Asymptotic Analysis of RLS-based Digital Precoder with Limited PAPR in Massive MIMO

Xiuxiu Ma, Abla Kammoun, Member, IEEE, Ayed M. Alrashdi, Member, IEEE, Tarig Ballal, Member, IEEE, Tareq Y. Al-Naffouri, Senior Member, IEEE and Mohamed-Slim Alouini, Fellow, IEEE


ded, conventional precoding approaches based on simple linear precoders such as maximum ratio transmission (MRT) and regularized zero-forcing (RZF), the precoders in this paper are obtained by solving a convex optimization problem. To be specific, these precoders are designed so that the power of each precoded symbol entry is restricted, and the PAPR at each antenna can be tunable. By using the Convex Gaussian Min-max Theorem (CGMT), we analytically characterize the empirical distribution of the precoded vector and the vector's empirical distribution of the precoder vector. The main difference with RZF is that it constrains the absolute value of the precoded vector to not exceed a certain threshold. Referred to as the RLS-based precoder with limited peak-to-average power ratio (PAPR), the precoder in [9] allows for achieving the two sought-for goals, that is a lower number of RF chains together with a limited PAPR, making it possible to use inexpensive power amplifiers.

A. Contributions and related works

Performance analysis of non-linear precoders. In this paper, we carry out a rigorous, asymptotic characterization of the performance of multi-user downlink transmission when an RLS precoder with limited PAPR is employed. More precisely, we study the asymptotic behavior of the precoder with limited peak-to-average distortion power, per-user distortion power, signal-to-noise and distortion ratio (SINAD), bit error probability when the number of antennas and the number of served users grow large at the same pace. A similar problem has been recently studied in [11], [12] where asymptotic expressions for the distortion error and a lower bound on the achievable rate have been derived. Compared to these works, our contribution differs as follows. On a methodological level, while the works in [11], [12] are based on the non-rigorous replica method, the main tool in our work is the recently developed Convex Gaussian Min-max Theorem (CGMT) [13] framework. Using the CGMT, our analysis goes beyond the performance metrics studied in [11], [12]. Particularly, we assume BPSK modulation while [11], [12] rely on Gaussian signaling. Furthermore, we derive accurate characterization of the joint distribution between the transmitted symbol vector and the distortion error. This characterization allows us to analyze the bit error probability and a tight approximation of the SINAD. On an operational level, we derive several insights from our analysis by studying the obtained asymptotic expressions in different regimes describing small numbers of served users or small/large values of the power control parameter. Particularly, we show that the reflect-array and transmit-array antennas scheme. Aside from power consumption, it is of interest to control the power of each RF chain allowing for cheap system modules. In light of this observation, the work in [9] proposes a precursor with fewer RF chains that constrains the power of each signal entry to be below a certain threshold.

The proposed precoder in [9] reminds the conventional regularized zero-forcing precoder (RZF) [10], that in it builds on the regularized least squares (RLS) method to minimize a penalty of the residue sum of squares (RSS). The main difference with RZF is that it constrains the absolute value of the precoded vector to not exceed a certain threshold. Referred to as the RLS-based precoder with limited peak-to-average power ratio (PAPR), the precoder in [9] allows for achieving the two sought-for goals, that is a lower number of RF chains together with a limited PAPR, making it possible to use inexpensive power amplifiers.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems are recognized among the key enabling technologies for next-generation communication systems [1]–[3]. However, there are still major implementation issues to address for massive MIMO systems to become a reality. First, the number of antennas that can be supported is limited by the transceiver’s form factor. In practice, this issue can be handled by moving the operating frequency to mmWave frequency bands [4]. Second, it requires equipping each antenna with a dedicated radio frequency (RF) chain, which allows the pass-band communication signals to be processed in the base-band [5], thereby leading to a prohibitively high cost and power consumption, calling into question the practicality of such systems. One possible solution is to reduce the number of RF chains by employing hybrid-precoding [6], [7]. However, the work in [8] shows that the power consumption of some hybrid precoding architectures still scales with the number of antennas, and proposed as a solution a linear and highly energy efficient

X. Ma, A. Kammoun, T. Ballal, T. Y. Al-Naffouri and M. Alouini are with the Division of Computer, Electrical and Mathematical Science & Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, KSA. A. M. Alrashdi is with the Department of Electrical Engineering, College of Engineering, University of Ha’il, PO. Box 2440, Ha’il, 81441, Saudi Arabia. E-mail: {xiuxiu.ma; abla.kammoun; tarig.ahmed; tareq.alnaffouri; slim.alouini} @kaust.edu.sa; am.alrashdi@uoh.edu.sa}

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information. Authorized licensed use limited to: KAUST. Downloaded on November 02,2022 at 08:42:33 UTC from IEEE Xplore. Restrictions apply.
performance of the RLS precoder with limited PAPR is not always better when the transmit per-antenna power increases, for higher transmit power may also imply higher distortion power. In other words, there is an optimal per-antenna transmit power that maximizes the performance in terms of SINAD and bit error probability. It can be achieved by properly setting the power control parameter.

**Convex Gaussian Min-max Theorem.** The main ingredient of the proof of our main results is the Convex Gaussian Min-max Theorem (CGMT). This framework has been initiated by Stojnic [14] before being formally developed in [13] and [15]. It has been applied to characterize the asymptotic behavior of convex-optimization-based estimators with application to high-dimensional regression problems as well as binary classification problems. In this line, the work in [15] applied the CGMT to quantify the performances of several estimators including the Least Absolute Shrinkage and Selection Operator (LASSO). As far as wireless communications are concerned, the CGMT has been applied to characterize the performance of non-explicit decoders. In this context, under the assumption of real Gaussian channels, the CGMT was used to derive closed-form approximations of the bit error probability of convex-optimization-based decoders termed box relaxation decoders under Binary Phase Shift Keying (BPSK) signaling [16], [17] as well as M-ary Pulse Amplitude Modulation (MAPAM) signaling [18]. All these works have focused on the design of non-linear algorithms from the decoder perspective. The application of the CGMT for the design of non-linear precoders has not, to the best of our knowledge, been studied, which motivates our work.

**B. Paper Organization**

In Section II, we introduce the system model and formulate the problem. Then, in Section III we state our main results characterizing the statistics of the precoded vector and the distortion, based on which we get the asymptotic behavior of the studied specifications and performance metrics, namely, the transmit per-antenna power, the per-user distortion power, the received SINAD, and the bit error probability. In Section IV, numerical simulations are provided to confirm the accuracy of our results before concluding the paper in Section V.

**C. Notations**

For simplicity, we make use of the following notations onwards.

Our work is to characterize the behaviors of the RLS-based precoder with limited PAPR, in the large dimensional regime where $m,n \to \infty$ with a fixed ratio $\delta = m/n$, and to keep the notations short we simply write $n \to \infty$. We say that an event $\xi$ holds with probability approaching $1$ (w.p.a.1) if $\lim_{n \to \infty} \mathbb{P}[\xi] = 1$. If a sequence of random variables $X_n$ converges to a constant $X$, we write $X_n \to^p X$. For any vector $x$, we use $x_i$ or $[x]_i$ to denote its $i$-th element. We also use the notation $\| \cdot \|$ to denote the Euclidean norm, and the notation $(\cdot),_+$ to denote $\max(x,0)$. We write $f(x)$ as $O(g(x))$ if there are constants $M$ such that $|f(x)| \leq M g(x)$ for all $x$ going to the limiting value in the analysis.

The empirical distribution of a vector $t \in \mathbb{R}^m$ is given by $\frac{1}{m} \sum_{i=1}^{m} \delta_{t_i}$, where $\delta_{t_i}$ is the Dirac delta mass at $t_i$. For $q \in \mathbb{N}$, a function $f : \mathbb{R}^q \to \mathbb{R}$ is said to be pseudo-Lipschitz of order $k$ if for all $x$ and $y$ in $\mathbb{R}^q$, $|f(x) - f(y)| \leq C(1 + \|x\|^{k-1} + \|y\|^{k-1}) \|x - y\|$. The Wasserstein-$k$ distance [19] between two measures $\mu$ and $\nu$ is defined as $W_k(\mu, \nu) = (\inf_{\mathbb{P}\in\mathcal{P}(\mu)} \mathbb{E}_\mathbb{P}[|X - Y|^{k}])^{\frac{1}{k}}$ where the infimum is over all random variables $(X,Y)$ such that $X \sim \mu$ and $Y \sim \nu$ marginally. A sequence of probability distributions $\nu_p$ converges in $W_k$ to $\nu$ if $W_k(\nu_p, \nu) \to 0$ as $p \to \infty$. An equivalent definition of the convergence in $W_k$ is that, for any $f$ pseudo-Lipschitz of order $k$, $\lim_{p\to\infty} \mathbb{E}[f(X_p)] = \mathbb{E}[f(X)]$ where the expectation is with respect $X_p \sim \nu_p$ and $X \sim \nu$.

**II. System Model and Problem Formulation**

Consider a conventional multiuser downlink, slow narrow band transmission between a base station equipped with $n$ transmit antennas and $m$ single antenna user terminals. The precoding scheme is a function that maps the user information symbols, collected in $s = [s_1, s_2, ..., s_m]^T = \{\pm 1\}^m$ and assumed to be drawn uniformly from the BPSK constellation, into an $n$-dimensional signal $x = [x_1, x_2, ..., x_n]^T$. Since the signal here is BPSK, we assume a real wireless channel and additive noise. Letting $h_k$ denote the channel vector between the base station and user $k$, the received signal at the $k$-th user writes as $y_k = h_k^T x + z_k$, (1)

where $z_k$ is the additive noise, assumed to follow a Gaussian distribution with mean zero and variance $\sigma^2$. Stacking the received signals into a vector $y = [y_1, ..., y_m]^T$ yields $y = Hx + z$.

The main goal of precoding is to remove the effect of the channel by minimizing the error between the channel-distorted received vector $Hx$ and the information vector $s$. To meet this requirement, the non-linear least squares precoder proposed in [12] is formulated as the solution to the following regularized least squares problem:

$$
\hat{x} = \arg \min_{x \in \mathbb{C}^n} \|Hx - \sqrt{\rho}s\|^2 + \lambda \|x\|^2,
$$

(2)

where $\rho$ is a positive power control factor, $\lambda$ is a positive regularization parameter and $X$ being a predefined set containing admissible values for the precoded signal. The formulation in (2) defines a whole class of precoded vectors for different choices of the set $X$ and parameter $\lambda$. For example, if $X = \Re$ and $\lambda > 0$, we obtain the RZF precoding given by

$$
\hat{x}_{\text{RZF}} = \sqrt{\rho} \left( H^T H + \lambda I_n \right)^{-1} H^T s,
$$

which for $\lambda = 0$ reduces to the zero-forcing (ZF) precoding (assuming $m > n$)

$$
\hat{x}_{\text{ZF}} = \sqrt{\rho} \left( H^T H \right)^{-1} H^T s.
$$

When $X = \{-\sqrt{P}, \sqrt{P}\}$, the precoder in (2) does not admit a closed-form expression, characterizing the performance of the precoder is a challenging task. In this paper, we aim to study
its performance in the large dimensional regime in which the number of antennas \( n \) and the number of users \( m \) grow large at the same pace. More formally, in our analysis, we rely on the following assumptions.

**Assumption 1.** The number of antennas \( n \) and the number of users \( m \) grow to infinity at a fixed ratio \( \delta := \frac{m}{n} \).

**Assumption 2.** The channel matrix \( \mathbf{H} \) has independent and identically distributed Gaussian entries with zero mean and a variance equal to \( \frac{1}{n} \).

We are interested in characterizing the performance of the precoder in (2) with respect to the following specifications and performance metrics.

**Per-antenna power:** We define the per-antenna transmit power as

\[
P_b := \frac{\|\mathbf{x}\|^2}{n}. \tag{3}
\]

**Per-user distortion error power:** By expressing the received signal at user \( k \) as

\[
y_k = \sqrt{\rho} s_k + \mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k + z_k, \tag{4}
\]

the distortion error observed by user \( k \) is represented by the quantity \( \mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k \). We define the per-user distortion error power as

\[
P_d := \frac{\|\mathbf{H}\mathbf{x} - \sqrt{\rho}\mathbf{s}\|^2}{m}. \tag{5}
\]

**Average per-user SINAD:** From (4), we can easily see that the SINAD at user \( k \) is given by

\[
\text{SINAD}_k = \frac{\rho}{\mathbb{E}_{s_k} \|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 + \sigma^2}.
\]

We define the average per user SINAD as:

\[
\text{SINAD} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{E} [\text{SINAD}_k]. \tag{6}
\]

**Average per-user SINAD upper bound and lower bound:** From Jensen’s inequality, we can easily check that the expected value of the SINAD at user \( k \) can be upper bounded and lower bounded as

\[
\mathbb{E}[\text{SINAD}_k] \leq \mathbb{E} \left[ \frac{\rho}{\|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 + \sigma^2} \right], \tag{7}
\]

\[
\mathbb{E}[\text{SINAD}_k] \geq \mathbb{E} \left[ \frac{\rho}{\|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 + \sigma^2} \right]. \tag{8}
\]

From (7) and (8), we can prove that the following quantities define upper and lower bounds for the average per-user SINAD:

\[
\text{SINAD}_{\text{up}} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{E} \left[ \frac{\rho}{\|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 + \sigma^2} \right], \tag{9}
\]

\[
\text{SINAD}_{\text{lb}} = \mathbb{E} \left[ \frac{1}{m} \sum_{k=1}^{m} \frac{\rho}{\|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 + \sigma^2} \right]. \tag{10}
\]

In practice, we predict the SINAD lower bound in (10) to provide a tight approximation for the SINAD. Indeed, under Assumption 2, all users experience the same channel statistics, we expect that

\[
\mathbb{E}_{s_k} \|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 
\]

is asymptotically close to \( \frac{1}{m} \sum_{k=1}^{m} \|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 \). This latter term should converge to its expectation \( \mathbb{E} \left[ \frac{1}{m} \sum_{k=1}^{m} \|\mathbf{h}_k^T \mathbf{x} - \sqrt{\rho} s_k\|^2 \right] \), and hence substituting it by its expectation leads to \( \text{SINAD}_{\text{lb}} \). Therefore, in Section IV, we compare the empirical SINAD with our approximation for \( \text{SINAD}_{\text{lb}} \) and validate its accuracy.

**Bit error rate and bit error probability:** The bit error rate (BER) is defined as

\[
\text{BER} := \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\{\text{sign}(y_i) \neq s_i\}}, \tag{11}
\]

where \( \mathbb{1}_{\{\cdot\}} \) denotes the indicator function. Another related quantity of interest is the bit error probability \( P_e \), which is defined as the expectation of the BER, i.e.,

\[
P_e := \mathbb{E} [\text{BER}] = \frac{1}{m} \sum_{i=1}^{m} \mathbb{P}[\text{sign}(y_i) \neq s_i]. \tag{12}
\]

Although both the RZF and ZF precoding admit a closed-form expression, the PAPR at each RF chain is not restricted. Taking the definition in [9], the per-antenna PAPR is defined as follows:

\[
\text{PAPR}_i := \left( \frac{1}{J} \sum_{j=1}^{J} |x_i(j)|^2 \right)^{-1} \max_{j \in [J]} |x_i(j)|^2, \tag{13}
\]

where \( x(j) \) is the \( j \)-th realization of the transmit vector, and \( J \) is the number of samples. We claim that

\[
\lim_{J \to \infty} \mathbb{E} \left[ \frac{1}{J} \sum_{j=1}^{J} |x_i(j)|^2 \right] = \lim_{n \to \infty} \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} |x_i(j)|^2 \right], \tag{14}
\]

which means the average PAPR over antennas equals that over time. The above result, verified by simulations, follows because the channel statistics over all antennas are the same, so there is no reason that one antenna experiences more power than any other one. As illustrated later in the following sections, the limiting of the per-antenna power represented by the right-hand side of (14) can be tuned to any target value by properly choosing the power control parameter. Moreover, by construction, the entries of the precoded vector are in the set \( \mathbb{X} = [-\sqrt{P}, \sqrt{P}] \), where \( P \) is carefully chosen so that the peak value of the precoded vector is restricted. As a consequence, the studied precoder achieves a PAPR that is less than \( P/\mathbb{E} \) where \( \mathbb{E} \) is the target power value. (2) will be thus referred to as the RLS-based precoder with limited PAPR and for simplicity termed as limited PAPR-RLS precoder.

**III. MAIN RESULTS**

**A. Distributional characterization of the precoded vector and the distortion error**

A major result of our study is the theoretical characterization of the empirical distributions of the elements of the precoded
vector \( \hat{x} \) and the joint empirical distribution of the distortion error vector given by
\[
\hat{e} = (H\hat{x} - \sqrt{S}s)
\]
and the transmitted symbol \( s \). As shown next, both of these distributional characterizations will be instrumental in sharply characterizing the convergences of the specifications and performance metrics introduced in the previous section.

**Theorem 1** (Distributional characterization of the precoded vector). Consider the following max-min optimization problem:
\[
\bar{\phi} = \max_{\beta \geq 0} \min_{\tau \geq 0} \frac{\tau \beta \delta}{2} + \frac{\rho \beta}{2\tau} - \frac{\beta^2}{4} + Y(\beta, \tau),
\]
(15)
where
\[
Y(\beta, \tau) = \frac{\beta}{\alpha} \left( \mathbb{E}_{H \sim N(0,1)} \left[ (H - \sqrt{P}\alpha)^2 \mathbb{1}_{\{H \geq \sqrt{P}\alpha\}} \right] - \frac{1}{2} \right),
\]
with \( \alpha = 1/\tau + 2\lambda/\beta \).

(i) The optimization problem in (15) admits a unique finite saddle-point \((\beta^*, \tau^*)\) if and only if \( \lambda > 0 \) or \( \lambda = 0 \) and \( \delta > 1 \).

(ii) When \( \lambda = 0 \) and \( \delta > 1 \), the saddle point \((\beta^*, \tau^*)\) of (15) is given by
\[
\tau^* = \arg \min_{\tau \geq 0} \left( \frac{\tau \delta}{2} + \frac{\rho}{2\tau} + \tilde{Y}(\tau) \right) + ,
\]
(16)
\[
\beta^* = \left( \tau^* \delta + \frac{\rho}{\tau^*} + 2\tilde{Y}(\tau^*) \right) + ,
\]
(17)
where
\[
\tilde{Y}(\tau) := \tau \left( \mathbb{E} \left[ (H - \sqrt{P}\tau)^2 \mathbb{1}_{\{H \geq \sqrt{P}\tau\}} \right] - \frac{1}{2} \right).
\]
Moreover, \( \bar{\phi} \) reduces to
\[
\bar{\phi} = \left( \frac{\tau \delta}{2} + \frac{\rho}{2\tau} + \tilde{Y}(\tau) \right)^2 + .
\]
(iii) Let \( \hat{x} \) be the solution of (2), and consider its associated empirical density function
\[
\hat{\mu}(\hat{x}) := \frac{1}{n} \sum_{i=1}^{n} \delta_{\hat{x}_i}.
\]
Further, let the function \( \theta : \mathbb{R} \to [-\sqrt{P}, \sqrt{P}] \),
\[
\theta(\gamma) := \begin{cases} 
-\sqrt{P} & \text{if } \gamma \leq -\sqrt{P}\alpha^* , \\
\frac{\gamma}{\sqrt{P}} & \text{if } -\sqrt{P}\alpha^* \leq \gamma \leq \sqrt{P}\alpha^* , \\
\sqrt{P} & \text{if } \gamma \geq \sqrt{P}\alpha^* ,
\end{cases}
\]
where \( \alpha^* = 1/\tau^* + 2\lambda/\beta^* \). Assume either \( \lambda > 0 \) or \( \lambda = 0 \) and \( \delta > 1 \). Then, under Assumption 1 and Assumption 2, for any pseudo-Lipschitz function \( f \) of order \( k \), it holds that
\[
\frac{1}{n} \sum_{i=1}^{n} f(\hat{x}_i) \xrightarrow{P} \mathbb{E}_H[f(\theta(H))],
\]
where \( H \sim N(0,1) \). Particularly, the empirical density function \( \hat{\mu}(\hat{x}) \) converges in Wasserstein–\( k \) distance to \( \theta(H) \).

**Proof.** See Appendix A. \( \square \)

**Theorem 2** (Distributional characterization of the distortion). Consider the setting of Theorem 1. Let \( f : \mathbb{R}^2 \to \mathbb{R} \) be a pseudo-Lipschitz function of order 2 and let \( \alpha^* = 1/\tau^* + 2\lambda/\beta^* \). Assume either \( \lambda > 0 \) or \( \lambda = 0 \) and \( \delta > 1 \). Then, under Assumption 1 and Assumption 2, the following convergence holds true:
\[
\frac{1}{m} \sum_{i=1}^{m} f(\hat{e}_i, s_i) \xrightarrow{P} \mathbb{E}_{H,S}[f \left( \frac{\beta^* \sqrt{(\tau^*)2\delta - \rho H - \sqrt{\rho S}}}{\tau^* \delta}, S \right)],
\]
(18)
where \( H \) is a standard normal scalar variable and \( S \) a discrete binary variable taking 1 and \(-1\) with equal probabilities. Equivalently, letting
\[
\hat{\mu}(e, s) := \frac{1}{m} \sum_{i=1}^{m} \delta_{(e_i, s_i)},
\]
then \( \hat{\mu}(e, s) \) converges in Wasserstein–2 distance to the distribution of \( \frac{\beta^*}{2} \sqrt{(\tau^*)2\delta - \rho H - \sqrt{\rho S}} S \).

**Proof.** The proof of Theorem 2 shares similarities with that of Theorem 1, so we do not provide such a proof for Theorem 2 here due to space limitations. We refer the interested reader to Section VI-D of the full version of this paper [20]. \( \square \)

**B. Characterizations of specifications and performance metrics**

As an application of Theorem 1 and Theorem 2, we derive closed-form approximations for the specifications and performance metrics defined in Section II:

**Corollary 1** (Convergence of the average SINAD upper and lower bounds). Consider Assumption 1 and 2, then \( \text{SINAD}_{up} \) converges to:
\[
\text{SINAD}_{up} \xrightarrow{P} \text{SINAD}_{up} := \mathbb{E}_{H,S} \left[ \frac{\rho}{4(\beta^*)2 \left( \frac{(\sqrt{\tau^*)2\delta - \rho H - \sqrt{\rho S}}}{\tau^* \delta} \right)^2 + \sigma^2} \right],
\]
(19)
and \( \text{SINAD}_{lb} \) converges to:
\[
\text{SINAD}_{lb} \xrightarrow{P} \text{SINAD}_{lb} := \frac{\rho}{(\beta^*2)^2 + \sigma^2}.
\]
(20)

**Proof.** Function \( x \mapsto \frac{\rho}{x^2 + \sigma^2} \) is a Lipschitz function. Applying Theorem 2 yields:
\[
\frac{1}{m} \sum_{k=1}^{m} \frac{\rho}{|e_k|^2 + \sigma^2} \xrightarrow{P} \mathbb{E}_{H,S} \left[ \frac{\rho}{4(\beta^*)2 \left( \frac{(\sqrt{\tau^*)2\delta - \rho H - \sqrt{\rho S}}}{\tau^* \delta} \right)^2 + \sigma^2} \right].
\]
(21)
Finally, since \( x \mapsto \frac{\rho}{x^2 + \sigma^2} \) is bounded by \( \frac{\rho}{\sigma^2} \), the convergence in (19) follows from the dominated convergence theorem. To
prove (20), we use the fact that $x \mapsto x^2$ is a pseudo-Lipschitz function of order 2. Hence, we may again use Theorem 2 to obtain

$$\frac{1}{m} \sum_{k=1}^{m} |e_k|^2 \Rightarrow (\beta^*)^2 \frac{\rho}{4\delta}.$$  

To prove the convergence in (20), it suffices to check that $\frac{1}{m} \sum_{k=1}^{m} |e_k|^2$ is bounded. Indeed, if this is true then one can in a similar way as before use the dominated convergence theorem to prove the convergence of the expectation of $\frac{1}{m} \sum_{k=1}^{m} |e_k|^2$ to its probability limit. Using the fact that $\hat{x}$ minimizes the cost in (2), the following inequality holds:

$$\frac{1}{m} \| \mathbf{H} \hat{x} - \sqrt{\rho} \mathbf{s} \|^2 + \frac{\lambda}{m} \| \hat{x} \|^2 \leq \frac{1}{m} \| \mathbf{H} \hat{x} - \sqrt{\rho} \mathbf{s} \|^2$$

and hence,

$$\frac{1}{m} \| \mathbf{H} \hat{x} - \sqrt{\rho} \mathbf{s} \|^2 \leq \rho.$$  

Recalling that $\frac{1}{m} \sum_{k=1}^{m} |e_k|^2 = \frac{1}{m} \| \mathbf{H} \hat{x} - \sqrt{\rho} \mathbf{s} \|^2$ we establish that $\frac{1}{m} \sum_{k=1}^{m} |e_k|^2$ is bounded.  

**Corollary 2** (Convergence of the per-antenna power and the per-user distortion error power). Under the setting of Theorem 1, the per-antenna and the per-user distortion error power satisfy the following convergences:

$$P_b \Rightarrow P_b^* := \delta (\tau^*)^2 - \rho, \quad (22)$$

and

$$P_d \Rightarrow P_d^* := \frac{(\beta^*)^2}{4\delta}. \quad (23)$$

**Proof.** Note that $P_d = \frac{1}{m} \sum_{k=1}^{m} |e_k|^2$. The convergence of $P_d$ to its probability limit has been established in the proof of Corollary 1. The convergence of $P_b$ to the limit in (22) follows directly by applying Theorem 1 along with the first-order optimality condition for the variable $\tau$.  

Corollary 2 allows us to provide an interpretation of the parameters $\tau^*$ and $\beta^*$. From the convergences stated in this Corollary, it appears that $\tau^*$ is related to how much power is devoted to the precoded vector $\hat{x}$, while $\beta^*$ allows for quantifying the amount of distortion experienced by the PAPR precoder. The control factor $\rho$ can always be adjusted to fix the power $P_b^*$ to a given feasible value. However, this would lead to varying the coefficient $\beta^*$ which determines the distortion level. More details on the role of the control factor $\rho$ on the performance will be given in this section and in section IV.

**Corollary 3** (Convergence of the bit error probability). Under the setting of Theorem 1, the bit error probability defined in (12) converges to

$$P_e \Rightarrow P_e^* := \Phi \left( \frac{\sqrt{\rho} - \frac{\beta^*}{2\pi^*} \sqrt{\frac{\rho}{2\pi^*}}}{\sqrt{\frac{(\beta^*)^2 (\tau^*)^2 (\pi^*)^2}{4} - \frac{(\beta^*)^2 (\tau^*)^2 (\pi^*)^2}{\pi^*} + \sigma^2}} \right).$$

**Proof.** The symbol $s_k$ is decoded erroneously if

$$\mathbf{h}_k^T \hat{x} + \sqrt{\rho} + z_k \leq -\sqrt{\rho}$$

when $s_k = 1$ and

$$\mathbf{h}_k^T \hat{x} + \sqrt{\rho} + z_k \geq \sqrt{\rho}$$

when $s_k = -1$. So

$$P_e = \frac{1}{2} \mathbb{P} \left[ \mathbf{h}_k^T \hat{x} - \sqrt{\rho} s_k + z_k \leq -\sqrt{\rho} \mid s_k = 1 \right] + \frac{1}{2} \mathbb{P} \left[ \mathbf{h}_k^T \hat{x} - \sqrt{\rho} s_k + z_k \geq \sqrt{\rho} \mid s_k = -1 \right]. \quad (24)$$

Using Theorem 2 along with the Portemanteau Lemma [21], we prove that

$$P_e \Rightarrow \frac{1}{2} \mathbb{P} \left[ \frac{\beta^*}{2} \sqrt{\frac{(\tau^*)^2 (\pi^*)^2 - \rho H - \sqrt{\rho} S}{\pi^*}} + \sigma Z \leq -\sqrt{\rho} S = 1 \right] + \frac{1}{2} \mathbb{P} \left[ \frac{\beta^*}{2} \sqrt{\frac{(\tau^*)^2 (\pi^*)^2 - \rho H - \sqrt{\rho} S}{\pi^*}} + \sigma Z \geq \sqrt{\rho} S = -1 \right] \quad (25)$$

$$= \Phi \left( \frac{\sqrt{\rho} - \frac{\beta^*}{2\pi^*} \sqrt{\frac{\rho}{2\pi^*}}}{\sqrt{\frac{(\beta^*)^2 (\tau^*)^2 (\pi^*)^2}{4} - \frac{(\beta^*)^2 (\tau^*)^2 (\pi^*)^2}{\pi^*} + \sigma^2}} \right). \quad (26)$$

where in (25) $Z$ follows a standard normal distribution with mean zero and variance 1.  

It is important to note that although a BPSK modulation is assumed, (26) is different from the asymptotic bit error probability $P_e = Q(\sqrt{2SNR})$. The reason lies in the fact that the latter relation holds in the case of additive Gaussian noise that is independent of the transmitted symbols. In our case, we have not only noise but also the distortion $\hat{x}$ which is correlated with the transmitted symbols, as evidenced by Theorem 2.

**C. Special cases: RZF and ZF precoding** ($P \to \infty$)

The analysis of the RZF and ZF precoding in multi-user downlink systems has been the focus of several studies in the literature. Among these studies, we cite the work in [22] which considered this problem with sophisticated channel models involving different correlations across users. However, to the best of our knowledge, none of the existing works studied the bit error probability approximation (all the focus being on the asymptotic characterization of the SINAD). In the sequel, we show that by taking $P \to \infty$ in the asymptotic expressions of Theorem 1, we can simplify the expression (20) to reach the same results for the asymptotic SINAD performance obtained in the literature. Additionally, we obtain new asymptotic approximations for the bit error probability.  

For the sake of scientific rigor, since our proofs in Theorem 1 and Theorem 2 relied on the assumption of finite values of $P$, we do not claim the convergence in probability of the specifications and performance metrics to the limits of their asymptotic equivalents when $P \to \infty$, although we believe this to be the case. A rigorous proof of the convergence would require us to re-consider the case where $P \to \infty$ separately. However, we do not provide such a proof since the analysis of the RZF or the ZF can be conducted using tools from random matrix theory [23] and is thus less worthy of consideration.
Theorem 3 \((P \to \infty \text{ and } \lambda > 0)\). For a given value of \(P\), denote by \(\tau^*(P)\) and \(\beta^*(P)\) the solutions to the max-min problem in (15). Assume \(\lambda > 0\), then as \(P \to \infty\), the following convergences hold true:
\[
\lim_{P \to \infty} \tau^*(P) = \frac{\sqrt{\rho}}{\sqrt{\delta - \frac{1}{\lambda+1}}}.
\]
(27)
\[
\lim_{P \to \infty} \beta^*(P) = \frac{2\sqrt{\rho}}{s^*\sqrt{\delta - \frac{1}{\lambda+1}}}.
\]
(28)
where \(s^*\) is given by:
\[
s^* = \frac{\sqrt{(\delta - \lambda - 1)^2 + 4\delta \lambda - \delta + \lambda + 1}}{2\delta \lambda}.
\]
(29)
Particularly, in this regime, the asymptotic values of the per-antenna power, the distortion power, the SINAD lower bound \(\text{SINAD}_{lb}\), and the bit error probability converge to
\[
\lim_{P \to \infty} P^*_b = \frac{\rho}{\delta (1 + \lambda s^*)^2 - 1}.
\]
(30)
\[
\lim_{P \to \infty} P^*_d = \frac{(s^*)^2 \delta (\delta - \frac{1}{1 + \lambda s^*})}{(1 + \lambda s^*)^2}.
\]
(31)
\[
\lim_{P \to \infty} \text{SINAD}^*_{lb} = \frac{(s^*)^2 \delta (\delta - \frac{1}{1 + \lambda s^*})}{1 + \frac{2\rho}{\rho} (s^*)^2 \delta (\delta - \frac{1}{1 + \lambda s^*})}.
\]
(32)
\[
\lim_{P \to \infty} P^*_e = Q\left(\frac{\sqrt{\rho (\delta^* - 1)}}{\sqrt{\delta^2 (s^*)^2 + \frac{2\rho}{\rho} (\delta^* - 1)}}\right).
\]
(33)
Proof. See Appendix C in the supplementary material. \(\square\)

Theorem 4 \((P \to \infty, \lambda = 0 \text{ and } \delta > 1)\). For a given value of \(P\), denote by \(\tau^*(P)\) and \(\beta^*(P)\) the solutions to the max-min problem in (15). Assume \(\lambda = 0\) and \(\delta > 1\). Then as \(P \to \infty\),
\[
\lim_{P \to \infty} \tau^*(P) = \sqrt{\frac{\rho}{\delta - 1}},
\]
(34)
\[
\lim_{P \to \infty} \beta^*(P) = 2\sqrt{\rho (\delta - 1)}.
\]
(35)
Particularly, in this regime, the asymptotic values of the per-antenna power, the distortion power, the SINAD lower bound \(\text{SINAD}^*_{lb}\), and the bit error probability converge to
\[
\lim_{P \to \infty} P^*_b = \frac{\rho}{\delta - 1},
\]
(36)
\[
\lim_{P \to \infty} P^*_d = (\rho - 1)\frac{1}{\delta},
\]
(37)
\[
\lim_{P \to \infty} \text{SINAD}^*_{lb} = \rho (\rho - 1)\frac{1}{\delta} + \sigma^2,
\]
(38)
\[
\lim_{P \to \infty} P^*_e = Q\left(\frac{\sqrt{\rho}}{\sqrt{\rho (\delta - 1) + \sigma^2}}\right).
\]
(39)
Proof. By setting \(\lambda = 0\) in the proof of Theorem 3, we directly obtain (34) and (35), from which the approximations in (36)-(39) follow easily. \(\square\)

D. Limiting cases

The expressions derived so far are useful to characterize the performance of the limited PAPR precoder in terms of the design parameters, that is the ratio of \(m\) to \(n\), and the power control parameter \(\rho\). To gain more insight into the impact of these parameters on the performance of the limited PAPR precoder, next we study the following limiting cases.

Theorem 5 \((\text{The number of users much smaller than the number of antennas } (m \ll n))\). For a given value of \(\delta\), denote by \(\tau^*(\delta)\) and \(\beta^*(\delta)\) the solutions to the max-min problem in (15). Assume \(\lambda > 0\). Then as \(\delta \to 0\), the following convergences hold true:
\[
\lim_{\delta \to 0} \tau^*(\delta) = 1,
\]
(40)
\[
\lim_{\delta \to 0} \beta^*(\delta) = 1.
\]
(41)
Particularly, in this regime, the asymptotic values of the per-antenna power, the distortion power, the SINAD lower bound \(\text{SINAD}^*_{lb}\), and the bit error probability converge to
\[
\lim_{\delta \to 0} P^*_b = 0,
\]
(42)
\[
\lim_{\delta \to 0} P^*_d = \frac{\rho}{(1 + \frac{1}{\delta})^2},
\]
(43)
\[
\lim_{\delta \to 0} \text{SINAD}^*_{lb} = \frac{\rho}{(1 + \frac{1}{\delta})^2} + \sigma^2,
\]
(44)
\[
\lim_{\delta \to 0} P^*_e = Q(\sqrt{\rho}).
\]
(45)
Proof. See Section X-A of the full version of this paper [20]. \(\square\)

Theorem 5 allows us to understand the behavior of the limited PAPR precoder when the number of available antennas largely exceeds the number of users. As an important remark, we note that, interestingly, all specifications and performance metrics do not asymptotically depend on \(P\). In other words, considering the regimes where \(\delta \to 0\), regardless of the maximum power at each antenna, the performance is almost the same. However, the results depend on \(\lambda\) when \(\delta \to 0\). This behavior can be attributed to the fact that the limited PAPR precoder becomes close to the RZF precoder when \(\delta \to 0\).

Below, we provide arguments supporting these claims.

In this case, the limited PAPR precoder solving (2) becomes close to the RZF precoder if the latter satisfies the per-antenna constraint. It turns out that when \(\delta \to 0\), the Frobenius norm of \(\textbf{H}^T\) scales as \(\sqrt{m}\) and so does the per antenna power of the RZF \(^1\), which leads to the RZF becoming asymptotically feasible with respect to the optimization problem (2). However, in this case, it is advisable to select a small regularization parameter. The reason lies in that a sufficient number of degrees of freedom are available to find \(\textbf{x}\) such that \(|\textbf{Hx} - \sqrt{7}\textbf{s}|^2\) is as small as desired. Using a large regularization parameter will thus increase the bias, thereby deteriorating the performance.

\(^1\)Here we used the fact that \(|\text{SINAD}^*_{lb}| \leq \frac{1}{2}\|\textbf{H}^T\textbf{s}\|\) and that \(|\|\textbf{H}^T\textbf{s}\|\) can be approximated by \(\sqrt{\text{tr}(\textbf{H}^T\textbf{H})}\) with high probability when \(m \to \infty\).

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.Authorized licensed use limited to: KAUST. Downloaded on November 02,2022 at 06:42:33 UTC from IEEE Xplore. Restrictions apply.
Fig. 1: The theoretical per-antenna power, distortion power, SINAD lower bound and bit error probability versus $\delta$, for $\sigma = 0.05$ and different parameter combinations.

**Numerical illustration.** Figure 1 plots the theoretical values for all the studied specifications and performance metrics versus $\delta$. As expected from Theorem 5, the per-antenna power goes to zero as $\delta$ tends to zero. Moreover, the performance in terms of bit error probability and SINAD becomes the best when $\delta$ is close to zero due to the excess in the number of spatial degrees of freedom.

**Theorem 6** (Power control parameter $\rho \to 0$). For a given value of $\rho$, denote by $\tau^*(\rho)$ and $\beta^*(\rho)$ the solutions to the max-min problem in (15). Then the following statements hold true:

1) Assume $\lambda > 0$. Then, as $\rho \to 0$, the following convergences hold

$$\lim_{\rho \to 0} \frac{\tau^*(\rho)}{\sqrt{\delta}} \sqrt{\frac{1}{1+\lambda s^*}} = 1,$$

where $s^*$ is the same as in (29). Particularly, in this regime, the asymptotic values for the per-antenna power, the distortion power, the SINAD lower bound $\text{SINAD}_{lb}$, and the bit error probability become

$$\lim_{\rho \to 0} \frac{\tau^*(\rho)}{\sqrt{\delta}} \sqrt{\frac{1}{1+\lambda s^*}} = 1,$$

$$\lim_{\rho \to 0} \frac{\beta^*(\rho)}{s^* \sqrt{\delta}} \sqrt{\frac{1}{1+\lambda s^*}} = 1,$$

$$\lim_{\rho \to 0} \frac{\text{SINAD}_{lb}}{\text{SINAD}_{lb}} = 1,$$

$$\lim_{\rho \to 0} \frac{P_b^*}{P_b} = 1.$$
and the bit error probability converge to:

\[
\lim_{\rho \to 0} \frac{P_b^*}{\rho} = 1, \\
\lim_{\rho \to 0} \frac{P_d^*}{\delta (\delta - 1) \sigma^2} = 1, \\
\lim_{\rho \to 0} \frac{\text{SINAD}_{lb}^*}{\sigma^2} = 1, \\
\lim_{\rho \to 0} \frac{1}{2} - \frac{1}{\sqrt{2\pi} \sqrt{\delta (\delta - 1) + \sigma^2 \delta^2}} = 1.
\]

2) Assume \( \lambda = 0 \) and \( \delta > 1 \). Then, as \( \rho \to 0 \), the following convergences hold:

\[
\lim_{\rho \to 0} \frac{\tau^*(\rho)}{\sqrt{\delta}} = 1, \\
\lim_{\rho \to 0} \frac{\beta^*(\rho)}{2\sqrt{\delta (\delta - 1) \sigma^2}} = 1.
\]

Particularly, in this regime, the asymptotic values for the per-antenna power, the distortion power, the SINAD lower bound \( \text{SINAD}_{lb}^* \), and the bit error probability converge to:

\[
\lim_{\rho \to 0} \frac{P_b^*}{\delta - 1} = 1, \\
\lim_{\rho \to 0} \frac{P_d^*}{\rho (\delta - 1)} = 1, \\
\lim_{\rho \to 0} \frac{\text{SINAD}_{lb}^*}{\sigma^2} = 1, \\
\lim_{\rho \to 0} \frac{1}{2} - \frac{1}{\sqrt{2\pi} \sqrt{\delta (\delta - 1) + \sigma^2 \delta^2}} = 1.
\]

**Proof.** See Section X-C of the full version of this paper [20].

Theorem 6 and Theorem 7 allow us to shed light on the behavior of the limited PAPR precoder when the control parameter \( \rho \) goes to either zero or infinity. As an interesting remark, we note that in the case where \( \rho \to 0 \), the performance becomes independent of \( P \) but dependent on the regularization parameter \( \lambda \). In this case, we claim that the limited PAPR precoder becomes close to the RZF. This is because as \( \rho \to 0 \), the entries of the RZF precoder tend to zero and thus the RZF becomes feasible with respect to the optimization problem in (2). On the other hand, when \( \rho \to \infty \), the performance depends on \( P \) but not on the regularization parameter. To explain such behavior, we argue that, in this case, the limited PAPR precoder becomes close to the non-linear least squares (LS) precoder given by:

\[
\hat{x}_{\text{LS}} = \arg \min_{-\sqrt{P} \leq x \leq \sqrt{P}} \|Hx - \sqrt{\rho} s\|^2.
\]

To see this, it suffices to note that for all \( x \), the term \( \|Hx - \sqrt{\rho} s\|^2 \) becomes dominant in the minimization of (2) since for all feasible \( x \),

\[
\|Hx - \sqrt{\rho} s\|^2 \geq \|Hx_{\text{ZF}} - \sqrt{\rho} s\|^2 = \rho \| (I - H(H^T H)^{-1} H^T) s \|^2
\]

while the second term \( \lambda \|x\|^2 \) remains bounded by \( Pn \) as \( \rho \) grows large.

By combining the observations in both regimes (\( \rho \to 0 \) and \( \rho \to \infty \)), we get a more precise idea of the role of the control parameter \( \rho \) on the per-antenna power \( P_b^* \). Setting \( \rho \) to small values makes the per-antenna power close to zero while using large values for \( \rho \) leads the precoder to use the maximum allowed power at each antenna. Such behavior is illustrated in Figure 2, which plots \( P_b^* \) against \( \rho \) for several values of \( P \). As can be seen, by varying \( \rho \), the per-antenna power varies accordingly, becoming small for small \( \rho \) values and close to the maximum allowed power for very large \( \rho \) values. Note that, unlike RZF and ZF, the value of \( \rho \) that achieves a fixed asymptotic per-antenna power \( P_b^* \) cannot be determined in an explicit form. In this respect, when it comes to comparing precoders, it is necessary to require the same \( P_b^* \) value. This can be done for each precoder by using the value of \( \rho \) that achieves the target \( P_b^* \). A plot like the one in Figure 2 can be used to determine numerically the corresponding values of the power control parameter.

\[
\text{SINAD}_{lb}^*, \text{ and the bit error probability can be approximated as:}
\]

\[
P_b^* = P + O \left( \frac{1}{\sqrt{\rho}} \right),
\]

\[
P_d^* = \rho - 2\left[ \frac{2P}{P} \right] + O(1),
\]

\[
\text{SINAD}_{lb}^* = 1 + 2\left[ \frac{2P}{P} \right] + O(1),
\]

\[
P_e^* = Q \left( \left[ \frac{2P}{P} \right] + O \left( \frac{1}{\sqrt{\rho}} \right) \right).
\]
By expanding $\|Hx - \sqrt{\rho}s\|^2$ in (64) and neglecting the quantities independent of $\rho$ or $x$, we may claim that for large $\rho$ values, the precoder in (64) would be close to the one-bit precoding $\hat{x} := \sqrt{P}\text{sign}(H^Ts)$. Such a finding, although making sense, calls into question the main interest behind solving the optimization problem in (2) to obtain the limited PAPR precoder. If for $\rho \to \infty$, its behavior would be equivalent to the precoder $\hat{x}$ which uses the maximum allowed power, one can rightly think that it should be less complex and more efficient to use $\hat{x}$ rather than solving the involved problem in (2). Such a conclusion would be correct if more power necessarily implies better performance. As evidenced later in the simulation section, it is possible for a precoder using a lower per-antenna power to perform better than the one-bit precoding scheme given by $\hat{x}$ (See Figure 7 and Figure 8).

**Numerical illustration.** Figure 3 plots the theoretical values for all studied specifications and performance metrics against $\rho$. In agreement with the results of Figure 2, the per-antenna power is an increasing function of $\rho$, approaching $P$ when $\rho$ tends to infinity. However, when $\rho$ becomes very large, the distortion power increases, resulting in the saturation of the SINAD and the bit error probability. Interestingly, there is an optimal finite $\rho$, and hence an optimal $P^\star_b$ for which the performances in terms of SINAD and bit error probability are maximized.

**IV. Numerical simulations**

In this section, we numerically investigate the performance of the limited PAPR precoders under different settings. We study the following specifications and performance metrics: the average per-user SINAD defined in (6), the average per-antenna power, the average per-user distortion power, and the bit error probability. We compare the results with the theoretical predictions derived in Section III. In all the figures below, solid lines represent the theoretical predictions, while markers show the simulated results averaged over 50 realizations of random quantities $H$, $s$ and $z$.

**A. Impact of the regularization parameter $\lambda$**

Figure 4 illustrates the behavior of various specifications and performance metrics with a varying regularization parameter $P = 8, 10$ and $12$. We note that the per-antenna power decreases with $\lambda$, since a large $\lambda$ value would penalize the term $\|x\|^2$ in (2) more. However, as it can be seen from the plot of the bit error probability, setting $\lambda$ to smaller values does not always translate into better performance. Indeed, a non-zero optimal value of $\lambda$ exists, which is small for a large $P$, but becomes larger as $P$ decreases. This shows that regularization is more important when $P$ is small to compensate for the bias caused by restricting the per-antenna power of the precoded vector. In a second experiment, we investigate whether this behavior still holds when all precoders have the same average.
Fig. 4: Impact of the regularization parameter $\lambda$ for different $P$ values. $n = 512$, $\delta = 0.84$, $\rho = 2$ and $\sigma = 0.05$. (Markers show the simulated results averaged over 50 realizations of random quantities $H$, $s$ and $z$. We do not show simulated results for $P_e$ because they require too many runs to simulate such a low $P_e$.)

B. Impact of the number of users to the number of antennas ratio ($\delta$)

In Figure 6, we investigate the impact of the number of users to the number of antennas ratio $\delta$ on the performance of the limited PAPR precoder. As in Figure 5, for each plot, we leverage our asymptotic analysis to set the power control parameter $\rho$ at the value ensuring the target asymptotic per-antenna power $P^*_b$. As expected, we note that the power distortion increases with $\delta$. This is because a higher $\delta$ translates into serving more users and thus causes higher distortion error levels. However, it is curious to note that the distortion error reaches very high levels as $P^*_b$ becomes of an order of magnitude of $P$. To explain this, we refer to the findings of Theorem 7 and Figure 2, which suggest that a higher value of $\rho$ is required to reach higher values of $P^*_b$. But, when $\rho$ is large, the distortion error automatically increases as it becomes difficult to approximate $\sqrt{\rho}s$ by $Hx$ when $x$ is constrained to a compact set. An important consequence of this behavior is that the SINAD performance and the bit error probability do not always improve by increasing the average per-antenna power $P^*_b$. As shown in Figure 6, the performance is worse for $P^*_b = 10$ than for $P^*_b = 3$. This is because, for $P^*_b = 10$, the higher transmit power could not compensate for the higher distortion caused by using a higher value for $\rho$.

C. Comparison between precoders with optimal regularization

Figure 7 demonstrates the performance variation with $P_b$ for $P = 25, 30$, and $35$, when the regularization parameter $\lambda$ is set to the value optimizing the SINAD. As in Figure 5 and Figure 6, $\rho$ is tuned to achieve the target $P^*_b$. As an important remark, we note that there exists a $P^*_b$ for which the performances in terms of bit error probability and SINAD are optimal. This value is below $P$. Indeed, as $P^*_b$ approaches $P$, the distortion power significantly increases, resulting in a large performance deterioration. In a final experiment, we compare in Figure 8 the bit error probability performances of the limited PAPR precoder using optimal regularization with the one-bit precoding $\hat{x}_{\rho} = \sqrt{P^*_b}\text{sign}(H^Ts)$. Complexitywise, the one-bit precoding possesses an explicit form and thus is more computationally efficient than the limited PAPR precoder which is based on solving a convex optimization problem. However, when it comes to bit error probability performance,
we can easily see that the limited PAPR precoder is more efficient for all values of $P^*_b$. As expected from Theorem 7, the performance gap is small when $P^*_b$ approaches $P$ but becomes much more pronounced when the limited PAPR precoder uses the optimal value of $P^*_b$.

V. CONCLUSION

In this paper, we studied the asymptotic behavior of the limited PAPR precoder for multi-user communication systems in the regime in which the number of antennas and that of users grow large at the same pace. Contrary to the previous studies in [9] and [11], we rely on the CGMT framework and present approximations for other important performance metrics including the bit error probability and the average per user SINAD. To get more insights, we particularized our results to specific regimes in which the number of antennas is much larger than that of users, or the power control parameter takes very small or very high values. As a major outcome, our analysis demonstrates the existence of an optimal transmit power that maximizes the SINAD, and the bit error probability.
Appendix A: Proof of Theorem 1

A. The CGMT framework

The main technical ingredient is the CGMT. Before delving into the technical details of the proof, we provide a brief overview of the CGMT tool.

Fig. 7: The distortion power, SINAD and bit error probability versus $P_b$, for different $P$ values. $n = 256$, $\delta = 1$ and $\sigma = 0.05$. The control parameter $\rho$ is tuned to achieve the target $P_b^*$ and the regularization parameter $\lambda$ is set to the value that maximizes the SINAD.

The CGMT is a mathematical framework that allows us to study the asymptotic behavior of high-dimensional optimization problems that can be written in the form of

$$\Phi(G) := \min_{w \in S_w} \max_{u \in S_u} u^T G w + \psi(w, u),$$

(65)

where $G \in \mathbb{R}^{m \times n}$ is a standard Gaussian matrix, $\psi$ is a real-valued function possibly random but independent of $G$, and $S_w$ and $S_u$ are two compact sets. The problem defined in (65) is known as the primary optimization problem (PO). The CGMT infers the behavior of the PO by considering the following associated auxiliary optimization problem (AO):

$$\phi(g, h) := \min_{w \in S_w} \max_{u \in S_u} \|w\| g^T u - \|u\| h^T w + \psi(w, u),$$

(66)

where $g \in \mathbb{R}^m$ and $h \in \mathbb{R}^n$ two standard Gaussian vectors. More formally the CGMT is stated as follows:

**Theorem 8 (CGMT).** Consider the optimization problems in (65) and (66) The following statements hold true:

- For all $t \in \mathbb{R},$

$$P[\Phi(G) \leq t] \leq 2P[\phi(g, h) \leq t].$$

(67)

- If additionally $S_w$ and $S_u$ are convex and $\psi$ is convex-concave, then for all $t \in \mathbb{R},$

$$P[\Phi(G) \geq t] \leq 2P[\phi(g, h) \geq t].$$

(68)

Particularly, for any $\nu \geq 0$ it holds that:

$$P[|\Phi(G) - \nu| \geq t] \leq 2P[|\phi(g, h) - \nu| \geq t].$$

According to the first statement in Theorem 8,

$$P[\Phi(G) \leq t] \leq 2P[\phi(g, h) \leq t].$$

Equivalently stated, this implies that a high-probability lower bound of the AO cost is also a high probability lower bound of the PO. Such a result holds even when the sets or the function $\psi$ are not convex.
However, the main interest in the CGMT lies in the second statement of Theorem 8, which affirms that under convexity conditions of the PO, the AO can be used to infer properties on the PO’s asymptotic cost. More precisely, if for some \( \nu \), the AO cost concentrates around \( \nu \), so does cost of the PO. Moreover, as shall be shown next, under appropriate strong-convexity conditions with respect to the solutions of the AO, the CGMT shows that concentration of Lipschitz functions of the solution of the AO implies concentration of that of the PO.

In the sequel, we make use of the CGMT framework to analyze the performance of the PAPR precoding scheme. As a first step, we express the PAPR precoding problem as a PO problem.

**B. Relating the PAPR precoding problem to POs**

**Formulation of the POs.** For \( X = [−\sqrt{T}, \sqrt{T}] \), the solution of the regularized least squares problem is given by

\[
\hat{x} = \arg\min_{\|z\|^2} \frac{1}{n} \|Hx - \sqrt{n}s\|^2 + \frac{\lambda}{n} \|x\|^2,
\]

(69)

where compared to (2), we normalized the optimization cost by \( \frac{1}{n} \). Using the following identity:

\[
\|z\|^2 = \max_{u \in \mathbb{R}^m} u^Tz - \frac{\|u\|^2}{4},
\]

which holds for any vector \( z \in \mathbb{R}^m \), we can write the optimization problem in (69) as

\[
\min_{\|z\|^2} \max_{u \in \mathbb{R}^m} \frac{\sqrt{n}u^THx}{n} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n}.
\]

(70)

The above problem is in the form of the PO, except that the constraint set over \( u \) is not bounded. From first-order optimality conditions, we can easily check that the optimal \( u \) is given by

\[
u^* = 2 \left( \frac{1}{\sqrt{n}}Hx - \frac{\sqrt{n}s}{\sqrt{n}} \right).
\]

Hence,

\[
\|\nu^*\| \leq 2\sqrt{T} \|H\| + \frac{\sqrt{n}m}{\sqrt{n}}.
\]

Also, using standard inequalities of the spectral norm of Gaussian matrices, we can prove that \( \|H\| \leq B \) with probability approaching 1 for some positive constant \( B \). All this shows that \( \|\nu^*\| \) is bounded with probability approaching 1. Thus the analysis would not thus change if we instead consider the following problem:

\[
\min_{\|z\|^2} \max_{u \in \mathbb{R}^m} \frac{\sqrt{n}u^THx}{n} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n}.
\]

(71)

where \( S_u = \{ u \in \mathbb{R}^m, \|u\| \leq B \} \) for some \( B > 0 \) is a high-probability upper bound on \( \|u^*\| \). Our interest is to characterize the asymptotic behavior of the solutions in \( x \) and \( u \) to (71), which perfectly agrees with the conditions required by the CGMT. For that, we introduce the following cost functions:

\[
C_{\lambda,\rho}(x) = \max_{u \in S_u} \frac{\sqrt{n}u^THx}{n} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n},
\]

(72)

\[
V_{\lambda,\rho}(u) = \min_{x^* \in S_u} \frac{\sqrt{n}u^THx}{n} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n},
\]

(73)

and consider the following primary problems:

\[
\Phi_{\lambda,\rho}(H) := \min_{x^* \in S_u} C_{\lambda,\rho}(x),
\]

(74)

\[
\hat{\Phi}_{\lambda,\rho}(H) := \max_{u \in S_u} V_{\lambda,\rho}(u).
\]

(75)

Since the objective function in (71) is convex in \( x \) and concave in \( u \), then

\[
\Phi_{\lambda,\rho}(H) = \hat{\Phi}_{\lambda,\rho}(H),
\]

(76)

and the solutions \( \hat{x}_{PO} \) and \( \hat{u}_{PO} \) to (71) are given by \(^2\)

\[
\hat{x}_{PO} := \arg\min_{x^* \in S_u} C_{\lambda,\rho}(x),
\]

(77)

\[
\hat{u}_{PO} := \arg\max_{u \in S_u} V_{\lambda,\rho}(u).
\]

(78)

**Formulation of the AOs.** With the PO problems in (74) and (75), we associate the following AO problems:

\[
\phi_{\lambda,\rho}(g, h) := \min_{x^* \in S_u} L_{\lambda,\rho}(x),
\]

(79)

\[
\hat{\phi}_{\lambda,\rho}(g, h) := \max_{u \in S_u} F_{\lambda,\rho}(u),
\]

(80)

where \( L_{\lambda,\rho}(x) \) and \( F_{\lambda,\rho}(u) \) are given by

\[
L_{\lambda,\rho}(x) := \max_{u \in S_u} \frac{1}{n} \|x\|g^Tu - \frac{1}{n} \|u\|h^Tx - \frac{\sqrt{n}u^Ts}{\sqrt{n}} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n},
\]

(81)

\[
F_{\lambda,\rho}(u) := \min_{x^* \in S_u} \frac{1}{n} \|x\|g^Tu - \frac{1}{n} \|u\|h^Tx - \frac{\sqrt{n}u^Ts}{\sqrt{n}} - \frac{\|u\|^2}{4} + \frac{\lambda \|x\|^2}{n}.
\]

(82)

Similarly, we define the solutions \( \hat{x}^{AO} \) and \( \hat{u}^{AO} \) as

\[
\hat{x}^{AO} := \arg\min_{x^* \in S_u} L_{\lambda,\rho}(x),
\]

(83)

\[
\hat{u}^{AO} := \arg\max_{u \in S_u} F_{\lambda,\rho}(u).
\]

(84)

The objective of the CGMT is to prove that the properties of the solutions of the PO defined in (77) and (78) can be transferred to the solutions of the AO defined in (83) and (84). This can be performed by using the following inequalities which directly follow as a direct application of Theorem 8.

---

\(^2\)Note that the objective in (71) is convex in \( x \) and concave in \( u \). Hence, the solutions \( x \) and \( u \) are unique.
As a first step, we show that Statement 1 holds if for an appropriate choice of \( \varepsilon > 0 \), the function \( \mathcal{L} \) is \( \frac{1}{\varepsilon} \)-strongly convex in a neighborhood of \( \Phi^{AO} \). Moreover, according to the second statement, for some \( \varepsilon \) sufficiently small, we have:

\[
\hat{\mathcal{L}}_{\lambda,\rho}(\Phi^{AO}) \leq \min_{x} \mathcal{L}_{\lambda,\rho}(x) + \varepsilon.
\]

Hence, from Lemma B1 in [24],

\[
\forall x, \frac{a}{\beta n} \|x - \Phi^{AO}\|^2 \geq \varepsilon \implies \hat{\mathcal{L}}_{\lambda,\rho}(x) \geq \min_{x^* \leq P} \hat{\mathcal{L}}_{\lambda,\rho}(x) + \varepsilon.
\]

Consequently, if for every \( x \in S \), (92) holds, then

\[
\forall x \in S, \quad \hat{\mathcal{L}}_{\lambda,\rho}(x) \geq \min_{x^* \leq P} \hat{\mathcal{L}}_{\lambda,\rho}(x) + \varepsilon \geq \phi_{\lambda,\rho}(g, h) + \varepsilon.
\]

Hence,

\[
\min_{x \in S} \mathcal{L}_{\lambda,\rho}(x) \geq \phi_{\lambda,\rho}(g, h) + \varepsilon.
\]

With this inequality at hand, we use the fact that for any \( \eta > 0 \), we have with probability approaching 1

\[
\phi(g, h) \geq \phi_{\lambda,\rho}(g, h) + \varepsilon.
\]

Hence, for any \( \eta \leq \frac{\varepsilon}{2} \), we obtain

\[
\min_{x \in S} \mathcal{L}_{\lambda,\rho}(x) \geq \phi_{\lambda,\rho}(g, h) + \frac{\varepsilon}{2} - \eta.
\]

and hence, Statement 3 holds with \( \bar{\phi}_{\lambda,\rho} = \bar{\phi} + \frac{\varepsilon}{2} \).

Proof of (92). In what follows, we will thus consider proving (92). As a first step, we use the weak law of large numbers to prove the following convergence:

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(x^i) - \bar{\alpha} \mathcal{L} = 0.
\]
Equivalently, for any $\hat{\epsilon} > 0$, with probability approaching 1, we have
\[
|F(\mathbf{x}^{AO}) - \kappa| \leq \hat{\epsilon}.
\]
(93)

To continue, let $\mathbf{x} \in S$, then,
\[
|F(\mathbf{x}) - \kappa| \geq 2\epsilon.
\]
Hence, using (93) yields
\[
|F(\mathbf{x}) - F(\mathbf{x}^{AO})| \geq |F(\mathbf{x}) - \kappa| - |F(\mathbf{x}^{AO}) - \kappa| \geq 2\epsilon - \hat{\epsilon}.
\]
(94)

Since $f$ is pseudo-Lipschitz of order $k$, there exists a constant $C$ such that:
\[
\frac{C}{n} \sum_{i=1}^{n} |x_i|^{k-1} + \left| \mathbf{x}^{AO} \right|^{k-1} |x_i| - \left| \mathbf{x}^{AO} \right|
\]
(95)

Since the absolute values of elements of $\mathbf{x}$ and $\mathbf{x}^{AO}$ are bounded by $\sqrt{n}$, then
\[
\max \left( \sqrt{n} \sum_{i=1}^{n} |x_i|^{2(k-1)}, \sqrt{n} \sum_{i=1}^{n} |\mathbf{x}^{AO}|^{2(k-1)} \right) \leq P^{k-1}.
\]
Hence,
\[
|F(\mathbf{x}) - F(\mathbf{x}^{AO})| \leq \frac{C}{n} \left| \mathbf{x} - \mathbf{x}^{AO} \right| \left( 1 + 2P^{k-1} \right).
\]
When combined with (94), this shows that
\[
|F(\mathbf{x}) - F(\mathbf{x}^{AO})| \leq \frac{C}{n} \left| \mathbf{x} - \mathbf{x}^{AO} \right| \left( 1 + 2P^{k-1} \right).
\]
(96)

Proof of Statement 1). Recall that from the proof of Statement 3) we have
\[
\tilde{\phi}_S = \tilde{\phi} + \frac{1}{2} \hat{\epsilon}.
\]
Hence, with $\eta \leq \frac{1}{2} \epsilon$, we have
\[
3\eta \leq \frac{1}{2} \hat{\epsilon} \Rightarrow \tilde{\phi} + 3\eta \leq \tilde{\phi} + \tilde{\epsilon} = \frac{1}{2} \hat{\epsilon}.
\]

REFERENCES


