Polynomial Expansion of the Precoder for Power Minimization in Large-Scale MIMO Systems

Houssem Sifaou†, Abla Kammoun†, Luca Sanguinetti⋆⋄, Mérouane Debbah⋄∗, Mohamed-Slim Alouini†

† Electrical Engineering Department, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
⋆ Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
⋄Large Networks and System Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France
∗ Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France.

***Abstract*—This work focuses on the downlink of a single-cell large-scale MIMO system in which the base station equipped with M antennas serves K single-antenna users. In particular, we are interested in reducing the implementation complexity of the optimal linear precoder (OLP) that minimizes the total power consumption while ensuring target user rates. As most precoding schemes, a major difficulty towards the implementation of OLP is that it requires fast inversions of large matrices at every new channel realizations. To overcome this issue, we aim at designing a linear precoding scheme providing the same performance of OLP but with lower complexity. This is achieved by applying the truncated polynomial expansion (TPE) concept on a per-user basis. To get a further leap in complexity reduction and allow for closed-form expressions of the per-user weighting coefficients, we resort to the asymptotic regime in which M and K grow large with a bounded ratio. Numerical results are used to show that the proposed TPE precoding scheme achieves the same performance of OLP with a significantly lower implementation complexity.**

I. Introduction

Large-scale multiple-input multiple-output (MIMO) systems, also known as massive MIMO systems, are considered as a promising technology to sustain the exponential growth of data traffic in cellular networks [1]–[4]. By deploying large-scale arrays with very many antennas at base stations (BSs), an exceptional array gain and spatial precoding resolution can be obtained. This can be exploited to achieve higher rates and serve more user equipments (UEs) simultaneously. In the downlink, this is usually achieved using different linear precoding techniques such as (among others) maximum ratio transmission (MRT), zero-forcing (ZF) and regularized ZF (RZF). In this work, we focus on the problem of designing the optimal linear precoding (OLP) scheme that minimizes the total transmit power while satisfying a set of target UE rates [5]–[7]. This problem is receiving a renewed interest nowadays due to the emerging research area of green cellular networks [8]. In particular, we consider the downlink of a single-cell multi-user MIMO system in which the BS makes use of M antennas to communicate with K single-antenna UEs under the assumption that perfect channel state information is available. Within this setting, the structure of OLP is well-known and can be derived using different approaches [9]–[11]. As most linear precoding schemes (a notable exception is the maximum ratio transmission scheme), a major difficulty towards the implementation of OLP is that it requires to compute matrix inversions in every channel coherence period with a computational complexity proportionally to MK^2. This makes it unsuited for scenarios with highly varying channels or with large values of M and K.

A possible solution to overcome this issue relies on using a truncated polynomial expansion (TPE) of the matrix inverse. Roughly speaking, the idea of TPE is to approximate the matrix inverse by a matrix polynomial with J terms, where J needs not to scale with the system dimensions to maintain a certain approximation accuracy. Applications of TPE to linear precoding schemes for large-scale MIMO systems can be found in [12], [13]. In [12], the J polynomial coefficients are designed so as to maximize the achievable rate subject to a power constraint. In [13], the authors aim at mimicking the performance of the RZF scheme presented in [14], [15]. It is worth observing that TPE techniques have been also applied in other contexts such as channel estimation [16] and multi-user detection [17], [18].

All the aforementioned works assume that the same TPE coefficients apply to all UEs. This reduces the degrees of freedom and results in some performance loss for small values of J. Moreover, it allows to apply the TPE technique only to some particular cases [12], [13]. To overcome this drawback, in this work the TPE concept is applied to OLP on a per-UE basis. In other words, the polynomial truncation artifice is applied separately to each precoding vector rather than to the all precoding matrix. This allows to formulate the optimization problem of the weighting coefficients in a convex form. A further reduction in complexity is obtained by assuming that M and K grow large with a bounded ratio. Such an assumption allows us to approximate the signal-to-noise-plus-interference ratio (SINR) and the transmit power by deterministic quantities that depend only on the channel statistics rather than on its instantaneous realizations [12], [19], [20]. All this leads to a novel TPE precoder scheme, which is shown (by means of numerical results) to achieve the same performance of OLP but with much lower complexity.

The remainder of this work is organized as follows. Next section introduces the system model. Section III formulates the power minimization problem and revises the structure of OLP. Section IV introduces the main concept of TPE and derives the

L. Sanguinetti and M. Debbah were supported by the ERC Starting Grant 305123 MORE. L. Sanguinetti was also partially supported by the research project 5GIOTTO funded by the University of Pisa.

proposed precoding scheme. Numerical results are shown in Section V while some conclusions and implications are drawn in Section VI.

II. System Model

We consider a single-cell large-scale MIMO system in which the BS equipped with M antennas serves K single antenna UEs. The K active UEs are randomly selected from a larger set of UEs within the coverage area. The location of UE k is characterized by its own distance d_k from the BS. The large-scale channel fading corresponding to UE k is assumed to be dominated by the pathloss and is denoted by β_k. We further assume that β_k is the same for all BS antennas. Such an assumption is reasonable since the distances among BS and UEs are much larger than the distances among BS antennas. The channel vector $\mathbf{h}_k$ associated to UE k is thus modeled as:

$$\mathbf{h}_k = \sqrt{\beta_k}\mathbf{z}_k \quad (1)$$

where $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$. For notational convenience, we collect the channel attenuation coefficients into the diagonal matrix $\mathbf{D} = \mathrm{diag}(\beta_1, \ldots, \beta_K)$ and the UEs channel vectors in the matrix $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$.

We assume that the BS employs Gaussian codebook and linear precoding. The precoding matrix is denoted by $\mathbf{G} = [\mathbf{g}_1, \cdots, \mathbf{g}_K]$ while the data symbol vector for all UEs is called $\mathbf{s} = [s_1, \cdots, s_K]^{\mathrm{T}}$ with $s_k \sim \mathcal{CN}(0, 1)$. The transmit signal vector $\mathbf{x}$ is $\mathbf{x} = \mathbf{G}\mathbf{s}$ whereas the received signal at UE k can be written as:

$$y_k = \mathbf{h}_k^H \mathbf{g}_k s_k + \sum_{i=1, i\neq k}^{K} \mathbf{h}_k^H \mathbf{g}_i s_i + n_k \quad (2)$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ accounts for thermal noise. The SINR at UE k is thus given by:

$$\mathrm{SINR}_k = \frac{|\mathbf{h}_k^{\mathrm{H}} \mathbf{g}_k|^2}{\sum_{i=1, i\neq k}^{K} |\mathbf{h}_k^{\mathrm{H}} \mathbf{g}_i|^2 + \sigma^2} \quad (3)$$

whereas the total transmit power at the BS is computed as $P = \mathrm{tr}(\mathbf{G}\mathbf{G}^{\mathrm{H}})$.

III. Problem statement

The objective of this work is to design a low-complexity linear precoding scheme that minimizes the transmit power at the BS while maintaining target user rates $\{r_k\}$. More formally, the power minimization problem is formulated as:

$$\begin{aligned}\mathcal{P}: \ &\underset{\mathbf{G}}{\text{minimize}} \quad \mathrm{tr}(\mathbf{G}\mathbf{G}^{\mathrm{H}}) \\ &\text{subject to} \quad \mathrm{SINR}_k \geq \gamma_k \quad k = 1, \cdots, K.\end{aligned}$$

where $\gamma = [\gamma_1, \ldots, \gamma_K]^T$ is the vector of target SINRs given by $\gamma_k = 2^{r_k} - 1$ (under the assumption of Gaussian codebooks). The solution of ($\mathcal{P}$) is well-known and given by [5]–[7]:

$$\mathbf{G}_{\mathrm{OLP}} = \left(\sum_{i=1}^{K} q_i^* \mathbf{h}_i \mathbf{h}_i^{\mathrm{H}} + M\mathbf{I}_M \right)^{-1} \mathbf{H}\sqrt{\mathrm{diag}\left(\mathbf{p}^*\right)} \quad (4)$$

where $\{q_k^*\}$ are obtained as the unique positive solution to the following set of fixed point equations:

$$\left(1 + \frac{1}{\gamma_k}\right) q_k^* = \frac{1}{\mathbf{h}_k^{\mathrm{H}} \left(\sum_{i=1}^{K} q_i^* \mathbf{h}_i \mathbf{h}_i^{\mathrm{H}} + M\mathbf{I}_M \right)^{-1} \mathbf{h}_k} \ \forall k \quad (5)$$

while $\mathbf{p}^* = [p_1^*, \ldots, p_K^*]^T$ is such that the SINR constraints are all satisfied with equality. This yields $\mathbf{p}^* = \sigma^2 \mathbf{A}^{-1} \mathbf{1}_K$ where the (k, i)-th element of $\mathbf{A}$ is

$$[\mathbf{A}]_{k,i} = \begin{cases} \frac{1}{\gamma_k} |\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_k^*|^2 & \text{if} \quad k = i \\ -|\mathbf{h}_k^{\mathrm{H}} \mathbf{v}_i^*|^2 & \text{if} \quad k \neq i \end{cases} \quad (6)$$

with $\mathbf{v}_k^*$ being the k-th column of $\mathbf{V}^* = (\sum_{i=1}^{K} q_i^* \mathbf{h}_i \mathbf{h}_i^{\mathrm{H}} + M\mathbf{I})^{-1}\mathbf{H}$. The major difficulties towards the implementation of $\mathbf{G}_{\mathrm{OLP}}$ are as follows. Firstly, $\mathbf{G}_{\mathrm{OLP}}$ is parameterized by $\mathbf{q}^*$ and $\mathbf{p}^*$. Both require matrix inversions every coherence period. As for $\mathbf{q}^*$, it needs also to be evaluated by an iterative procedure due to the fixed-point equations in (5). Once $\mathbf{q}^*$ and $\mathbf{p}^*$ are computed, the application of $\mathbf{G}_{\mathrm{OLP}}$ to the data vector requires a matrix inversion. All this is a computationally demanding task when M and K grow large as envisioned in large-scale MIMO systems.

A possible way to reduce the complexity in the computation of $\mathbf{q}^*$ and $\mathbf{p}^*$ is to resort to the asymptotic analysis. Indeed, when M and K grow large with a bounded ratio we have that [10], [11], [15]

$$\max_{1\leq k\leq K} |q_k^* - \bar{q}_k| \xrightarrow{\text{a.s.}} 0 \quad (7)$$

$$\max_{1\leq k\leq K} |p_k^* - \bar{p}_k| \xrightarrow{\text{a.s.}} 0 \quad (8)$$

with

$$\bar{q}_k = \frac{\gamma_k}{\beta_k \xi} \quad (9)$$

$$\bar{p}_k = \frac{\gamma_k}{\beta_k \xi^2} \left(\bar{P} + \frac{\sigma^2}{\beta_k}(1+\gamma_k)^2 \right) \quad (10)$$

where $\bar{P} = \frac{\sigma^2 \sum_{i=1}^{K} \frac{\gamma_i}{\beta_i}}{M\xi}$ and $\xi = 1 - \frac{1}{M}\sum_{i=1}^{K} \frac{\gamma_i}{1+\gamma_i}$. From the above results, it follows that both $\mathbf{q}^*$ and $\mathbf{p}^*$ admit simple closed-form deterministic approximations that depend solely on the large-scale channel statistics. This information can be easily observed and estimated accurately at the BS because it changes slowly with time (relative to the small-scale fading).

Replacing $\{q_k^*\}$ and $\{p_k^*\}$ with $\{\bar{q}_k\}$ and $\{\bar{p}_k\}$ yields the so-called asymptotically OLP (A-OLP) given by:

$$\mathbf{G}_{\mathrm{A-OLP}} = \left(\sum_{i=1}^{K} \bar{q}_i \mathbf{h}_i \mathbf{h}_i^{\mathrm{H}} + M\mathbf{I}_M \right)^{-1} \mathbf{H}\sqrt{\mathrm{diag}\left(\bar{\mathbf{p}}\right)} \quad (11)$$

where $\bar{\mathbf{p}} = [\bar{p}_1, \ldots, \bar{p}_K]^T$. Although the parameters $\{\bar{q}_k\}$ and $\{\bar{p}_k\}$ are no longer computed at every channel realization, the computation of $\mathbf{G}_{\mathrm{A-OLP}}$ can be a task of a prohibitively high complexity when M and K are large. To address this issue, a TPE approach will be adopted in the next section.

IV. User Specific TPE Precoding Design

The common way to apply the TPE concept to precoding design consists in replacing the matrix inverse by a weighted matrix polynomial with J terms [12], [13], [20]. Differently from the traditional approach, we propose to apply the polynomial truncation artifice separately to each vector of the precoding scheme.

A. Principle of User Specific TPE (US-TPE) precoding

Inspired by [12], [13], [20], the TPE of the precoding vector associated with the k-th UE is computed as:

$$\mathbf{g}_k = \sum_{\ell=0}^{J-1} w_{\ell,k} \left(\frac{\mathbf{H}\mathbf{R}\mathbf{H}^H}{K}\right)^{\ell} \frac{\mathbf{h}_k}{K} \quad \forall k \tag{12}$$

where J is the truncation order and $\mathbf{R} = \mathrm{diag}(\bar{q}_1, \cdots, \bar{q}_K)$. Plugging $\mathbf{g}_k$ in (12) into (3) and denoting by $\mathbf{w}_k = [w_{0,k}, \cdots, w_{J-1,k}]^T$, the SINR associated with the k-th UE can be written as:

$$\mathrm{SINR}_k = \frac{\mathbf{w}_k^H \mathbf{a}_k \mathbf{a}_k^H \mathbf{w}_k}{\sum_{i\neq k} \frac{1}{K} \mathbf{w}_i^H \mathbf{B}_{k,i} \mathbf{w}_i + \sigma^2} \quad \forall k \tag{13}$$

where $\mathbf{a}_k \in \mathbb{C}^J$ and $\mathbf{B}_{k,i} \in \mathbb{C}^{J\times J}$ are random quantities given by:

$$[\mathbf{a}_k]_\ell = \frac{1}{K}\mathbf{h}_k^H \left(\frac{\mathbf{H}\mathbf{R}\mathbf{H}^H}{K}\right)^{\ell} \mathbf{h}_k \tag{14}$$

$$[\mathbf{B}_{k,i}]_{\ell,m} = \frac{1}{K}\mathbf{h}_k^H \left(\frac{\mathbf{H}\mathbf{R}\mathbf{H}^H}{K}\right)^{\ell} \mathbf{h}_i \mathbf{h}_i^H \left(\frac{\mathbf{H}\mathbf{R}\mathbf{H}^H}{K}\right)^{m} \mathbf{h}_k. \tag{15}$$

The transmit power at the BS can be expressed as:

$$P = \frac{1}{K}\sum_{k=1}^{K} \mathbf{w}_k^H \mathbf{E}_k \mathbf{w}_k \tag{16}$$

where $\mathbf{E}_k \in \mathbb{C}^{J\times J}$ has entries given by

$$[\mathbf{E}_k]_{\ell,m} = \frac{1}{K}\mathbf{h}_k^H \left(\frac{\mathbf{H}\mathbf{R}\mathbf{H}^H}{K}\right)^{(\ell+m)} \mathbf{h}_k. \tag{17}$$

As M and K grow simultaneously large, we have that (the details are omitted for space limitations, only the main steps are described in Appendix)

$$\mathrm{SINR}_k - \overline{\mathrm{SINR}}_k \xrightarrow{\text{a.s.}} 0, \quad P - \overline{P} \xrightarrow{\text{a.s.}} 0 \tag{18}$$

with

$$\overline{\mathrm{SINR}}_k = \frac{\mathbf{w}_k^H \tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^H \mathbf{w}_k}{\sum_{i\neq k} \frac{1}{K} \mathbf{w}_i^H \tilde{\mathbf{B}}_{k,i} \mathbf{w}_i + \sigma^2} \tag{19}$$

and

$$\overline{P} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{w}_k^H \tilde{\mathbf{E}}_k \mathbf{w}_k \tag{20}$$

where the entries of $\tilde{\mathbf{a}}_k \in \mathbb{C}^J$, $\tilde{\mathbf{B}}_{k,i} \in \mathbb{C}^{J\times J}$ and $\tilde{\mathbf{E}}_k \in \mathbb{C}^{J\times J}$ are given in Appendix and depend only on the pathloss coefficients. The asymptotic analysis above is used in the sequel to compute the weighting vector $\{\mathbf{w}_k\}$ that minimizes the transmit power while satisfying SINR constraints.

B. Optimization of the US-TPE precoding

We look for the solution of the following optimization problem:

$$\mathcal{P}_1: \underset{\mathbf{w}_1,\cdots,\mathbf{w}_K}{\text{minimize}} \ \frac{1}{K}\sum_{k=1}^{K} \mathbf{w}_k^T \tilde{\mathbf{E}}_k \mathbf{w}_k \tag{21}$$
$$\text{subject to} \ \frac{1}{\gamma_k}\mathbf{w}_k^T \tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^T \mathbf{w}_k \geq \sum_{i\neq k} \frac{1}{K}\mathbf{w}_i^T \tilde{\mathbf{B}}_{k,i} \mathbf{w}_i + \sigma^2$$

which can be easily shown to be non-convex. To reformulate it in a convex form, we use the same trick of [5] (see also [7]). In particular, we note that the power and the SINRs keep the same values if $\forall k$ $\mathbf{w}_k$ is replaced by $-\mathbf{w}_k$. Therefore, we can assume that $\mathbf{w}_k^T \tilde{\mathbf{a}}_k$ is positive $\forall k$. It is worth mentioning that this artifice cannot be used if the same weight vector is used for all UEs, i.e., $\forall k$ $\mathbf{w}_k = \mathbf{w}$. This easily follows from observing that if $\forall k$ $\mathbf{w}_k = \mathbf{w}$ then we cannot ensure that $\forall k$ $\mathbf{w}^T \tilde{\mathbf{a}}_k > 0$. This explains why the traditional TPE can be applied only to some particular cases [12], [13]. By rewriting the SINR constraints as [7]:

$$\frac{1}{\gamma_k}\mathbf{w}_k^T \tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^T \ \mathbf{w}_k \geq \gamma_k \Big(\sum_{i\neq k} \frac{1}{K}\mathbf{w}_i^T \tilde{\mathbf{B}}_{k,i}\mathbf{w}_i + \sigma^2\Big) \Leftrightarrow$$
$$\frac{1}{\sqrt{\gamma_k \sigma^2}} \tilde{\mathbf{a}}_k^T \ \mathbf{w}_k \geq \sqrt{\sum_{i\neq k} \frac{1}{\sigma^2 K}\mathbf{w}_i^T \tilde{\mathbf{B}}_{k,i}\mathbf{w}_i + 1} \tag{22}$$

the problem $\mathcal{P}_1$ becomes convex. Moreover, it can be easily shown that the Slater condition is fulfilled [7]. Therefore, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for the reformulated problem and thus for the original one [7]. The Lagrangian of $\mathcal{P}_1$ is defined as:

$$\mathcal{L}(\mathbf{w}_1, \cdots, \mathbf{w}_K, \lambda_1, \cdots, \lambda_K) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{w}_k^T \tilde{\mathbf{E}}_k \mathbf{w}_k +$$
$$\sum_{k=1}^{K} \lambda_k \Big(\sum_{i\neq k} \frac{1}{\sigma^2 K}\mathbf{w}_i^T \tilde{\mathbf{B}}_{k,i}\mathbf{w}_i + 1 - \frac{1}{\gamma_k \sigma^2}\mathbf{w}_k^T \tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^T \mathbf{w}_k\Big) \tag{23}$$

It is known that the optimal $\{\mathbf{w}_k\}$ satisfy $\forall k$ $\frac{\partial \mathcal{L}}{\partial \mathbf{w}_k} = 0$. After some algebraic manipulations, we get

$$\mathbf{w}_k = \frac{K\lambda_k}{\sigma^2 \gamma_k}\Big(\tilde{\mathbf{E}}_k + \sum_{i\neq k} \frac{\lambda_i}{\sigma^2} \tilde{\mathbf{B}}_{i,k}\Big)^{-1} \tilde{\mathbf{a}}_k (\tilde{\mathbf{a}}_k^T \mathbf{w}_k). \tag{24}$$

Since $\mathbf{a}_k^T \mathbf{w}_k$ is a scalar, the optimal $\mathbf{w}_k$ corresponding to UE k is proportional to the vector $\big(\tilde{\mathbf{E}}_k + \sum_{i\neq k} \frac{\lambda_i}{\sigma^2}\tilde{\mathbf{B}}_{i,k}\big)^{-1}\mathbf{a}_k$. The positive scalars λ_k are obtained as the unique solution of the following set of equations:

$$\lambda_k = \frac{\gamma_k \sigma^2}{K \mathbf{a}_k^T \left(\tilde{\mathbf{E}}_k + \sum_{i\neq k} \frac{\lambda_i}{\sigma^2}\tilde{\mathbf{B}}_{i,k}\right)^{-1} \tilde{\mathbf{a}}_k}. \tag{25}$$

Putting all the above results together, the optimal vector $\mathbf{w}_k^*$ is found to be:

$$\mathbf{w}_k^* = \sqrt{p_k^*}\frac{\left(\tilde{\mathbf{E}}_k + \sum_{i\neq k}\frac{\lambda_i}{\sigma^2}\tilde{\mathbf{B}}_{i,k}\right)^{-1}\tilde{\mathbf{a}}_k}{\left\|\left(\tilde{\mathbf{E}}_k + \sum_{i\neq k}\frac{\lambda_i}{\sigma^2}\tilde{\mathbf{B}}_{i,k}\right)^{-1}\tilde{\mathbf{a}}_k\right\|} = \sqrt{p_k^*}\bar{\mathbf{w}}_k \quad (26)$$

where $\bar{\mathbf{w}}_k$ are the beamforming directions and p_k^* are such that the SINR constraints in $\mathcal{P}_1$ are all satisfied with equality:

$$p_k^*\bar{\mathbf{w}}_k^T\tilde{\mathbf{a}}_k\tilde{\mathbf{a}}_k^T\bar{\mathbf{w}}_k = \gamma_k\sum_{i\neq k}\frac{p_i^*}{K}\bar{\mathbf{w}}_i^T\tilde{\mathbf{B}}_{k,i}\bar{\mathbf{w}}^T + \gamma_k\sigma^2 \quad \forall k. \quad (27)$$

This yields:

$$\mathbf{p}^* = \sigma^2\mathbf{F}^{-1}\mathbf{1}_K \quad (28)$$

where the (k,i)-th element of the matrix $\mathbf{F}$ is

$$[\mathbf{F}]_{k,i} = \begin{cases}\frac{1}{\gamma_k}\bar{\mathbf{w}}_k^T\tilde{\mathbf{a}}_k\tilde{\mathbf{a}}_k^T\bar{\mathbf{w}}_k & \text{if} \quad k=i\\ -\frac{1}{K}\bar{\mathbf{w}}_i^T\tilde{\mathbf{B}}_{k,i}\bar{\mathbf{w}}_i^T & \text{if} \quad k\neq i.\end{cases} \quad (29)$$

The precoding vectors obtained as in (26) are referred to as US-TPE in the sequel.

V. Numerical results

Monte Carlo simulations are now used to make comparisons among US-TPE, OLP and the TPE precoding proposed in [13]. The results are obtained averaging over 1000 different channel realizations. We assume that UEs are uniformly distributed in a cell with radius $D = 250$ m and minimum distance $d_{\min} = 35$ m. The pathloss β_k of UE k is modeled as:

$$\beta_k = \frac{d_0}{\|d_k\|^\eta} \quad (30)$$

where d_k accounts for the distance from the BS whereas $\eta = 3.6$ is the pathloss coefficient and $d_0 = 10^{-3.53}$ is a constant that regulates the channel attenuation at $d_{\min}$. Besides, the transmission bandwidth is fixed to $W = 20$ MHz and the total noise power $W\sigma^2$ is -97.8 dBm. The UEs target rates $\{r_k\}$ are randomly chosen from the interval [0.1, 3] bits/sec/Hz.

Fig. 1 plots the total transmit power vs. the number of BS antennas M of all the investigated precoders. Comparisons are also made with the PA-RZF precoding proposed in [14]. As it is seen, US-TPE with $J = 3$ requires almost the same amount of power of OLP and provides a marginal saving with respect to both PA-RZF in [14] and TPE in [13]. As far as computational complexity is concerned, it requires, like the traditional TPE, $\mathcal{O}(KM)$ arithmetic operations. This has to be compared with the PA-RZF and the OLP, which involve (K^2M) arithmetic operations per coherence period.

Fig. 2 illustrates the total transmit power vs. M for different values of the truncation order J. Obviously, as J increases, the average transmitted power of the US-TPE decreases, but with a slower rate. From the results of Fig. 2, choosing $J = 3$ seems to be a good option since it leads to sufficiently close performance as the OLP.

The impact of UE target rates is investigated in Fig 3, where we report the transmit power vs. the maximum allowed rate

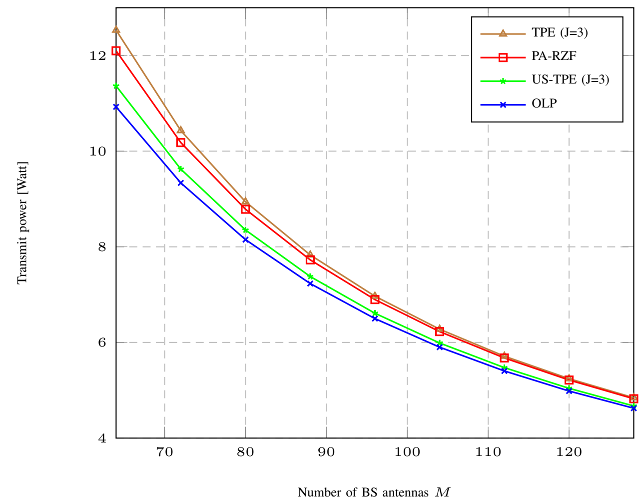

Fig. 1. Average per UE transmit power in Watt vs. M when $K = 32$.

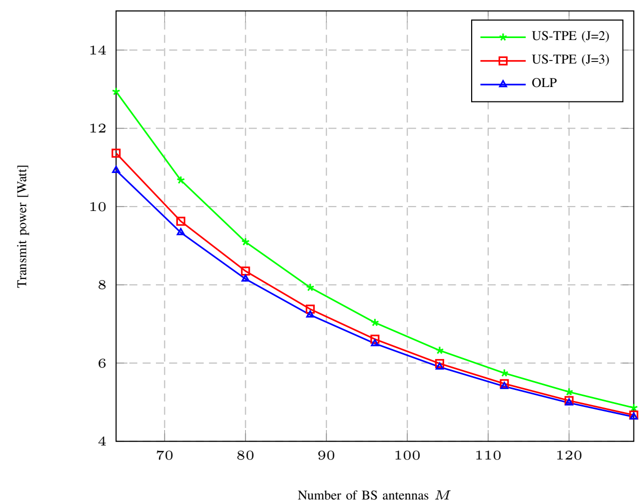

Fig. 2. Transmit power in Watt vs. M when $K = 32$.

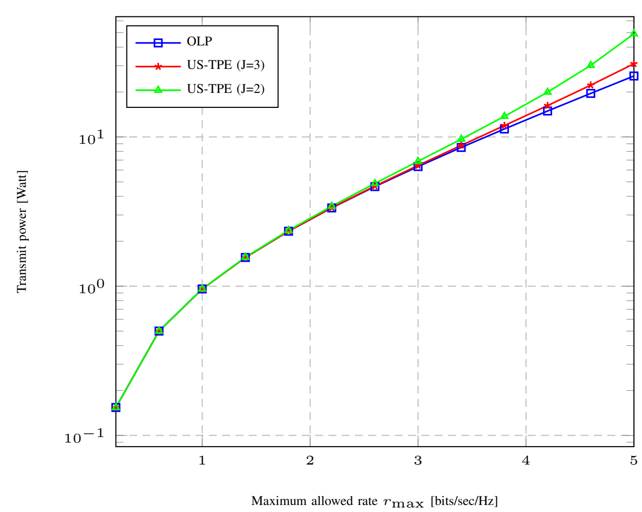

Fig. 3. Transmit power in Watt vs. maximal allowed rate $r_{\max}$ when $M = 128$ and $K = 32$.

$r_{\max}$. The UEs rates are chosen randomly in the interval [0.1, $r_{\max}$]. As seen, for values of $r_{\max}$ smaller than 4 [bit/s/Hz], $J = 3$ is enough to achieve the same performance of OLP. On the other hand, higher values of the truncation order J must be used to further enhance performance when $r_{\max} \geq 4$.

VI. Conclusions

In this work, we focused on the design of a linear TPE precoder to achieve the same performance of the optimal linear precoder (OLP) minimizing the total power consumption while ensuring target UE rates. Differently from traditional TPE schemes, the proposed one applied the truncated polynomial expansion concept on a per-UE basis. This allowed to determine in closed-form the optimal TPE coefficients that approximate with the best accuracy the performance of OLP. In order to further facilitate the design of the proposed precoder, we considered the asymptotic regime in which the number of UEs and BS antennas grow simultaneously large with a bounded ratio. Such an assumption allows to approximate the transmit power and the SINRs by deterministic quantities that depend only on the channel statistics. Numerical results showed that the proposed TPE provided the same power consumption of OLP under different operating conditions.

Appendix

Here, we provide the details for the computation of vectors $\tilde{\mathbf{a}}_k$ and matrices $\tilde{\mathbf{B}}_{k,i}$ and $\tilde{\mathbf{E}}_k$ described in section IV. The determiniation of closed-form expressions for these quantities is based on the observation that they are related to a set of deterministic quantities that are extensively used in random matrix theory. As a matter of fact, let $\left(\delta(t), \tilde{\delta}(t)\right)$ be the unique solution of the following system of equations [adapted from [21]]:

$$\begin{cases} \delta(t) &= \frac{M}{K}\left(1 + t\tilde{\delta}(t)\right)^{-1} \\ \tilde{\delta}(t) &= \frac{1}{K}\operatorname{tr}\left(\mathbf{R}\mathbf{D}\left(\mathbf{I}_K + t\delta(t)\mathbf{R}\mathbf{D}\right)^{-1}\right) \end{cases}$$

and consider $\mathbf{T}(t) = (1 + \frac{t}{1+t\delta(t)})^{-1}\mathbf{I}_M$ and $\tilde{\mathbf{T}}(t) = (\mathbf{I}_K + t\delta(t)\mathbf{R}\mathbf{D})^{-1}$. Then, the elements of $\tilde{\mathbf{a}}_k$, $\tilde{\mathbf{B}}_{k,i}$ and $\tilde{\mathbf{E}}_k$ are given by:

$$[\tilde{\mathbf{a}}_k]_\ell = \frac{(-1)^\ell}{\ell!}\beta_k\delta^{(\ell)} \tag{31}$$

$$[\tilde{\mathbf{E}}_k]_{\ell,m} = \frac{(-1)^{\ell+m}}{(\ell+m)!}\beta_k\delta^{(\ell+m)} \tag{32}$$

$$[\tilde{\mathbf{B}}_{k,i}]_{\ell,m} = \frac{(-1)^{\ell+m}}{\ell!m!}Y_{k,i}^{(\ell,m)} \tag{33}$$

where $\delta^{(\ell)}$ is the ℓ-th derivative of $\delta(t)$ at $t = 0$ and $Y_{k,i}^{(\ell,m)}$ is the (ℓ, m)-th derivative at $(t = 0, u = 0)$ of $Y_{k,i}(t, u)$ defined as :

$$Y_{k,i}(t,u) = \frac{1}{K}g_{k,i}(t)g_{k,i}(u)\phi(t,u) \tag{34}$$

where

$$g_{k,i}(t) = \frac{\sqrt{\beta_k\beta_i}}{(1 + t\bar{q}_k\beta_k\delta(t))(1 + t\bar{q}_i\beta_i\delta(t))} \tag{35}$$

$$\phi(t,u) = \frac{\frac{1}{K}\operatorname{tr}\left(\mathbf{T}(u)\mathbf{T}(t)\right)}{1 - tu\psi(t,u)} \tag{36}$$

and

$$\psi(t,u) = \frac{1}{K^2}\operatorname{tr}\left(\mathbf{T}(u)\mathbf{T}(t)\right)\operatorname{tr}\left(\mathbf{R}^2\mathbf{D}^2\tilde{\mathbf{T}}(u)\tilde{\mathbf{T}}(t)\right). \tag{37}$$

The computation of $\tilde{\mathbf{a}}_k$, $\tilde{\mathbf{B}}_{k,i}$ and $\tilde{\mathbf{E}}_k$ requires the numerical evaluation of the derivatives of $\delta(t)$, $\mathbf{T}$ and $\tilde{\mathbf{T}}$ at $t = 0$, which can be obtained by using the iterative algorithm given in [12], [17]. However, unlike $\tilde{\mathbf{a}}_k$ and $\tilde{\mathbf{E}}_k$, the computation of $Y_{k,i}(t, u)$ arising in the expression of $\tilde{\mathbf{B}}_{k,i}$ is not immediate, requiring further derivations based on the use of the Liebniz rule.

Indeed, using Leibniz formula, $Y_{k,i}^{(\ell,m)}$ can be computed as:

$$Y_{k,i}^{(\ell,m)} = \sum_{j=0}^{\ell}\sum_{n=0}^{m}\binom{\ell}{j}\binom{m}{n}\frac{1}{K}g_{k,i}^{(j)}g_{k,i}^{(n)}\phi^{(\ell-j,m-n)}$$

where $g_{k,i}^{(\ell)}, \ell = 0, 1, 2, ..,$ denote the derivatives of $g_{k,i}(t)$ at $t = 0$ and $\phi^{(n,m)}$ denote that of $\phi(t, u)$ at $t = u = 0$. The derivation of $g_{k,i}^{(\ell)}$ are obtained using the Leibniz formula, whereas $\phi^{(n,m)}$ can be evaluated iteratively using the following relation:

$$\phi^{(n,m)} = \frac{1}{K}\operatorname{tr}\left(\mathbf{T}^{(n)}\mathbf{T}^{(m)}\right) + \sum_{p_n=1}^{n}\sum_{p_m=1}^{m}p_np_m\binom{n}{p_n}\binom{m}{p_m}\phi^{(p_n-1,p_m-1)}\psi^{(n-p_n,m-p_m)}$$

where the derivatives of $\psi(t, u)$ at $(0, 0)$ are given by:

$$\psi^{(n,m)} = \sum_{p_n=0}^{n}\sum_{p_m=0}^{m}\binom{n}{p_n}\binom{m}{p_m}\frac{1}{K^2}\operatorname{tr}\left(\mathbf{T}^{(p_m)}\mathbf{T}^{(p_n)}\right)\operatorname{tr}\left(\mathbf{R}^2\mathbf{D}^2\tilde{\mathbf{T}}^{(m-p_m)}\tilde{\mathbf{T}}^{(n-p_n)}\right).$$

References

[1] T.L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] 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, Feb. 2013.

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

[4] 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, Feb. 2014.

[5] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*. L. C. Godara, Ed CRC Press, 2001.

[6] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Transactions on Signal Processing*, vol. 54, pp. 161–176, 2006.

[7] E. Björnson, E. G. Larsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Processing Mag.*, vol. 13, pp. 142–148, July 2014.

[8] Y. Chen, S. Zhang, S. Xu, and G.Y Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, 2011.

[9] H. Dahrouj and Wei Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748 – 1759, May 2010.

[10] R. Zakhour and S.V. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.

[11] L. Sanguinetti, R. Couillet, and M. Debbah, "Large system analysis of base station cooperation for power minimization," *IEEE Trans. Wireless Commun.*, Sept. 2015, submitted, [Online] http://arxiv.org/abs/1509.00731.

[12] A. Müller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on truncated polynomial expansion: Reducing complexity in massive MIMO," *EURASIP Journal on Wireless Communications and Networking*, 2015.

[13] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, and M-S. Alouini, "Power efficient low complexity precoding for massive MIMO systems," in *GlobalSIP 2014, Atalanta, Georgia*, 2014.

[14] L. Sanguinetti, E. Björnson, M. Debbah, and A.L Moustakas, "Optimal linear pre-coding in multi-user MIMO systems: A large system analysis," in *Proc. IEEE GLOBECOM*, Austin, USA, Dec. 2014.

[15] L. Sanguinetti, A.L. Moustakas, E. Bjornson, and M. Debbah, "Large system analysis of the energy consumption distribution in multi-user MIMO systems with mobility," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1730 – 1745, March 2015.

[16] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity channel estimation in large-scale MIMO using polynomial expansion," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun. (PIMRC)*, 2013.

[17] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Proc. Int. Symp. Inf. Theory (ISIT)*, 2011.

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

[19] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in miso broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[20] A. Kammoun, A. Müller, E. Björnson, and M. Debbah, "Linear precoding based on truncated polynomial expansion: Large-scale multi-cell systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 861–875, Oct. 2014.

[21] W. Hachem, O. Khorunzhy, P. Loubaton, J. Najim, and L. A. Pastur, "A new approach for capacity analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sept. 2008.