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 "ℓ_1-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

$$\boldsymbol{y} = \mathbf{H}\,\boldsymbol{x} + \boldsymbol{z} \tag{1}$$

with $\mathbf{H} \in \mathbb{C}^{k\times n}$, $\boldsymbol{x} \in \mathbb{X}^n$ and $\boldsymbol{z} \sim \mathcal{CN}(\mathbf{0}, \lambda_z \mathbf{I}_k)$, the nonlinear Least Square Error (LSE) precoder with the general penalty function $u(\cdot)$ is given by

$$\boldsymbol{x} = \arg\min_{\boldsymbol{v}\in\mathbb{X}^n} \|\mathbf{H}\boldsymbol{v} - \sqrt{\rho}\,\boldsymbol{s}\|^2 + u(\boldsymbol{v}). \tag{2}$$

The precoder maps the k-dimensional source vector $\boldsymbol{s}$, scaled with the power control factor ρ, to the n-dimensional input vector $\boldsymbol{x}$ whose entries are taken from the given support $\mathbb{X}$. The mapping is such that the distortion caused by the channel impact, i.e., $\|\mathbf{H}\,\boldsymbol{x} - \sqrt{\rho}\,\boldsymbol{s}\|^2$, 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 $\boldsymbol{x}$ is broaden. Consequently, several precoding schemes are considered as special cases of (2). To name some examples, let $u(\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2$; then, for $\mathbb{X} = \mathbb{C}$, the precoder reduces to the RZF precoder introduced in [3], and by considering $\mathbb{X} = \{v \in \mathbb{C} : |v| = \mathsf{K}\}$ 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 ℓ_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 $\mathbf{I}_k$, and the $k \times k$ matrix with all entries equal to one is denoted by $\mathbf{1}_k$. $\mathbf{H}^{\mathsf{H}}$ indicates the Hermitian of the matrix $\mathbf{H}$. 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 $s \in \mathbb{C}$, $\mathsf{Re}\{s\}$ and $\sphericalangle s$ identify the real part and argument, respectively. $\|\cdot\|$ and $\|\cdot\|_1$ denote the Euclidean and

This work was supported by the German Research Foundation, Deutsche Forschungsgemeinschaft (DFG), under Grant No. MU 3735/2-1.

ℓ_1-norm, respectively, and $\|\boldsymbol{x}\|_0$ 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 ρ is represented by $\phi(\cdot;\rho)$. Whenever needed, we assume the support $\mathbb{X}$ 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) $\mathbf{H}_{k\times n}$ is a random matrix whose eigendecomposition is $\mathbf{H}^{\mathsf{H}}\mathbf{H}=\mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{H}}$ with $\mathbf{U}_{n\times n}$ being a Haar distributed unitary matrix, and $\mathbf{D}_{n\times n}$ being a diagonal matrix with asymptotic eigenvalue distribution $\mathrm{p}_{\mathbf{D}}$.
(b) $\boldsymbol{s}_{k\times 1}$ has independent and identically distributed (i.i.d.) zero-mean and unit-variance complex Gaussian entries, i.e., $\boldsymbol{s}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_k)$ and is independent of $\mathbf{H}$.
(c) ρ is a non-negative real power control factor.
(d) $u(\cdot)$ is a general penalty function with decoupling property, i.e., $u(\boldsymbol{v})=\sum_{j=1}^{n}u(v_j)$.
(e) The dimensions of $\mathbf{H}$ grow large, such that the load factor, 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 $f(\cdot):\mathbb{X}\mapsto\mathbb{R}$. The marginal of $f(\boldsymbol{x})$ over $\mathbb{W}(n)\subseteq[1:n]$ is

$$\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{x};n):=\frac{1}{|\mathbb{W}(n)|}\,\mathsf{E}\sum_{w\in\mathbb{W}(n)}f(x_w) \tag{3}$$

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

The asymptotic marginal of $f(\boldsymbol{x})$ determines large-system characteristics of $\boldsymbol{x}$ 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

$$\mathsf{D}(\rho):=\lim_{k\uparrow\infty}\frac{1}{k}\mathsf{E}\,\|\mathbf{H}\boldsymbol{x}-\sqrt{\rho}\,\boldsymbol{s}\|^2. \tag{4}$$

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 $\mathrm{G}_t(s)=\mathsf{E}\,(t-s)^{-1}$. Denoting the inverse with respect to (w.r.t.) composition by $\mathrm{G}_t^{-1}(\cdot)$, the R-transform of p_t is defined as $\mathrm{R}_t(\omega)=\mathrm{G}_t^{-1}(-\omega)-\omega^{-1}$ such that $\lim_{\omega\downarrow 0}\mathrm{R}_t(\omega)=\mathsf{E}\,t$.

Moreover, let $\mathbf{M}_{n\times n}$ be decomposed as $\mathbf{M}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}$ where $\boldsymbol{\Lambda}_{n\times n}$ is the diagonal matrix of eigenvalues, and $\mathbf{U}_{n\times n}$ is the matrix of eigenvectors. Then $\mathrm{R}_t(\mathbf{M})$ is an $n\times n$ matrix defined as $\mathrm{R}_t(\mathbf{M})=\mathbf{U}\ \mathrm{diag}[\mathrm{R}_t(\lambda_1),\ldots,\mathrm{R}_t(\lambda_n)]\ \mathbf{U}^{-1}$.

Proposition 1 expresses $\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{x})$ and $\mathsf{D}(\rho)$ in terms of the R-transform of $\mathrm{p}_{\mathbf{D}}$. 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 nonlinear LSE precoder in Section II, and define $\mathbf{v}_{m\times 1}$ to be a random vector over $\mathbb{X}^m$ with the distribution $\mathrm{p}_{\mathbf{v}}^{\beta}(\mathbf{v};\mathbf{Q})$

$$\mathrm{p}_{\mathbf{v}}^{\beta}(\mathbf{v};\mathbf{Q})=\frac{e^{-\beta\left[\mathbf{v}^{\mathsf{H}}\mathbf{T}\mathrm{R}_{\mathbf{D}}(-\beta\mathbf{T}\mathbf{Q})\mathbf{v}+u(\mathbf{v})\right]}}{\sum_{\mathbf{v}}e^{-\beta\left[\mathbf{v}^{\mathsf{H}}\mathbf{T}\mathrm{R}_{\mathbf{D}}(-\beta\mathbf{T}\mathbf{Q})\mathbf{v}+u(\mathbf{v})\right]}}. \tag{5}$$

for some $m\times m$ matrix $\mathbf{Q}$ with real entries, non-negative real scalar β, and $\mathbf{T}:=\mathbf{I}_m-\dfrac{\beta\rho}{1+m\beta\rho}\mathbf{1}_m$. Let $\mathbf{Q}^{\star}$ satisfy

$$\mathbf{Q}^{\star}=\sum_{\mathbf{v}}\mathrm{p}_{\mathbf{v}}^{\beta}(\mathbf{v};\mathbf{Q}^{\star})\mathbf{v}\mathbf{v}^{\mathsf{H}}. \tag{6}$$

Then, under the replica continuity assumption, the asymptotic marginal of $f(\boldsymbol{x})$ is given by

$$\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{x})=\lim_{\beta\uparrow\infty}\lim_{m\downarrow 0}\sum_{\mathbf{v}}\mathrm{p}_{\mathbf{v}}^{\beta}(\mathbf{v};\mathbf{Q})\mathsf{M}_f^{\mathbb{T}}(\mathbf{v};m), \tag{7}$$

and $\mathsf{D}(\rho)=\rho+\alpha^{-1}\lim_{\beta\uparrow\infty}\mathcal{D}^{\mathsf{R}}(\beta)$ where $\mathcal{D}^{\mathsf{R}}(\cdot)$ is defined as

$$\begin{aligned}\mathcal{D}^{\mathsf{R}}(\beta):=&\frac{\partial}{\partial\beta}\left[\lim_{m\downarrow 0}\frac{1}{m}\mathrm{Tr}\left\{\int_0^{\beta}\mathbf{T}\mathbf{Q}^{\star}\mathrm{R}_{\mathbf{D}}(-\omega\mathbf{T}\mathbf{Q}^{\star})\mathrm{d}\omega\right\}\right]\\&-\beta\lim_{m\downarrow 0}\frac{1}{m}\mathrm{Tr}\left\{\mathbf{T}\mathrm{R}_{\mathbf{D}}(-\beta\mathbf{T}\mathbf{Q}^{\star})\frac{\partial\mathbf{Q}^{\star}}{\partial\beta}\right\}.\end{aligned} \tag{8}$$

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

To determine $\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{x})$ and $\mathsf{D}(\rho)$ in Proposition 1, one needs to determine the fixed-point $\mathbf{Q}^{\star}$ through (6), and then, find the function at the right hand side (r.h.s.) of (7) and $\mathcal{D}^{\mathsf{R}}(\beta)$ 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 (RS). The RS solution, however, may result in an invalid prediction of the performance. A more general structure is given by imposing the 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 expressed 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 $\boldsymbol{x}_{n\times 1}$, i.e., x_j for some $j \in [1:n]$, by $\mathrm{p}_x^{j(n)}$ where the superscript n indicates the dependency on the length of $\boldsymbol{x}$. Then, the asymptotic marginal distribution p_x^j is defined to be the limit of $\mathrm{p}_x^{j(n)}$ as $n \uparrow \infty$, i.e., $\mathrm{p}_x^j(t) \coloneqq \lim_{n\uparrow\infty} \mathrm{p}_x^{j(n)}(t)$.

General Marginal Decoupling Property: Consider the nonlinear LSE precoder with the constraints given in Section II. Then, under the replica continuity assumption, the asymptotic marginal distribution p_x^j converges to a deterministic distribution which is constant in j for any $j \in [1:n]$ regardless of the structure imposed on $\mathbf{Q}^\star$.

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 $\mathbf{Q}^\star$, and then recursively constructs new structures. After b steps of recursion, $\mathbf{Q}^\star$ becomes of the form

$$\mathbf{Q}^\star = \frac{\chi}{\beta}\mathbf{I}_m + \sum_{\kappa=1}^{b} \mathsf{c}_\kappa \, \mathbf{I}_{\frac{m\beta}{\mu_\kappa}} \otimes \mathbf{1}_{\frac{\mu_\kappa}{\beta}} + \mathsf{p}\mathbf{1}_m, \tag{9}$$

for some non-negative real scalars χ, β and p, and sequences $\{\mathsf{c}_\kappa\}$ and $\{\mu_\kappa\}$. The structure in (9) reduces to RS by setting $\{\mathsf{c}_\kappa\} \equiv 0$. By substituting (9) in Proposition 1, the b-steps RSB ansatz is determined. For cases that the RS ansatz gives the exact solution, the coefficients $\{\mathsf{c}_\kappa\}$ at the saddle points are equal to zero. However, in cases that RS fails, the sequence $\{\mathsf{c}_\kappa\}$ 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 1 hold, and consider $\mathbf{Q}^\star$ to be of the form (9) with $b = 1$. For given χ, p, μ and c, define ρ^{rs} and ρ_1^{rsb} as

$$\rho^{\mathsf{rs}} = \xi^2 \frac{\partial}{\partial \tilde{\chi}} \left[(\rho\tilde{\chi} - \mathsf{p}) \mathrm{R}_{\mathbf{D}}(-\tilde{\chi}) \right] \tag{10a}$$

$$\rho_1^{\mathsf{rsb}} = \xi^2 \mu^{-1} \left[\mathrm{R}_{\mathbf{D}}(-\chi) - \mathrm{R}_{\mathbf{D}}(-\tilde{\chi}) \right] \tag{10b}$$

where $\tilde{\chi} \coloneqq \chi + \mu\mathsf{c}$ and $\xi \coloneqq [\mathrm{R}_{\mathbf{D}}(-\chi)]^{-1}$. Let x be

$$\mathrm{x} = \arg\min_v |v - s^{\mathsf{rs}} - s_1^{\mathsf{rsb}}|^2 + \xi\, u(v). \tag{11}$$

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

$$\mathrm{p}_1^{\mathsf{rsb}}(u|t) = \frac{e^{-\frac{\mu}{\xi}\left[|\mathrm{x}-u-t|^2 - |u+t|^2\right] - \mu u(\mathrm{x})} \phi(u;\rho_1^{\mathsf{rsb}})}{\int_{\mathbb{C}} e^{-\frac{\mu}{\xi}\left[|\mathrm{x}-w-t|^2 - |w+t|^2\right] - \mu u(\mathrm{x})} \phi(w;\rho_1^{\mathsf{rsb}}) \mathrm{d}w} \tag{12}$$

Then, $\mathsf{M}_f^{\mathsf{W}}(\boldsymbol{x}) = \mathsf{E}\, f(\mathrm{x})$, and the asymptotic distortion reads

$$\mathsf{D}(\rho) = \rho + \alpha^{-1} \left\{ \frac{\partial}{\partial \tilde{\chi}} [(\mathsf{p} - \rho\tilde{\chi})\tilde{\chi} \mathrm{R}_{\mathbf{D}}(-\tilde{\chi})] + \frac{\xi\mathsf{p} - \tilde{\chi}\rho_1^{\mathsf{rsb}}}{\xi^2} \right\}. \tag{13}$$

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

$$\mathsf{c} + \mathsf{p} = \mathsf{E}\,|\mathrm{x}|^2 \tag{14a}$$

$$\mathsf{p} + \tilde{\chi} = \frac{\xi}{\rho_1^{\mathsf{rsb}}} \mathsf{E}\,\mathsf{Re}\left\{ \mathrm{x}^* s_1^{\mathsf{rsb}} \right\} \tag{14b}$$

$$\tilde{\chi} = \frac{\xi}{\rho^{\mathsf{rs}}} \mathsf{E}\,\mathsf{Re}\left\{ \mathrm{x}^* s^{\mathsf{rs}} \right\}. \tag{14c}$$

and μ satisfies the following fixed-point equation

$$\frac{\mu^2 \mathsf{p}}{\xi^2}\rho_1^{\mathsf{rsb}} + \frac{\mu\mathsf{c}}{\xi} + \mathcal{I} = \mathrm{I}\left(s_1^{\mathsf{rsb}}; s^{\mathsf{rs}}\right) + \mathrm{D}_{\mathsf{KL}}(\mathrm{p}_{s_1^{\mathsf{rsb}}} \| \phi(\cdot;\rho_1^{\mathsf{rsb}})) \tag{15}$$

where $\mathrm{p}_{s_1^{\mathsf{rsb}}}(u) = \int \mathrm{p}_1^{\mathsf{rsb}}(u|t)\phi(t;\rho^{\mathsf{rs}})\mathrm{d}t$, $\mathrm{D}_{\mathsf{KL}}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence, and $\mathcal{I} \coloneqq -\int_{\chi}^{\tilde{\chi}} \mathrm{R}_{\mathbf{D}}(-\omega)\mathrm{d}\omega$.

Remark: The ansatz in Corollary 1 reduces to RS [8], by enforcing the fixed-point solution to have $\mathsf{c} = 0$. 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 $\mathrm{p}_x^j \equiv \mathrm{p_x}$. The distribution can be described by an equivalent single-user system which we refer to as the "decoupled precoder", and is defined as

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \arg\min_v |v - s^{\mathsf{dec}}|^2 + \xi\, u(v). \tag{16}$$

The one-step RSB decoupled precoder is similar to RS; however, the "decoupled input" s^{dec}, which in RS is s^{rs}, is replaced by $s^{\mathsf{rs}} + s_1^{\mathsf{rsb}}$. Taking the same approach as in [12], it is shown that under b-steps of RSB, the decoupled precoder has a same form, and $s^{\mathsf{dec}} = s^{\mathsf{rs}} + \sum_{\kappa=1}^{b} s_\kappa^{\mathsf{rsb}}$. In this case, s_κ^{rsb} is obtained from s^{rs} and $\left\{ s_\varsigma^{\mathsf{rsb}} \right\}_{\varsigma=\kappa+1}^{b}$ through $\mathrm{p}_\kappa^{\mathsf{rsb}}(u_\kappa | u_{\kappa+1}, \dots, u_b, t)$.

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

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \begin{cases} \dfrac{s^{\mathsf{dec}}}{1+\xi\lambda} & |s^{\mathsf{dec}}| \geq \tau_0 \\ 0 & |s^{\mathsf{dec}}| < \tau_0 \end{cases} \tag{17}$$

for $\tau_0 \coloneqq \sqrt{\xi\lambda_0(1+\xi\lambda)}$. Here, the decoupled precoder is a hard thresholding operator. As $\lambda_0 \downarrow 0$, τ_0 tends to zero as well. For the case with limited peak power where for some $\mathrm{P} \in \mathbb{R}^+$

$$\mathbb{X} = \left\{ re^{\mathrm{j}\theta} : \ 0 \leq \theta \leq 2\pi \ \wedge \ 0 \leq r \leq \sqrt{\mathrm{P}} \right\}, \tag{18}$$

the decoupled precoder is given by

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \begin{cases} \dfrac{s^{\mathsf{dec}}}{|s^{\mathsf{dec}}|}\sqrt{\mathrm{P}} & \hat{\tau}_0 \leq |s^{\mathsf{dec}}| \\ 0 & \tilde{\tau}_0 \leq |s^{\mathsf{dec}}| < \hat{\tau}_0 \\ \dfrac{s^{\mathsf{dec}}}{1+\xi\lambda} & \tau_0 \leq |s^{\mathsf{dec}}| \leq \tilde{\tau}_0 \\ 0 & 0 \leq |s^{\mathsf{dec}}| < \tau_0 \end{cases} \tag{19}$$

where $\tilde{\tau}_0 = (1+\xi\lambda)\sqrt{\mathrm{P}}$ and $\hat{\tau}_0 = \max\left\{\tilde{\tau}_0, \tilde{\tau}_0/2 + \tau_0^2/2\tilde{\tau}_0\right\}$. 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 $\lambda_0 = 0$, τ_0 becomes zero and $\hat{\tau}_0 = \tilde{\tau}_0$.

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 ℓ_1-norm minimization.

B. TAS by ℓ_1-Norm Minimization

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

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \begin{cases} \dfrac{s^{\mathsf{dec}}}{1+\xi\lambda}\dfrac{|s^{\mathsf{dec}}|-\tau_1}{|s^{\mathsf{dec}}|} & |s^{\mathsf{dec}}| \geq \tau_1 \\ 0 & |s^{\mathsf{dec}}| < \tau_1 \end{cases} \tag{20}$$

with $\tau_1 \coloneqq \xi\lambda_1/2$. The decoupled precoder in this case is a soft thresholding operator. In fact, (20) is obtained from (17) by multiplying the factor $1 - \tau_1/|s^{\mathsf{dec}}|$. 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

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \begin{cases} \dfrac{s^{\mathsf{dec}}}{|s^{\mathsf{dec}}|}\sqrt{\mathrm{P}} & \tilde{\tau}_1 \leq |s^{\mathsf{dec}}| \\ \dfrac{s^{\mathsf{dec}}}{1+\xi\lambda}\dfrac{|s^{\mathsf{dec}}|-\tau_1}{|s^{\mathsf{dec}}|} & \tau_1 \leq |s^{\mathsf{dec}}| < \tilde{\tau}_1 \\ 0 & 0 \leq |s^{\mathsf{dec}}| < \tau_1 \end{cases} \tag{21}$$

for $\tau_1 \coloneqq \xi\lambda_1/2$ and $\tilde{\tau}_1 \coloneqq \sqrt{\mathrm{P}}(1+\xi\lambda) + \xi\lambda_1/2$. As in (19), the decoupled precoder in (21) is a two-steps thresholding. In the first step, s^{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 τ_1 reads $\tau_1 = 0$ and $\tilde{\tau}_1 = \sqrt{\mathrm{P}}(1+\xi\lambda)$.

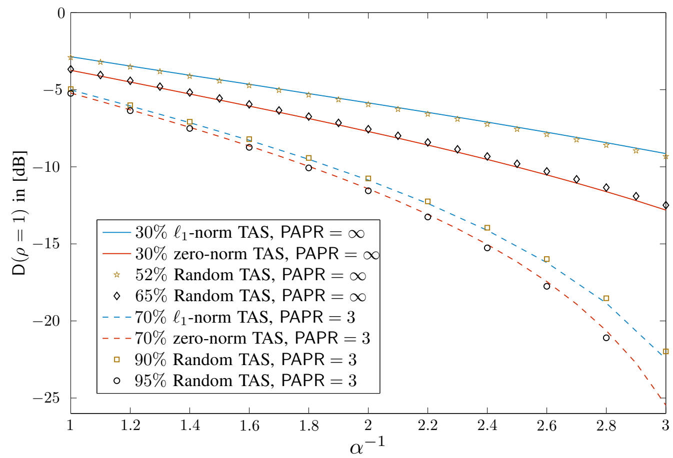

Fig. 1: RS-predicted $\mathsf{D}(\rho)$ vs. α^{-1} for $\mathsf{P} = 0.5$ considering no PAPR limitation and $\mathsf{PAPR} = 3$ dB. The zero-norm and ℓ_1-norm precoders save 35% and 22% of active antennas in case of no PAPR restriction, and about 25% and 20% when $\mathsf{PAPR} = 3$ dB, respectively.

C. TAS with M-PSK Signals on Antennas

Considering the precoding support as $\mathbb{X} = \{0, \sqrt{\mathrm{P}}e^{\mathrm{j}\frac{2k\pi}{\mathsf{M}}}\}$, for $k \in [1 : \mathsf{M}]$, 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, $\|\boldsymbol{x}\|^2 = \mathrm{P}\|\boldsymbol{x}\|_0$ 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 $u(\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2$. By defining the function $\psi(\cdot)$ as $\psi(k) \coloneqq \cos\left(\frac{2k\pi}{\mathsf{M}} - \sphericalangle s^{\mathsf{dec}}\right)$, the decoupled precoder in this case is derived as

$$\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) = \begin{cases} \sqrt{\mathrm{P}}e^{\mathrm{j}\frac{2k^\star\pi}{\mathsf{M}}} & |s^{\mathsf{dec}}| \geq \tau_{\mathsf{d}} \\ 0 & |s^{\mathsf{dec}}| < \tau_{\mathsf{d}} \end{cases} \tag{22}$$

where $\tau_{\mathsf{d}} \coloneqq \sqrt{\mathrm{P}}(1+\xi\lambda)\psi(k^\star)^{-1}/2$ for $k^\star \coloneqq \arg\max_k \psi(k)$. As in Sections IV-A and IV-B, (22) describes a thresholding operator over the M-PSK constellation. Here, by growth of λ, the threshold τ_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 η which is determined by $\eta = \mathsf{E}\,\mathbf{1}\left\{\mathrm{x}^{\mathsf{dec}}(s^{\mathsf{dec}}) \neq 0\right\}$ with $\mathbf{1}\{\cdot\}$ being the indicator function. The average transmit power is represented by P, and the PAPR is denoted by PAPR which reads $\mathsf{PAPR} = \mathrm{P}/\mathsf{P}$. We consider $\mathbf{H}$ to be a fading channel whose entries are i.i.d. with zero mean and variance $1/n$; thus, $\mathrm{p}_{\mathbf{D}}$ follows Marcenko-Pastur's law, and $\mathrm{R}_{\mathbf{D}}(\omega) = \alpha(1-\omega)^{-1}$ [14].

Considering Sections IV-A and IV-B, Fig. 1 shows the RS predicted asymptotic distortion at $\rho = 1$ in terms of the inverse load factor for two cases of $\mathsf{PAPR} = 3$ dB and no peak power constraint. In the PAPR-limited case, the curves have been sketched for $\eta = 0.7$, and in the other case, $\eta = 0.3$ has been considered; moreover, the average transmit power is set to be

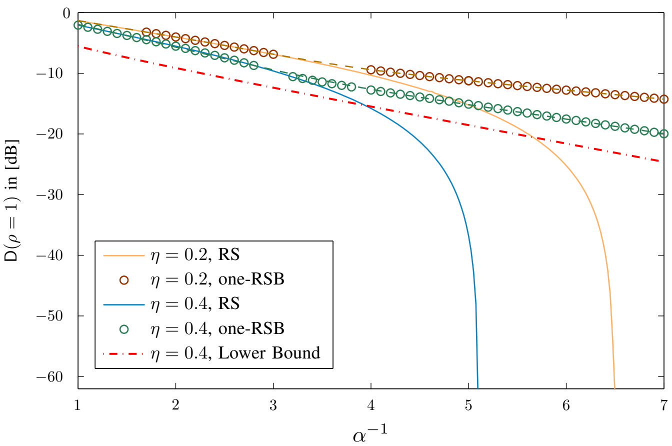

Fig. 2: RS- and one-step RSB-predicted $\mathsf{D}(\rho)$ for BPSK signals with $\mathsf{P} = 1$ under TAS. As α^{-1} grows, RS violates the lower bound. The RSB ansatz, however, is consistent with the lower bound.

$\mathsf{P} = 0.5$. As a benchmark, we have also plotted the points for random TAS which meet the corresponding curves. In fact, in the random TAS, the precoder selects a subset of transmit antennas randomly and precodes $\boldsymbol{s}$ using the penalty function $u(\boldsymbol{v}) = \lambda \|\boldsymbol{v}\|^2$. As the figure depicts, for the case of no peak power restriction, the zero-norm and ℓ_1-norm based precoders need respectively about 35% and 22% fewer active transmit antennas compared to the random TAS. The gains in the case of $\mathsf{PAPR} = 3$ dB reduce to 25% and 20% respectively.

In order to investigate the impact of RSB, we have also considered an example of antenna selection with BPSK transmission, i.e., $\mathsf{M} = 2$ in Section IV-C. Fig. 2 illustrates the RS as well as one-step RSB prediction of the asymptotic distortion at $\rho = 1$ for two cases of $\eta = 0.2$ and $\eta = 0.4$ when $\mathsf{P} = 1$. For sake of comparison, a theoretically rigorous lower bound for the case of $\eta = 0.4$ has been also sketched. The lower bound is derived as in [2, Appendix C]. As the figure shows, the RS ansatz starts to fail predicting the asymptotic distortion as α^{-1} grows, and it even violates the lower bound in large inverse load factors. For this regime of α^{-1}, however, the one-step RSB ansatz gives a theoretically valid prediction.

Appendix: Large-System Analysis

In the sequel, we briefly sketch the derivations. Consider the Hamiltonian $\mathcal{E}(\boldsymbol{v}|\boldsymbol{s}, \mathbf{H}) = \|\mathbf{H}\boldsymbol{v} - \sqrt{\rho}\,\boldsymbol{s}\|^2 + u(\boldsymbol{v})$, and define the partition function $\mathcal{Z}(\beta, h)$ to be

$$\mathcal{Z}(\beta, h) = \sum_{\boldsymbol{v}} e^{-\beta \mathcal{E}(\boldsymbol{v}|\boldsymbol{s},\mathbf{H}) + hn\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{v};n)}. \tag{23}$$

By a standard large deviation argument, it is shown that

$$\mathsf{M}_f^{\mathbb{W}}(\boldsymbol{x}) = \lim_{n\uparrow\infty} \lim_{\beta\uparrow\infty} \frac{\partial}{\partial h} \mathcal{F}(\beta, h)|_{h=0}, \tag{24}$$

in which $\mathcal{F}(\beta, h) \coloneqq n^{-1} \mathsf{E} \log \mathcal{Z}(\beta, h)$. Moreover, the asymptotic distortion reads $\alpha \mathsf{D}(\rho) + \mathsf{M}_u^{\mathbb{T}}(\boldsymbol{x}) = \tilde{\mathcal{E}}$ where we define $\mathbb{T}(n) \coloneqq [1:n]$, and $\tilde{\mathcal{E}} = \lim_{n\uparrow\infty} n^{-1} \mathsf{E}\, \mathcal{E}(\boldsymbol{x}|\boldsymbol{s}, \mathbf{H})$. $\mathsf{M}_u^{\mathbb{T}}(\boldsymbol{x})$ is determined in terms of $\mathcal{F}(\cdot)$ by setting $f(x) = u(x)$ in (24), and

$$\tilde{\mathcal{E}} = -\lim_{n\uparrow\infty} \lim_{\beta\uparrow\infty} \frac{\partial}{\partial \beta} \mathcal{F}(\beta, h)|_{h=0}. \tag{25}$$

Thus, the evaluation of $\mathsf{D}(\rho)$ and $\mathsf{M}_f^{\mathbb{T}}(\boldsymbol{x})$ reduce to determining $\mathcal{F}(\cdot)$; the task which we do via the replica method. Using the Riesz equality which states $\mathsf{E} \log \mathrm{x} = \lim_{m\downarrow 0} m^{-1} \log \mathsf{E}\, \mathrm{x}^m$,

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

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., $\mathsf{E} \left[\mathcal{Z}(\beta, h)\right]^m$ for $m \in \mathbb{Z}^+$, onto $\mathbb{R}^+$ equals to the non-negative real moment function, i.e., $\mathsf{E} \left[\mathcal{Z}(\beta, h)\right]^m$ 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 form (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 multiuser communication-Part I: channel inversion and regularization," *IEEE Trans. on Comm.*, vol. 53, no. 1, pp. 195–202, 2005.

[4] J. Hoydis, 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 multiuser 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. Verdú,"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 *IIEEE Trans. on Inf. 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.