Asymptotics of Nonlinear LSE Precoders with Applications to Transmit Antenna Selection

Ali Bereyhi, Mohammad Ali Sedaghat, Ralf R. Müller
Institute for Digital Communications (IDC), Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)
ali.bereyhi@fau.de, mohammad.sedaghat@fau.de, ralf.r.mueller@fau.de

Abstract—This paper studies the large-system performance of Least Square Error (LSE) precoders which minimize the input-output distortion over an arbitrary support subject to a general penalty function. The asymptotics are determined via the replica method in a general form which encloses the Replica Symmetric (RS) and Replica Symmetry Breaking (RSB) ansätze. As a result, the “marginal decoupling property” of LSE precoders for b-steps of RSB is derived. The generality of the studied setup enables us to address special cases in which the number of active transmit antennas are constrained. Our numerical investigations depict that the computationally efficient forms of LSE precoders based on “ℓ₁-norm” minimization perform close to the cases with “zero-norm” penalty function which have a considerable improvements compared to the random antenna selection. For the case with BPSK signals and restricted number of active antennas, the results show that RS fails to predict the performance while the RSB ansatz is consistent with theoretical bounds.

I. INTRODUCTION
For the Multiple-Input Multiple-Output (MIMO) channel

\[ y = Hx + z \]

with \( H \in \mathbb{C}^{k \times n}, x \in \mathbb{X}^{n} \) and \( z \sim \mathcal{C}\mathcal{N}(0, \sigma_{z}I_{n}) \), the nonlinear Least Square Error (LSE) precoder with the general penalty function \( u(\cdot) \) is given by

\[ x = \arg \min_{v \in \mathbb{X}^{n}} \|Hv - \sqrt{\rho}s\|^{2} + u(v). \]

The mapping is such that the distortion caused by the channel impact, i.e., \( \|Hx - \sqrt{\rho}s\|^{2} \), is minimized over the given input support \( \mathbb{X}^{n} \) subject to some constraints imposed by \( u(\cdot) \). The conventional precoding schemes such as Regularized Zero Forcing (RZF), Tomlinson-Harashima or vector precoding, mostly consider the average transmit power constraint and assume the set of possible input constellation points to be the complex plane, i.e., \( \mathbb{X} = \mathbb{C} \). The latter consideration was partially relaxed in [1] where authors studied the “per-antenna constant envelope precoding”. The set of possible constellation points was later generalized to an arbitrary set by introducing a class of power-limited nonlinear precoders [2]. The precoder in [2] generalizes the earlier schemes by letting different types of constraints be imposed on the precoded vector. In fact, due to the generality of the penalty function the scope of restrictions on \( x \) is broaden. Consequently, several precoding schemes are considered as special cases of [2]. To name some examples, let \( u(v) = \lambda \|v\|^{2} \); then, for \( \mathbb{X} = \mathbb{C} \), the precoder reduces to the RZF precoder introduced in [3], and by considering \( \mathbb{X} = \{v \in \mathbb{C} ; |v| = K\} \) for some constant \( K \), the precoder reduces to a constant envelope precoder [1].

This paper investigates the asymptotic performance of the precoder. Our motivation comes from recent promising results reported for massive MIMO systems [4]. For some choices of \( \mathbb{X} \) and \( u(\cdot) \), the system can be asymptotically analyzed via tools from random matrix theory [5]. The tools, however, fail to study the large-system performance of the precoder for many other choices. Therefore, we invoke the “replica method” developed in statistical mechanics. In the context of multiuser systems, the replica method was initially utilized by Tanaka in [6] to study the asymptotic performance of randomly spread CDMA detectors. The method was later widely employed for large-system analysis in communications and information theory; see for example [7] and the references therein.

Contributions
For nonlinear LSE precoders, we determine the input-output distortion, as well as the marginal distribution of output entries, in the large-system limit via the replica method. We deviate from our earlier replica symmetric study in [8], by determining the general replica ansatz which includes both the replica symmetry and symmetry breaking ansätze. Our general result furthermore depicts that under any assumed replicas’ structure, the output symbols of the precoder marginally decouple in the asymptotic regime. A brief introduction to the replica method is given in the appendix through the large-system analysis. As an application, we study special cases of the precoder with constraints on the number of active antennas. Our numerical investigations show that computationally efficient LSE precoders based on \( \ell_{1}\)-norm minimization perform significantly close to LSE precoders with zero-norm penalty. Moreover, the problem of BPSK transmission with constraint on the number of active antennas is shown to exhibit replica symmetry breaking.

Notation
We represent scalars, vectors and matrices with non-bold, bold lower case and bold upper case letters, respectively. A \( k \times k \) identity matrix is shown by \( I_{k} \), and the \( k \times k \) matrix with all entries equal to one is denoted by \( 1_{k} \). \( H^{\dagger} \) indicates the Hermitian of the matrix \( H \). The set of real and integer numbers are denoted by \( \mathbb{R} \) and \( \mathbb{Z} \), and their corresponding non-negative subsets by superscript +; moreover, \( \mathbb{C} \) represents the complex plane. For \( s \in \mathbb{C}, \operatorname{Re}\{s\} \) and \( \angle s \) identify the real part and argument, respectively. \( \|\cdot\| \) and \( \|\cdot\|_{1} \) denote the Euclidean and
\( \ell_1 \)-norm, respectively, and \( \|x\|_0 \) represents the zero
- norm defined as the number of nonzero entries. For a random variable
\( x \), \( p_x \) represents either the probability mass or density function.
Moreover, \( E \) identifies the expectation operator. For sake of
compactness, the set of integers \( \{1, \ldots, n\} \) is abbreviated as
\( [1 : n] \) and a zero-mean complex Gaussian distribution with
variance \( \sigma \) is represented by \( \phi(\cdot; \sigma) \). Whenever needed, we
assume the support \( \mathcal{X} \) to be discrete. The results, however, are
in full generality and hold also for continuous distributions.

II. PROBLEM FORMULATION

Consider the precoding scheme illustrated in (2) in which
(a) \( H_{k \times n} \) is a random matrix whose eigendecomposition is
\( H^H = UDU^H \) with \( U_{n \times n} \) being a Haar distributed
unitary matrix, and \( D_{n \times n} \) being a diagonal matrix with
asymptotic eigenvalue distribution \( p_D \).
(b) \( s_{k \times 1} \) has independent and identically distributed (i.i.d.)
zero-mean and unit-variance complex Gaussian entries,
i.e., \( s \sim \mathcal{CN}(0, I_k) \) and is independent of \( H \).
(c) \( \rho \) is a non-negative real power control factor.
(d) \( u(\cdot) \) is a general penalty function with decoupling prop-
terty, i.e., \( u(v) = \sum_{i=1}^n u(v_i) \).
(e) The dimensions of \( H \) grow large, such that the load fac-
tor, defined as \( \alpha := k/n \), is kept fixed in both \( k \) and \( n \).
For this setup, we define the asymptotic marginal as follows.

**Definition 1 (Asymptotic Marginal):** Consider the function
\( f((\cdot)) : \mathcal{X} \to \mathbb{R} \). The marginal of \( f(x) \) over \( \mathcal{W}(n) \subseteq [1 : n] \) is
\( M^W_{f}(x; n) := \frac{1}{|\mathcal{W}(n)|} E \sum_{w \in \mathcal{W}(n)} f(x_w) \) (3)

The asymptotic marginal of \( f(x) \) is then defined to be the
limit of \( M^W_{f}(v; n) \) as \( n \uparrow \infty \), i.e.,
\( M^W_{f}(x) := \lim_{n \uparrow \infty} M^W_{f}(x; n) \).

The asymptotic marginal of \( f(x) \) determines large-system characteristics of \( x \) including the marginal distribution of its
entries. In order to quantify the large-system performance, we
further define the asymptotic distortion as a measure.

**Definition 2 (Asymptotic Distortion):** For the precoder given in (2), the asymptotic input-output distortion is defined as
\( D(\rho) := \lim_{k \uparrow \infty} \frac{1}{k} E \|HX - \sqrt{\rho} s\|^2 \). (4)

III. MAIN RESULTS

We start by defining the R-transform of a distribution.

**Definition 3 (R-transform):** For \( t \) with distribution \( p_t \), the
Stieltjes transform over the upper complex half plane is given by
\( G_t(s) = E(t - s)^{-1} \). Denoting the inverse with respect to \( (w.r.t) \) the
composition by \( G^{-1}_t(\cdot) \), the R-transform of \( p_t \) is defined as
\( R_t(\omega) = G^{-1}_t(\omega) - \omega^{-1} \) such that \( \lim_{\omega \downarrow 0} R_t(\omega) = E_t \).
Moreover, let \( M_{n \times n} \) be decomposed as \( M = U \Lambda U^{-1} \) where
\( \Lambda_{n \times n} \) is the diagonal matrix of eigenvalues, and \( U_{n \times n} \) is the
matrix of eigenvectors. Then \( R_t(M) \) is an \( n \times n \) matrix defined as
\( R_t(M) = U \text{diag}[R_t(\lambda_1), \ldots, R_t(\lambda_n)] U^{-1} \).

Proposition 1 expresses \( M^W_{f}(x) \) and \( D(\rho) \) in terms of the
R-transform of \( p_D \). The result is determined for a general
structure of replicas, and only relies on the replica continuity assumption which is briefly explained in the appendix.

**Proposition 1 (General Replica Ansatz):** Consider the non-linear LSE precoder in Section II, and define \( v_{n \times 1} \) to be a random vector over \( \mathcal{X}^n \) with the distribution \( p_n^v(x; Q) \)
\( p_n^v(x; Q) = e^{-\beta[v^H\mathbb{T}_D(-\beta TQ)v + u(v)]} \sum_v e^{-\beta[v^H\mathbb{T}_D(-\beta TQ)v + u(v)]} \) (5)
for some \( m \times m \) matrix \( Q \) with real entries, non-negative real
scalar \( \beta \), and \( T := I_m - \frac{\beta \rho}{1 + m \beta \rho} I_m \). Let \( Q^* \) satisfy
\( Q^* = \sum_v p_n^v(v; Q^*)vv^H \). (6)

Then, under the replica continuity assumption, the asymptotic
marginal of \( f(x) \) is given by
\( M^W_{f}(x) = \lim_{\beta \uparrow \infty} \lim_{m \downarrow 0} \sum \rho v_n^v(v; Q)M^W_{f}(v; m) \) (7)
and \( D(\rho) = \rho + \alpha^{-1} \lim_{\beta \uparrow \infty} D^R(\beta) \) where \( D^R(\beta) \) is defined as
\( D^R(\beta) := \frac{\partial}{\partial \beta} \left[ \lim_{m \downarrow 0} \frac{1}{m} \text{Tr} \left\{ \int_0^\beta TQ^R(\omega - \omega TQ^*)d\omega \right\} \right] \)
\( - \beta \lim_{m \downarrow 0} \frac{1}{m} \text{Tr} \left\{ TQ^R(-\beta TQ^*)\frac{\partial Q^*}{\partial \beta} \right\} \) (8)

**Proof:** The proof is briefly addressed in the appendix. The de-
tails, however, are omitted due to lack of space and will be forthcoming in the extended version of the paper.

To determine \( M^W_{f}(x) \) and \( D(\rho) \) in Proposition 1, one needs to
determine the fixed-point \( Q^* \) through (6), and then, find the
function at the right hand side (r.h.s.) of (7) and \( D^R(\beta) \)
in an analytic form. Finding the solution of (6), however, is
notoriously difficult and possibly some of the solutions are not
of use. The trivial approach is to restrict the search to a set
of parameterized matrices. The most primary set is given by
Replica Symmetry Breaking (RSB) structure which we address in the sequel.

A. General Marginal Decoupling Property

Proposition 1 enables us to investigate a more general form of the
“asymptotic marginal decoupling property” introduced in [8]. The property indicates that in the large-system limit, the
marginal distribution of all output entries are identical and ex-
pressed as the output distribution of an equivalent single-user
system. In fact, it can be considered as a dual version of the
decoupling property investigated in the literature for different
classes of nonlinear estimators, e.g. [9–11]. As the analysis in [8]
was under the RS assumption, the result was limited to
the cases in which RS assumption gives a valid prediction. The
generality of Proposition 1, however, enables us to investigate
this property of the precoder for any structure of replicas. To illustrate the property, consider the following definition.

**Definition 4:** Denote the marginal distribution of the jth entry of $x_{n \times 1}$, i.e., $x_j$, for some $j \in [1 : n]$, by $p[n]_x(t)$ where the superscript $n$ indicates the dependency on the length of $x$. Then, the asymptotic marginal distribution $p[n]_x^*$ is defined to be the limit of $p[n]_x(t)$ as $n \uparrow \infty$, i.e., $p[n]_x^* (t) := \lim_{n \uparrow \infty} p[n]_x(t)$.

**General Marginal Decoupling Property:** Consider the non-linear LSE precoder with the constraints given in Section 11. Then, under the replica continuity assumption, the asymptotic marginal distribution $p[n]_x^*$ converges to a deterministic distribution which is constant in $j$ for any $j \in [1 : n]$ regardless of the structure imposed on $Q^n$.

**B. RSB Ansätze**

Parisi proposed the method of RSB to construct a set of parameterized matrices which recursively extends to larger classes. The method starts from the RS structure for $Q^n$, and then recursively constructs new structures. After $b$ steps of recursion, $Q^n$ becomes of the form

$$Q^n = \frac{\chi^n}{\beta} I_m + \sum_{k=1}^{b} c_n I_{nk} \odot 1_{mk} + 1_{pm},$$

(9)

for some non-negative real scalars $\chi$, $\beta$, $p$, and sequences $\{c_n\}$ and $\{\mu_n\}$. The structure in (9) reduces to RS by setting $\{c_n\} \equiv 0$. By substituting (9) in Proposition 11, the b-steps RSB ansatz is determined. For cases that the RS ansatz gives the exact solution, the coefficients $\{c_n\}$ at the saddle points are equal to zero. However, in cases that RS fails, the sequence $\{c_n\}$ has non-zero entries. The investigations in [2] show that the RS ansatz clearly fails giving a valid prediction of the performance in some cases. Therefore, the RSB ansätze are required to be considered further. For sake of compactness, we state the one-step RSB ansatz, i.e., $b = 1$, in this paper. The result, however, is extended to an arbitrary number of breaking steps by taking the approach in Appendix D of [12].

**Corollary 1 (One-step RSB Ansatz):** Let the assumptions in Proposition 11 hold, and consider $Q^n$ to be of the form (9) with $b = 1$. For given $\chi$, $p$, $\mu$, and $c$, define $\rho^s$ and $\rho_1^{rsb}$ as

$$\rho^s = \frac{\xi^2}{\beta} \frac{\partial}{\partial \chi} [(\rho \chi - p)R_D(-\chi)]$$

(10a)

$$\rho_1^{rsb} = \frac{\xi^2}{\beta} \mu^{-1} [R_D(-\chi) - R_D(-\chi)]$$

(10b)

where $\chi := \mu c$ and $\xi := \frac{1}{R_D(-\chi) - R_D(-\chi)}$. Let $x$ be

$$x = \arg \min_v \|v - s^c - s_1^{rsb}\|^2 + \xi u(v).$$

(11)

where $s^c \sim \phi(\cdot; \rho^s)$, and $s_1^{rsb}$ is obtained by passing $s^c$ through

$$p_1^{rsb}(u|t) = \frac{e^{-\mu_x|x-u|^2 - |u|^2} - \mu u(x)}{\int_{E} e^{-\mu_x|x-u|^2 - |u|^2} - \mu u(x)} \phi(w; \rho_1^{rsb})dw$$

Then, $M_f^n(x) = E f(x)$, and the asymptotic distortion reads

$$D(p) = \rho + \alpha - 1 \left\{ \frac{\partial}{\partial \chi} [(\rho - \rho \chi)R_D(-\chi)] + \frac{\xi - \chi}{\xi^2} \right\}.$$  (13)

In (12) and (13), $c$ and $p$ are determined via the equations

$$c + p = E|x|^2$$

(14a)

$$p + \chi = \frac{\xi}{\rho_1^{rsb}} E \Re \{x^*s_1^{rsb}\}$$

(14b)

and $\mu$ satisfies the following fixed-point equation

$$\frac{\mu^2}{\xi^2} \rho_1^{rsb} + \frac{\mu}{\xi} + I = 1(s_1^{rsb}; s^c) + D_{KL}(p_1^{rsb}(\cdot; \rho_1^{rsb}))$$

(15)

where $p_1^{rsb}(u) = \int p_1^{rsb}(u|t)\phi(t; \rho_1^{rsb})dt$. $D_{KL}(|\cdot|)$ denotes the Kullback-Leibler divergence, and $I := - \int_x R_D(-\omega) d\omega$.

**Remark:** The ansatz in Corollary 1 reduces to RS [8] by enforcing the fixed-point solution to have $c = 0$. The RS ansatz, however, is not necessarily valid. The valid solution here is chosen such that the corresponding free energy is minimized.

**RSB Marginal Decoupling Property:** Considering the one-step RSB ansatz, the asymptotic marginal distributions of the precoded symbols are described by $x$; more precisely, for any $j \in [1 : n]$ we have $p[n]_{x_j} \equiv p_c$. The distribution can be described by an equivalent single-user system which we refer to as the “decoupled precoder”, and is defined as

$$x^{\text{dec}}(s^{\text{dec}}) = \arg \min_v \|v - s^{\text{dec}}\|^2 + \xi u(v).$$

(16)

The one-step RSB decoder is similar to RS; however, the “decoupled input” $s^{\text{dec}}$, which in RS is $s^c$, is replaced by $s^c + s_1^{rsb}$. Taking the same approach as in [12], it is shown that under b-steps of RSB, the decoupled precoder has a same form, and $s^{\text{dec}} = s^c + \sum_{i=1}^{b} s_i^{rsb}$. In this case, $s_i^{rsb}$ is obtained from $s^c$ and $s_{n-1}^{rsb}$ through $p_1^{rsb}(u|k; u_{k+1}, \ldots, u_b)$.  

**IV. APPLICATIONS TO TRANSMIT ANTENNA SELECTION**

As we discussed, considering a general penalty function lets us investigate several transmit constraints. Restrictions on the number of active antennas is a constraint which arises in MIMO systems with Transmit Antenna Selection (TAS) [13]. The goal in these systems is to minimize the number of Radio Frequency (RF) chains which significantly reduces the overall RF-cost. The fundamental limits as well as efficient selection algorithms, however, have not been yet precisely addressed in the literature. In this section, we investigate the asymptotics of some special cases of the LSE precoder which imply TAS.

**A. TAS by Zero-Norm Minimization**

The LSE precoder with $u(v) = \lambda \|v\|^2 + \lambda_0 \|v\|_0$ imposes constraints on the average transmit power and number of active antennas. For $X = C$, the decoupled precoder reads

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} \frac{s^{\text{dec}}}{1 + \lambda_\delta} & \text{if } |s^{\text{dec}}| \geq \tau_0 \\ 0 & \text{if } |s^{\text{dec}}| < \tau_0 \end{cases}$$

(17)
for $\tau_0 := \sqrt{\xi \lambda_0 (1 + \xi \lambda)}$. Here, the decoupled precoder is a hard thresholding operator. As $\lambda_0 \downarrow 0$, $\tau_0$ tends to zero as well. For the case with limited peak power where for some $P \in \mathbb{R}^+$ we have

$$X = \left\{ r e^{j \theta} : 0 \leq \theta \leq 2\pi \wedge 0 \leq r \leq \sqrt{P} \right\},$$

(18)

the decoupled precoder is given by

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} s^{\text{dec}} \sqrt{P} & \tilde{\tau}_0 \leq |s^{\text{dec}}| < \tau_0 \\ 0 & |s^{\text{dec}}| \leq \tilde{\tau}_0 \end{cases}$$

(19)

where $\tilde{\tau}_0 = (1 + \xi \lambda)\sqrt{P}$ and $\tau_0 = \max \left\{ \tilde{\tau}_0, \tilde{\tau}_0/2 + \tau_0^2/2\tilde{\tau}_0 \right\}$. The decoupled precoder in (19) is a two-steps hard thresholding operator which in the first step constrains the transmit peak power, and in the second step, implies the TAS constraint. By setting $\lambda_0 = 0$, $\tau_0$ becomes zero and $\tilde{\tau}_0 = \tilde{\tau}_0$.

The LSE precoders with zero-norm penalty function need to minimize a non-convex function which has a high computational complexity. We therefore propose an alternative form of the precoder based on the $\ell_1$-norm minimization.

**B. TAS by $\ell_1$-Norm Minimization**

To reduce the complexity of the precoding schemes in Section IV-A we modify $u(\cdot)$ as $u(v) = \lambda \|v\|^2 + \lambda_1 \|v\|_1$. The objective function in this case is convex, and therefore, for convex choices of $X$, the resulting form of the LSE precoder is effectively implemented by employing computationally feasible algorithms. We start by considering $X = \mathcal{C}$ in which

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} s^{\text{dec}} \sqrt{P} & |s^{\text{dec}}| \geq \tau_1 \\ 0 & |s^{\text{dec}}| < \tau_1 \end{cases}$$

(20)

with $\tau_1 := \xi \lambda_1/2$. The decoupled precoder in this case is a soft thresholding operator. In fact, (20) is obtained from (17) by multiplying the factor $1 - \tau_1/|s^{\text{dec}}|$. Similar to (17), the threshold in (20) tends to zero as $\lambda_1 \downarrow 0$. For the case with limited peak transmit power, the decoupled precoder reads

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} s^{\text{dec}} \sqrt{P} & \tilde{\tau}_1 \leq |s^{\text{dec}}| < \tau_1 \\ 0 & |s^{\text{dec}}| \geq \tau_1 \end{cases}$$

(21)

for $\tau_1 := \xi \lambda_1/2$ and $\tilde{\tau}_1 := \sqrt{P(1 + \xi \lambda)} + \xi \lambda_1/2$. As in (19), the decoupled precoder in (21) is a two-steps thresholding. In the first step, $s^{\text{dec}}$ is constrained w.r.t. the peak power $P$ via a hard thresholding operator with level $\tilde{\tau}_1$, and then at the second step, the TAS constraint is imposed on the decoupled input by a soft thresholding operator as in (20). By setting $\lambda_1 = 0$, the threshold $\tau_1$ reads $\tau_1 = 0$ and $\tilde{\tau}_1 = \sqrt{P(1 + \xi \lambda)}$. The decoupled precoder is given by

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} s^{\text{dec}} \sqrt{P} & \tilde{\tau}_0 \leq |s^{\text{dec}}| < \tau_0 \\ 0 & |s^{\text{dec}}| \leq \tilde{\tau}_0 \end{cases}$$

(19)

C. TAS with M-PSK Signals on Antennas

Considering the precoding support as $X = \{0, \sqrt{P} e^{j \omega k M} \}$, for $k \in [1 : M]$, the precoder is constrained to map the source to a vector of M-PSK symbols over a subset of antennas while keeping the others silent. In this case, the transmit power on each active antenna is $P$, and therefore, $\|x\|^2 = P \|x\|_0$ which indicates that any restriction on the average transmit power imposes a proportional constraint on the number of active antennas. Consequently, TAS is applied via the LSE precoder by setting the penalty function as $u(v) = \lambda \|v\|^2$. By defining the function $\psi(\cdot)$ as $\psi(k) := \cos \left( \frac{2k\pi}{M} - \angle s^{\text{dec}} \right)$, the decoupled precoder in this case is derived as

$$x^{\text{dec}}(s^{\text{dec}}) = \begin{cases} \sqrt{P} e^{j \omega k M} & \|s^{\text{dec}}\| = \tau_d \\ 0 & \|s^{\text{dec}}\| < \tau_d \end{cases}$$

(22)

where $\tau_d := \sqrt{P(1 + \xi \lambda)} \psi(k^*)^{-1}/2$ for $k^* := \arg \max_k \psi(k)$. As in Sections IV-A and IV-B, (22) describes a thresholding operator over the M-PSK constellation. Here, by growth of $\lambda$, the threshold $\tau_d$ increases, and consequently, the number of active transmit antennas reduces.

D. Numerical Results

Throughout the numerical investigations, the asymptotic fraction of active antennas is denoted by $\eta$ which is determined by $\eta = \mathbb{E} \{ x^{\text{dec}}(s^{\text{dec}}) \neq 0 \}$ with $\{ \cdot \}$ being the indicator function. The average transmit power is represented by $P$, and the PAPR is denoted by $PAPR$ which reads $PAPR = P/P$. We consider $\mathbf{H}$ to be a fading channel whose entries are i.i.d. with zero mean and variance $1/\nu$; thus, $\rho_P$ follows Marcenko-Pastur’s law, and $\mathbb{P}(\omega) = \alpha(1 - \omega)^{-3}$.

Considering Sections IV-A and IV-B. Fig. 1 shows the RS-predicted D$(\rho)$ vs. $\alpha^{-1}$ for $P = 0.5$ considering no PAPR limitation and $PAPR = 3$ dB. The zero-norm and $\ell_1$-norm precoders save $35\%$ and $22\%$ of active antennas in case of no PAPR restriction, and about $25\%$ and $20\%$ when $PAPR = 3$ dB, respectively.
Thus, the evaluation of $D(\rho)$ and $\mathcal{M}_T(x)$ reduce to determining $\mathcal{F}(\cdot)$; the task which we do via the replica method. Using the Riesz equality which states $\mathbb{E} \log x = \lim_{m \uparrow 0} m^{-1} \log \mathbb{E} x^m$,

$$\mathcal{F}(\beta, h) = \frac{1}{n} \lim_{m \downarrow 0} \frac{1}{m} \log \mathbb{E} \left[ \mathcal{Z}(\beta, h) \right]^m. \quad (26)$$

**Replica Method:** Evaluating $\mathcal{F}(\beta, h)$ from (26) is not trivial, as $m \in \mathbb{R}^+$. The replica method determines the r.h.s. of (26) by conjecturing the replica continuity. The replica continuity indicates that the “analytic continuation” of the non-negative integer moment function, i.e., $\mathbb{E} \left[ \mathcal{Z}(\beta, h) \right]^m$ for $m \in \mathbb{Z}^+$, onto $\mathbb{R}^+$ equals to the non-negative real moment function, i.e., $\mathbb{E} \left[ \mathcal{Z}(\beta, h) \right]^m$ for $m \in \mathbb{R}^+$. The rigorous justification of the replica continuity has not been yet precisely addressed; however, the analytic results from the theory of spin glasses confirm the validity of the conjecture for several cases.

Considering the replica continuity assumption, Proposition 1 is concluded by taking some lines of calculations from (26) which have been left for the extended version of the manuscript due to the page limitation.

**REFERENCES**

[1] S. K. Mohammed and E. G. Larsson, “Per-antenna constant envelope precoding for large multi-user MIMO systems,” IEEE Trans. on Comm., vol. 61, no. 3, pp. 1059–1071, 2013.

[2] M. A. Sedaghat, A. Bereyhi, and R. Mueller, “LSE precoders for massive MIMO with hardware constraints: Fundamental limits,” arXiv preprint arXiv:1612.07902, 2016.

[3] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multuser communication-Part I: channel inversion and regularization,” IEEE Trans. on Comm., vol. 53, no. 1, pp. 195–202, 2005.

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

[5] D. A. Schmidt, M. Joham, and W. Utschick, “Minimum mean square error vector precoding,” European Transactions on Telecommunications, vol. 19, no. 3, pp. 219–231, 2008.

[6] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiantenna detectors,” IEEE Trans. on Inf. Theory, vol. 48, no. 11, pp. 2888–2910, 2002.

[7] B. M. Zaidel, R. Müller, A. L. Moustakas, and R. de Miguel, “Vector precoding for gaussian MIMO broadcast channels: Impact of replica symmetry breaking,” IEEE Trans. on Inf. Theory, vol. 58, no. 3, pp. 1413–1440, 2012.

[8] A. Bereyhi, M. A. Sedaghat, S. Asaad, and R. R. Müller, “Nonlinear precoders for massive MIMO systems with general constraints,” International ITG Workshop on Smart Antennas (WSA), 2017.

[9] D. Guo and S. Verdù, “Randomly spread CDMA: Asymptotics via statistical physics,” IEEE Trans. on Inf. Theory, vol. 51, no. 6, pp. 1983–2010, 2005.

[10] S. Rangan, A. K. Fletcher, and V. Goyal, “Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing,” in IEEE Trans. on Info. Theory, 2012, pp. 1902–1923.

[11] A. Bereyhi, R. Müller, and H. Schulz-Baldes, “RSB decoupling property of MAP estimators,” IEEE Inf. Theory Work. (ITW), pp. 379–383, 2016.

[12] A. Bereyhi, R. R. Müller, and H. Schulz-Baldes, “Statistical mechanics of MAP estimation: General replica ansatz,” arXiv preprint arXiv: 1612.01980, 2016.

[13] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, “Capacity of MIMO systems with antenna selection,” IEEE Trans. on Wireless Comm., vol. 4, no. 4, pp. 1759–1772, 2005.

[14] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” Mathematics of the USSR Sbornik, vol. 1, no. 4, pp. 457–483, 1967.