained how a simple change in power normalization can resemble two totally different behaviors, namely, sum-rate maximization or fairness provisioning. Moreover, we showed numerically how the choice of the normalization technique can change the main bottleneck of Massive MIMO systems.

**APPENDIX A**

We begin by plugging (7) and (10) into $\gamma_{jk}^{\text{VN}}$ given in (16) to obtain $\gamma_{jk}^{(\text{MRT-VN})}$. Then, we divide the numerator and denominator of $\gamma_{jk}^{(\text{MRT-VN})}$ by $N$ and define $d_{ii} = N d_{i}$. By applying the continuous mapping theorem [31] and replacing each component of $\gamma_{jk}^{(\text{MRT-VN})}$ by its DE, the DE of $\gamma_{jk}$ is obtained. Notice that DEs of signal power component, variance component, and interference components are given in [5]. Therefore, we only need to compute a DE of the VN coefficient. The latter is given by

$$d_{ii} = \frac{1}{E[h_{ii}^H h_{ii}]} = \frac{1}{\text{tr} E[h_{ii}^H h_{ii}]} = \frac{1}{\text{tr} \Phi_{ii}^{(MRT)}}$$

where in (a) we have applied Lemma 3 in Appendix E and used the fact that $h_{ii} \sim \mathcal{CN}(0, \Phi_{ii})$.

**APPENDIX B**

Let us define for convenience

$$\mathbf{H}_{l[k]}^H \mathbf{H}_{l[k]}^H = \mathbf{H}_l^H \mathbf{H}_l^H - \mathbf{H}_l^H \mathbf{H}_{l[k]}^H$$

and $C_l = N C_l$ and $C_{l[k]} = N C_{l[k]}$ with

$$C_l = (\mathbf{H}_l^H \mathbf{H}_l + Z_l + \alpha_1 N I_N)^{-1}$$

and

$$C_{l[k]} = (\mathbf{H}_{l[k]}^H \mathbf{H}_{l[k]} + Z_l + \alpha_1 N I_N)^{-1}.$$

Plugging (9) and (10) into (16) yields $\gamma_{jk}^{(\text{RZF-VN})}$. Then, we divide the numerator and denominator by $N$ and replace each term in $\gamma_{jk}^{(\text{RZF-VN})}$ with its DE. Notice that DEs of signal power component, variance component, and interference components are given in [5]. Therefore, we only need to compute a DE of the VN coefficient. From (10), we have that

$$d_{ii} = \frac{1}{N} d_{ii} = \frac{1}{N} (E[h_{ii}^H C_{ii}^2 h_{ii}])^{-1}.$$

An asymptotic approximation for $\mathbf{h}_{ii}^H C_{ii}^2 \mathbf{h}_{ii}$ can be computed as follows

$$\mathbf{h}_{ii}^H C_{ii}^2 \mathbf{h}_{ii} \approx \frac{a_N}{N \text{tr} \Phi_{ii}^{(MRT)}}$$

where $a_N \approx b_N$ is equivalent to $a_N - b_N \rightarrow 0$ as $N \rightarrow \infty$.

**APPENDIX C**

The main idea is to first compute a DE for RZF with $\alpha_l = \alpha$ for all $l \in L$ and then to obtain a DE for ZF by letting $\alpha_l \rightarrow 0$. By using the triangle inequality $|\gamma_{jk}^{(\text{ZF-VN})} - \gamma_{jk}^{(\text{RZF-VN})}|$ can be bounded as follows

$$|\gamma_{jk}^{(\text{ZF-VN})} - \gamma_{jk}^{(\text{RZF-VN})}| \leq |\gamma_{jk}^{(\text{ZF-VN})} - \gamma_{jk}^{(\text{ZF-VN})}| + |\gamma_{jk}^{(\text{RZF-VN})} - \gamma_{jk}^{(\text{RZF-VN})}|$$

Next, we show that each term in the right hand side of (59) can be made arbitrarily small (i.e. smaller than any given $\epsilon > 0$) when $\alpha_l \rightarrow 0$. Let us start with $|\gamma_{jk}^{(\text{ZF-VN})} - \gamma_{jk}^{(\text{RZF-VN})}|$. Notice that $\gamma_{jk}^{(\text{ZF-VN})}$ and $\gamma_{jk}^{(\text{RZF-VN})}$ are different because of the different form of $\mathbf{F}_l$ in (8) and (9). As $\alpha_l \rightarrow 0$ and for $Z_l = 0$, we have

$$\lim_{z_{i} \rightarrow 0} \frac{a_{i}^{(\text{ZF})}}{a_{i}^{(\text{RZF})}} = \lim_{\alpha \rightarrow 0} \left( \frac{\mathbf{H}_l \mathbf{H}_l^H + N \alpha_1 I_N}{\mathbf{H}_l \mathbf{H}_l} \right) = \mathbf{H}_l \mathbf{H}_l^{-1} e_k = a_{i}^{(\text{ZF})}.$$
Define $u_{lk} := \lim_{\alpha \to 0} \alpha u_{lk}$ for every $l$ and $k$. Based on [16] and by replacing $u_{lk}$ from Theorem 2 we have

$$u_{lk} = \lim_{\alpha \to 0} \alpha u_{lk} = \lim_{\alpha \to 0} \frac{1}{N} \text{tr} \Phi_{llk} \left( \frac{1}{N} \sum_{i=1}^{K} \Phi_{lii} + I_N \right)^{-1}$$

$$= \frac{1}{N} \text{tr} \left( \Phi_{llk} T_l \right)$$ (62)

with

$$T_l = \left( \frac{1}{N} \sum_{i=1}^{K} \Phi_{lii} \right)^{-1}.$$ (63)

Also defining $u_{ljk} := \lim_{\alpha \to 0} \alpha u_{ljk}$ for every $l$, $j$, and $k$ we have

$$u_{ljk} = \lim_{\alpha \to 0} \frac{1}{N} \text{tr} \Phi_{ljk} T_l = \frac{1}{N} \text{tr} \Phi_{ljk} T_l.$$ (64)

For the term $\lim_{\alpha \to 0} d_{ll}^2$, we obtain

$$\lim_{\alpha \to 0} d_{ll}^2 = \lim_{\alpha \to 0} \frac{u_{ll}^2}{\alpha^2} \frac{1}{N} \text{tr} \Phi_{ll} \left( \frac{1}{N} \sum_{i=1}^{K} \Phi_{lii} \right)^{-1}$$

$$= \lim_{\alpha \to 0} \frac{u_{ll}^2}{\alpha^2} \frac{1}{N} \text{tr} \Phi_{ll} T_l$$ (65)

where $T_l$ and $u_{ljk}$ are given in Theorem 8 in Appendix E. Notice that

$$\lim_{\alpha \to 0} \alpha^2 u_{ljk} = \frac{1}{N} \text{tr} \Phi_{ll} T_l = \frac{1}{N} \sum_{i=1}^{K} \left( \frac{u_{ljk}^2}{u_{ll}^2} + I_N \right) T_l$$

from which, by replacing $\lim_{\alpha \to 0} \alpha^2 u_{ljk} I_N$ with $u_{ljk}$, we have that (66) reduces to $d_{ll}^2 = u_{ljk}^2 \frac{1}{N} \text{tr} \Phi_{ll} T_l$. Therefore, we have that $\lim_{\alpha \to 0} \alpha^2 u_{ljk} = u_{ljk}^2$. From (65), we can thus conclude that $d_{ll}^2 = u_{ll}^2$. On the other hand, for $\epsilon_{i,j,k}$ we have

$$\epsilon_{i,j,k} = \lim_{\alpha \to 0} \alpha^2 \epsilon_{i,j,k}$$

$$= \frac{1}{N} \text{tr} \Theta_{i,j} T_l' \Phi_{lii} - \frac{2}{u_{lk}} \text{Re} \left( u_{l,j,k}^* \frac{1}{N} \text{tr} \Phi_{ljk} T_l' \Phi_{lii} \right)$$

$$+ \frac{|u_{ljk}|^2}{u_{ll}^2} \frac{1}{N} \text{tr} \Phi_{ll} T_l' \Phi_{lii}$$ (67)

where $T_l'$ and $\Phi_{lii}$ is

$$T_l' \Phi_{lii} = \lim_{\alpha \to 0} \alpha^2 T_l \Phi_{lii}$$

$$= T_l \left( \frac{1}{N} \sum_{i=1}^{K} \frac{u_{ljk}^2}{u_{ll}^2} + \Phi_{lii} \right) T_l$$ (68)

where $u_{ljk}^2 = \lim_{\alpha \to 0} \alpha^2 u_{ljk}$. From Theorem 8, we have

$$u_{ljk}^2 = \lim_{\alpha \to 0} \left( I_K - J_l \right)^{-1} \frac{1}{\alpha^2} \Phi_{ljk} = \left( I_K - J_l \right)^{-1} \Phi_{ljk}$$

where $J_l$ and $\Phi_{ljk}$ are given by (33) and (34), respectively. Therefore, $\epsilon_{l,j,k} = \lim_{\alpha \to 0} \alpha^2 \epsilon_{l,j,k}$ follows (30). Using all the above results in (61) completes the proof.

**APPENDIX D**

For brevity we only consider ZF with VN. The same steps can be used for ZF with MN. If the channel is modelled as in (40), then $\Theta_{i,j} = d_{ij} I_N$ and

$$\Phi_{ij} = \frac{d_{ijk} d_{ij}}{\alpha_{ij}} I_N$$ (69)

with $\alpha_{ij} = \sum_{n=1}^{L} d_{i,n} + \sigma_n^2$. Plugging (69) into (27) and (28) yields $u_{lk} = \frac{d_{lk} d_{lj}}{\alpha_{lj}} \frac{1}{N} \text{tr} \left( T_l \right)$ with

$$T_l = \left( \frac{1}{N} \sum_{i=1}^{K} \frac{1}{N} \text{tr} (T_l) + 1 \right)^{-1} I_N.$$ (70)

Call $\tilde{u} = \frac{1}{N} \text{tr} \left( T_l \right)$. Therefore, we have that

$$\tilde{u} = \frac{1}{N} \text{tr} \left( T_l \right) = \left( \frac{K}{N} \tilde{u} + 1 \right)^{-1}.$$ (71)

Solving with respect to $\tilde{u}$ yields $\tilde{u} = 1 - \frac{K}{N}$. Then, we eventually have that

$$u_{lk} = \frac{d_{lk} d_{lj}}{\alpha_{lj}} \tilde{u}$$ (72)

and also $u_{ljk} = \frac{d_{lk} d_{lj}}{\alpha_{lj}} \tilde{u}$. Therefore, the pilot contamination term in $\sum_{T_{jk}}$ reduces to

$$\sum_{l=1}^{L} \frac{u_{ljk}}{d_{lj}} \sum_{l=1, l \neq j}^{L} \frac{d_{lk}}{d_{lj}} \tilde{u}.$$ (73)

Let’s now compute $[J_l]_{i,i}$ defined as in (33). Using the above results yields

$$[J_l]_{i,i} = \frac{1}{N^2} \frac{d_{li}}{\alpha_{li}} \frac{d_{lj}}{\alpha_{lj}} \frac{1}{N} \text{tr} (T_l^2) = \frac{1}{N} \frac{d_{li}}{\alpha_{li}} d_{lj}.$$ (74)

Similarly, we have that

$$[v_{l,i}] = \frac{d_{li}}{\alpha_{li}} d_{lj}.$$ (75)

In compact form, we may write $J_l$ and $v_{l,i}$ as

$$J_l = \frac{1}{N^2} \frac{d_{li}}{\alpha_{li}} \frac{d_{lj}}{\alpha_{lj}} \text{tr} (T_l^2) = \frac{1}{N} \frac{d_{li}}{\alpha_{li}} d_{lj}.$$ (76)

with $[a_l]_l = [d_{li}/\alpha_{li}]$ and $[b_l]_l = 1/\alpha_{li}$. Then, we have that (by applying Lemma 2)

$$u_{ljk} = \frac{d_{lk} d_{lj}}{\alpha_{lk}} \frac{1}{N} \text{tr} (T_l)$$ (77)

Plugging the above result into (31) produces

$$T_l' \Phi_{lii} = \frac{d_{li}}{\alpha_{li}} \frac{1}{N} \text{tr} (T_l) \left( \frac{K}{N} \tilde{u} + 1 \right) T_l = \frac{d_{li}}{\alpha_{li}} \tilde{u} I_N = u_{ljk} I_N.$$ (78)

We are thus left with evaluating (30). Using the above results yields

$$\epsilon_{l,j,k} = \frac{d_{lj}}{N} \text{tr} (T_l' \Phi_{lii}) - \frac{1}{N} \frac{d_{lj}}{d_{li}} \frac{1}{\alpha_{lj}} \text{tr} (\Phi_{ljk} T_l' \Phi_{lii})$$

$$+ \frac{d_{lj}^2}{d_{li}} \frac{1}{N} \text{tr} (\Phi_{lk} T_l' \Phi_{lii})$$ (79)
from which, using (69) and (78), we obtain

\[ \epsilon_{li,jk} = d_{lj,k} u_{li} - \frac{d_{lj,k}^2}{\alpha_{lk}} u_{li}. \]

Therefore, we have that

\[ \frac{1}{N} \epsilon_{li,jk} = \frac{1}{N} d_{lj,k} \left( 1 - \frac{d_{lj,k}}{\alpha_{lk}} \right). \]  \(\text{(80)}\)

Plugging (72), (73) and (80) into (26) produces

\[ \frac{1}{N} \sum_{l=1}^{L} \sum_{i=1}^{K} \epsilon_{li,jk} u_{li} = \frac{K}{N} \sum_{l=1}^{L} d_{lj,k} \left( 1 - \frac{d_{lj,k}}{\alpha_{lk}} \right). \]  \(\text{(81)}\)

Collecting all the above results together completes the proof.

**APPENDIX E**

**Useful Results**

**Theorem 7.** [16, Theorem 1] Let \( B_l = \frac{1}{N} H_l H_l^H + S_l \) with \( H_l \in \mathbb{C}^{N \times K} \) be random with independent column vectors \( h_{lk} \sim \mathcal{CN}(0, \Phi_{lk}) \) for \( k \in \{1, \ldots, K\} \), \( S_l \in \mathbb{C}^{N \times N} \) and \( Q_l \in \mathbb{C}^{N \times N} \) be Hermitian nonnegative definite. Assume that \( Q_l \) and the matrices \( \Phi_{lk} \) for \( k \in \{1, \ldots, K\} \) have uniformly bounded spectral norms (with respect to \( N \)). Define

\[ m_{B_l, Q_l}(-\alpha_l) \triangleq \frac{1}{N} \text{tr} Q_l (B_l + \alpha_l I_N)^{-1} \]

Then, for any \( \alpha_l > 0 \), as \( N \) and \( K \) grow large with \( \beta = \frac{N}{K} \) such that \( 0 < \liminf \beta \leq \limsup \beta < \infty \) we have that

\[ m_{B_l, Q_l}(-\alpha_l) - m_{B_l, Q_l}^0(-\alpha_l) \propto 0 \]  \(\text{(83)}\)

where \( m_{B_l, Q_l}(-\alpha_l) \) is given by

\[ m_{B_l, Q_l}^0(-\alpha_l) = \frac{1}{N} \text{tr} Q_l T_l \]

with \( T_l \) is given by

\[ T_l = \left( \frac{1}{N} \sum_{i=1}^{K} \frac{\Phi_{li}}{1 + u_{li}(-\alpha_l)} + S_l + \alpha_l I_N \right)^{-1} \]

where the elements of \( u_t(-\alpha_l) = [u_{t1}(-\alpha_l), \ldots, u_{tK}(-\alpha_l)]^T \) are defined as \( u_{ti}(-\alpha_l) = \lim_{t \to \infty} u_{ti}^{(t)}(-\alpha_l) \), where for \( t \in \{1, 2, \ldots\} \)

\[ u_{ti}^{(0)}(-\alpha_l) = \frac{1}{N} \text{tr} \Phi_{li} \left( \frac{1}{N} \sum_{i=1}^{K} \frac{\Phi_{li}}{1 + u_{li}^{(t-1)}(-\alpha_l)} + S_l + \alpha_l I_N \right)^{-1} \]

with initial values \( u_{ti}^{(0)}(-\alpha_l) = \frac{1}{N} \) for all \( k \).

**Theorem 8.** [16] Let \( \Omega_l \in \mathbb{C}^{N \times N} \) be Hermitian nonnegative definite with uniformly bounded spectral norm (with respect to \( N \)). Under the conditions of Theorem 1

\[ \frac{1}{N} \text{tr} Q_l \left( \frac{1}{N} H_l H_l^H + S_l + \alpha_l I_N \right)^{-1} \Omega_l \]

\[ \left( \frac{1}{N} H_l \bar{H}_l^H + S_l + \alpha_l I_N \right)^{-1} - \frac{1}{N} \text{tr} Q_l T_l' \Omega_l = 0 \]

where \( T_l' \Omega_l \in \mathbb{C}^{N \times N} \) is defined as

\[ T_l' \Omega_l = T_l \times \left( \frac{1}{N} \sum_{j=1}^{K} u_{lj}(\alpha_l) \Phi_{lj} + \Omega_l \right) \times T_l \]

where \( T_l \) and \( u_l(-\alpha_l) \) are given by theorem 1, and \( u_l(-\alpha_l) = [u_{l1}(-\alpha_l), \ldots, u_{lK}(-\alpha_l)]^T \) is computed from

\[ u_l(-\alpha_l) = (I_K - J_l)^{-1} v_l \]

where \( J_l = \mathbb{C}^{K \times K} \) and \( v_l \in \mathbb{C}^{K} \) are:

\[ [J_l]_{mn} = \frac{\text{tr} \Phi_{llm} T_l \Phi_{llm} T_l}{N^2 (1 + u_{llm}(-\alpha_l))^2} \quad 1 \leq m, n \leq K \]  \(\text{(90)}\)

\[ [v_l, \Omega_l]_{ll} = \frac{1}{N} \text{tr} \Phi_{llm} T_l \Omega_l T_l \quad 1 \leq t \leq K \]  \(\text{(91)}\)

**Lemma 2** (Matrix Inversion Lemma). Let \( U \) be an \( N \times N \) invertible matrix and \( x, y \sim \mathcal{CN}(0, \frac{1}{N} I_N) \). Assume that \( A \) has uniformly bounded spectral norm (with respect to \( N \)) and that \( x \) and \( y \) are mutually independent and independent of \( A \). Then, for all \( p \geq 1 \),

\[ x^H (U + c xx^H)^{-1} = \frac{x^H U^{-1}}{1 + c x^H U^{-1} x}. \]  \(\text{(92)}\)

**Lemma 3** (Trace Lemma). Let \( A \in \mathbb{C}^{N \times N} \) and \( x, y \sim \mathcal{CN}(0, \frac{1}{N} I_N) \). Assume that \( A \) has uniformly bounded spectral norm (with respect to \( N \)) and that \( x \) and \( y \) are mutually independent and independent of \( A \). Then, for all \( p \geq 1 \),

\[ x^H A x - \frac{1}{N} \text{tr} A x \propto 0 \quad \text{and} \quad x^H A y \propto 0. \]  \(\text{(93)}\)

**Lemma 4** (Rank-1 perturbation lemma). Let \( A_1, A_2, \ldots, \) with \( A_N \in \mathbb{C}^{N \times N} \), be deterministic with uniformly bounded spectral norm and \( B_1, B_2, \ldots, \) with \( B_N \in \mathbb{C}^{N \times N} \), be random Hermitian, with eigenvalues \( \lambda^*_1 \leq \lambda^*_2 \leq \ldots \leq \lambda^*_N \) such that, with probability 1, there exist \( c > 0 \) for which \( \lambda^*_N > \epsilon \) for all large \( N \). Then for \( v \in \mathbb{C}^N \)

\[ \frac{1}{N} \text{tr} A_N B_N^{-1} - \frac{1}{N} \text{tr} A_N (B_N^{-1} + vv^H)^{-1} \propto 0. \]

where \( B_N^{-1} \) and \( (B_N^{-1} + vv^H)^{-1} \) exist with probability 1.

**References**

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. on Wireless Commun., vol. 9, no. 11, pp. 3590–3600, November 2010.

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

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, 2014.

[4] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and Y. Su, “Convergence analysis and assurance for gaussian message passing iterative detector in massive MU-MIMO systems,” IEEE Trans. Wireless Commun., vol. 15, no. 9, pp. 6487–6501, Sep. 2016.

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

[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” IEEE Trans. Commun., vol. 61, no. 4, pp. 1436–1449, 2013.

[7] J. Zhao, J. Zhang, C. Yuan, W. Jiang, and W. Luo, “Energy-efficient downlink transmission for multicell massive DAS with pilot contamination,” IEEE Trans. Veh. Technol., vol. 66, no. 2, pp. 1209–1221, Feb. 2017.
