Towards Designing Optimal Sensing Matrices for Non-linear Inverse Problems

Anonymous Author(s)
Affiliation
Address
email

Abstract

We consider a nonlinear inverse problem $\boldsymbol{y} = f(\boldsymbol{A}\boldsymbol{x})$, where observations $\boldsymbol{y} \in \mathbb{R}^m$ are the componentwise nonlinear transformation of $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} \in \mathbb{R}^n$ is the signal of interest and $\boldsymbol{A}$ is a known linear mapping. By properly specifying the nonlinear processing function, this model can be particularized to many signal processing problems, including compressed sensing and phase retrieval.

Our main goal in this paper is to understand the impact of sensing matrices, or more specifically the spectrum of sensing matrices, on the difficulty of recovering $\boldsymbol{x}$ from $\boldsymbol{y}$. Towards this goal, we study the performance of one of the most successful recovery methods, i.e. the expectation propagation algorithm (EP). We define a notion for the spikiness of the spectrum of $\boldsymbol{A}$ and show the importance of this measure in the performance of the EP. Whether the spikiness of the spectrum can hurt or help the recovery performance of EP depends on f. We define certain quantities based on the function f that enables us to describe the impact of the spikiness of the spectrum on EP recovery. Based on our framework, we are able to show that for instance, in phase-retrieval problems, matrices with spikier spectrums are better for EP, while in 1-bit compressed sensing problems, less spiky (flatter) spectrums offer better recoveries. Our results unify and substantially generalize the existing results that compare sub-Gaussian and orthogonal matrices, and provide a platform toward designing optimal sensing systems.

1 Introduction

1.1 Problem statement and contributions

Consider the problem of estimating a signal $\boldsymbol{x} \in \mathbb{R}^n$ from the nonlinear measurements:

$$\boldsymbol{y} = f\big(\boldsymbol{A}\boldsymbol{x}\big), \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a measurement (or sensing) matrix and $f : \mathbb{R} \mapsto \mathcal{Y}$ is a function accounting for possible nonlinear effect of the measuring process. Here, the function $f(\cdot)$ is applied to $\boldsymbol{A}\boldsymbol{x}$ in a component-wise manner. The above model arises in many applications of signal processing [12, 9, 38], communications [51, 8, 23], and machine learning [44, 37]. For instance, the phase retrieval problem, which is a special case of (1) with $f(z) = |z|$, has received significant interest in recent years [12, 11, 14, 49, 55, 22, 24, 3, 43, 30, 4].

The main goal of our work is to understand the impact of the sensing matrix, or more specifically the spectrum of the sensing matrix, on the difficulty of recovering the signal $\boldsymbol{x}$ from its measurements $\boldsymbol{y}$. In many applications, one has certain level of freedom in designing the sensing matrix (e.g., transmitter design in communications or the masks used in phase retrieval application) and hence understanding the impact of the sensing matrix on the recovery algorithms is the first step toward

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

the optimal design of such systems. Rather than studying the information theoretic limits, where the computational complexity of the recovery algorithm is ignored, we would like to study the impact of the spectrum of the sensing matrix on efficient algorithms that are used in applications. For this reason, we consider one of the most successful recovery algorithms that has received substantial attention in the last few years, i.e. expectation propagation (EP) [33, 35] (referred to as GLM-EP in this paper[1]), and study the impact of the spectrum of the sensing matrix on the performance of this algorithm. The EP algorithm studied here is an instance of the algorithm introduced in [20, 26] and is closely related to the orthogonal AMP (OAMP) [28] and vector AMP (VAMP) [39] algorithms (in that all these algorithms use divergence-free denoising functions [28]).

Similar to the approximate message passing (AMP) algorithm [15], GLM-EP has two distinguishing features: (i) Its asymptotic performance could be characterized exactly by a simple dynamical system (with very few states) called the state evolution (SE). (ii) It is conjectured that AMP or GLM-EP achieve the optimal performance among polynomial time algorithms [2, 13]. Based on the SE framework, we investigate the impact of the spectrum of the sensing matrix $\boldsymbol{A}$ on the performance of GLM-EP. It turns out that the "spikiness" (or conversely "flatness") of the spectrum of the sensing matrix spectrum has a major impact on the performance of GLM-EP. To formalize this statement, we first define a measure of "spikiness" of the spectrum based on Lorenz partial order [1]. We show that whether the spikiness of the spectrum benefits or hurts GLM-EP depends on the choice of the nonlinear mapping f. For instance, spikier spectrums help the performance of phase retrieval problem (where $f(x) = |x|$) but hurt the performance of 1-bit compressed sensing (where $f(x) = \text{sign}(x)$). We will characterize the classes of functions on which spikiness hurts or helps GLM-EP based on the monotonicity of a function (which is related to the scalar minimum mean square error) that will be defined in the next section. As a byproduct of our studies, we will also show that more spiky spectrums are better in terms of the number of measurements required to achieve perfect recovery (i.e., zero MSE) for any f. Furthermore, when the spectrum is spiky enough, the number of measurements required by GLM-EP to achieve perfect recovery approaches the information theoretical lower bound.

1.2 Related Work.

Approximate message passing (AMP) [15, 6, 7, 38, 5, 42, 45, 10, 18, 19] and expectation propagation (EP) [33, 35, 28, 20, 39, 46, 26, 47, 21] have been used extensively for solving the estimation problems similar to the one we have in (1). As a result of such studies, it is for instance known that partial orthogonal matrix is better than iid Gaussian matrix for noisy compressed sensing [29], and the spectral methods for phase retrieval perform better with iid Gaussian sensing matrices than coded diffraction pattern matrices [27, 34, 31, 17]. However, studying the impact of spectrum of the sensing matrix in the generality of our paper has not been done to the best of our knowledge. Recently, Maillard et al. [32] considered the phase retrieval problem and a sensing matrix which can be written as the product of Gaussian and another matrix. They reached the conclusion that the weak recovery threshold with this type of matrices can be made arbitrarily close to zero. As a special case of our results, we will also show that if we make the spectrum of the sensing matrix spiky GLM-EP can reach the information theoretic lower bounds in the phase retrieval problem.

Another venue of research that is also related to our work is the derivation of the information theoretic limits for analog compression schemes. Analog compression framework was first introduced in Wu and Verdú [53, 54] for compressed sensing. It was shown in Wu and Verdú [53, 54] that the minimum number of measurements required for successful signal reconstruction in an information theoretic framework is related to the Rényi information dimension of the signal distribution. Riegler and Tauböck [41] studied the phase retrieval problem using the analog compression framework and proved that (real-valued) phase retrieval has the same fundamental limit as that of compressed sensing. In order to compare the performance of GLM-EP on matrices with different spectrums we generalize the work of [53, 54] and [41] and obtain information theoretic limit for our sensing model. Note that while we are using such information theoretic tools, the problem we are studying in this paper is fundamentally different from the one studied in [53, 54, 41]. Here we are interested in the impact of the spectrum of the sensing matrix on the performance of GLM-EP, and information theoretic limits are mainly derived for comparison purposes (and evaluating the optimality of GLM-EP).

[1]The name GLM-EP is chosen because the model (1) is an instance of generalized linear models (GLM).

1.3 Definitions

In this section, we mention some definitions that will be frequently used throughout this paper. We first start with the Rényi information dimension of a random variable.

Definition 1 (Information dimension [40, 53])**.** *Let X be a real-valued random variable, and $\langle X \rangle_M = \lfloor MX \rfloor / M$ be a quantization operator.*[2] *Suppose the following limit exists*

$$d(X) = \lim_{M \to \infty} \frac{H\left(\langle X \rangle_M\right)}{\log M},$$

where $H(\cdot)$ is the entropy of a discrete random variable. The limit $d(X)$ is called the information dimension of X. Further, if $H(\lfloor X \rfloor) < \infty$, then $0 \leq d(X) \leq 1$.

As will be discussed later, $d(X)$ plays a critical role in the information theoretic lower bounds we derive for the recovery algorithms. The next lemma shows how $d(X)$ can be calculated for the simple distributions we observe in our applications.

Lemma 1 (Information dimension of mixed distribution [40, 53])**.** *Let X be a random variable such that $H(\lfloor X \rfloor)$ is finite. Suppose the distribution of X can be represented as*

$$P_X = (1-\rho) P_d + \rho P_c,$$

where P_d is a discrete measure and P_c is an absolutely continuous measure with respect to Lebesgue, and $0 \leq \rho \leq 1$. Then,

$$d(X) = \rho.$$

As we discussed before the spikiness of the spectrum can help or hurt GLM-EP depending on the nonlinearity f. The next notions we define, enable us to characterize the class of functions for which spikiness helps GLM-EP.

Definition 2 (MMSE for AWGN channel [25])**.** *Let (Z, U) be a pair of random variables. The MMSE $\mathsf{mmse}(Z, \mathsf{snr})$ and the conditional MMSE $\mathsf{mmse}(Z, \mathsf{snr}|U)$ given U are defined as*

$$\begin{aligned} \mathsf{mmse}(Z, \mathsf{snr}) &= \mathbb{E}\left[\left(Z - \mathbb{E}[Z|\sqrt{\mathsf{snr}}Z + N]\right)^2\right], \\ \mathsf{mmse}(Z, \mathsf{snr}|U) &= \mathbb{E}\left[\left(Z - \mathbb{E}[Z|\sqrt{\mathsf{snr}}Z + N, U]\right)^2\right], \end{aligned} \tag{2}$$

where $N \sim \mathcal{N}(0,1)$ is independent of (Z, U), and the outer expectation is taken over all random variables involved.

2 Information-theoretic limit for signal recovery

As we discussed before, our main objective is to evaluate the impact of the spectrum of the sensing matrices on the performance of GLM-EP. However, it is still useful to obtain the information theoretic lower bounds and compare what GLM-EP achieves (for different spectrums) with the information theoretic lower bounds. Hence, we devote this section to deriving the information theoretic lower bounds.

2.1 Assumptions

Before we proceed to the technical part of the paper, let us review and discuss some of the assumptions we will be making throughout the paper.

(A.1) The elements of $\boldsymbol{x}$ are independently drawn from P_X, which is an absolutely continuous distribution with respect to the Lebesgue measure. We assume $\mathbb{E}[X^2] = 1$.

(A.2) Let the SVD of $\boldsymbol{A}$ be $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\mathsf{T}$. $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ are independent Haar matrices, which are further independent of $\boldsymbol{\Sigma}$. Let $\{\sigma_i\}_{i=1}^n$ be the diagonal entries of $\boldsymbol{\Sigma}$ and define $\lambda_i \triangleq \sigma_i^2$. We assume that the empirical distribution of $\{\lambda_i\}_{i=1}^n$ converges almost surely to a deterministic limit P_λ with a compact support $[a, b]$ where $a > 0$. Further, we assume that $n^{-1} \sum_{i=1}^n \lambda_i \overset{a.s.}{\to} \mathbb{E}[\lambda] = m/n > 1$, where the expectation is with respect to P_λ.

[2] The notation $\lfloor z \rfloor$ denotes the largest integer that is smaller than z.

(A.3) $f : \mathbb{R} \mapsto \mathcal{Y}$ is a piecewise monotone function. Specifically, the domain $\mathbb{R}$ can be decomposed into $K \in \mathbb{N}_+$ non-overlapping intervals, and f is continuous and monotone on each interval. Furthermore, we assume $H(\lfloor f(Z) \rfloor) < \infty$ where $Z \sim \mathcal{N}(0,1)$.

Please note that Assumption (A.2) is a standard assumption in the theoretical analysis of GLM-EP [39, 46, 18]. Furthermore, all the nonlinearities that we observe in applications satisfy Assumption (A.3). We consider generic signal and do not impose any structural assumption. Finally, the independence assumption we have made in the prior of $\boldsymbol{x}$ is again standard in the literature of approximate message passing and expectation propagation [38, 20, 6, 52, 50]. One can weaken this assumption at the expense of making assumptions about the recovery algorithm.

2.2 Perfect reconstruction in a noiseless setting

As discussed before in this section we derive the information theoretic lower boud on the number of measurements required by Lipschitz recovery scheme to achieve vanishing error probability. Note that the computational complexity of the recovery algorithm is *not* of any concern in these lower bounds. We will later compare our results for GLM-EP with these information theoretic lower bounds.

Theorem 1 (Perfect reconstruction under Lipschitz decoding)**.** *Suppose Assumptions (A.1)-(A.3) hold. Suppose that there exists a Lipschitz continuous decoder $g : \mathcal{Y} \mapsto \mathbb{R}^n$ such that $\mathbb{P}\{\boldsymbol{x} \neq g(f(\boldsymbol{A}\boldsymbol{x}))\} \to 0$ as $m, n \to \infty$ and $m/n \to \delta \in (1, \infty)$, then necessarily we have*

$$\delta > \delta_{\text{opt}}^{\mathsf{p}} \triangleq \frac{1}{d(f(Z))}, \tag{3}$$

where $d(f(Z))$ is the information dimension of $f(Z)$, where $Z \sim \mathcal{N}(0,1)$. Here, the error probability is taken with respect to both $\boldsymbol{x}$ and $\boldsymbol{A}$.

The proof of this result can be found in the Supplementary Material. Intuitively speaking, $d(f(Z)) \in [0,1]$ may be interpreted as a measurement discount factor and the total number of effective measurements is $m \cdot d(f(Z))$.

2.3 Stable reconstruction in the noisy setting

Theorem 1 focuses on signal reconstruction for model (1) without any noise. For practical considerations, it is desirable to make sure that a small amount of measurement noise does not cause major performance degradation. In this paper, we consider the following noisy model[3]

$$\boldsymbol{y} = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}), \tag{4}$$

where $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I})$ is independent of $\boldsymbol{A}$ and $\boldsymbol{x}$. Define the *noise sensitivity* [16, 54] of the minimum mean square error (MMSE) estimator by

$$M^*(X, f, \delta) \triangleq \sup_{\sigma_w} \limsup_{n \to \infty} \frac{\frac{1}{n}\mathsf{mmse}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A})}{\sigma_w^2}, \tag{5}$$

where $\mathsf{mmse}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}) \triangleq \mathbb{E}\big[\,(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}])^2\,\big]$ is the MMSE of estimating $\boldsymbol{x}$ from $\boldsymbol{y}$, and in the above limit $n \to \infty$ is understood as $n \to \infty$ and $m/n \to \delta$. Theorem 2 below shows that to achieve bounded noise sensitivity, one needs $\delta > 1/d(f(Z))$, the same necessary condition for achieving vanishing error probability in the noiseless setting. Its proof can be found in the Supplementary Material.

Theorem 2 (Noise sensitivity)**.** *Suppose Assumptions (A.1)-(A.3) hold. Additionally, assume $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. A necessary condition for achieving bounded noise sensitivity, namely $M^*(X, f, \delta) < \infty$, is $\delta > \delta_{\text{opt}}^{\mathsf{p}}$, where $\delta_{\text{opt}}^{\mathsf{p}}$ is defined in* (3)*.*

2.4 Discussion of Theorems 1 and 2

Theorems 1 and 2 show that that the quantity $d(f(Z))$ determines the fundamental limit for signal recovery from the nonlinear model (1). Notice that $f(Z)$ is a mixed discrete-continuous distribution

[3] Other types of noisy models are possible, e.g., $\boldsymbol{y} = f(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{w}$. Extending our results to these models is beyond the scope of the current paper.

(by Assumption (A.3)), where the discrete component in $f(Z)$ corresponds to "flat" sections of f; see Figure 1 for illustration. According to Lemma 1, $d(f(Z))$ is simply the weight in the continuous component of the distribution of $f(Z)$, which is the probability of $Z \sim \mathcal{N}(0,1)$ falling into the non-flat sections of f. For illustration, Figure 1 shows three representative examples of f.

Type I: f is a piece-wise monotone function without flat sections; see the left panel of Figure 1 for illustration. This type of functions includes the absolute value function $f = |z|$, which appears in phase retrieval problems. For such functions, $f(Z)$ has an absolutely continuous distribution when $Z \sim \mathcal{N}(0,1)$, and hence $d(f(Z)) = 1$ according to Lemma 1.

Type II: f consists of purely flat sections. A special case is the quantization function. Clearly, $f(Z)$ has a discrete distribution and $d(f(Z)) = 0$.

Type III: f consists of both flat and non-flat sections, e.g., the function shown on the right panel of Figure 1. Note that such scenarios happen when sensors saturate for instance in the phase retrieval application. In this case, $f(Z)$ has a mixed discrete-continuous distribution and $0 < d(f(Z)) < 1$.

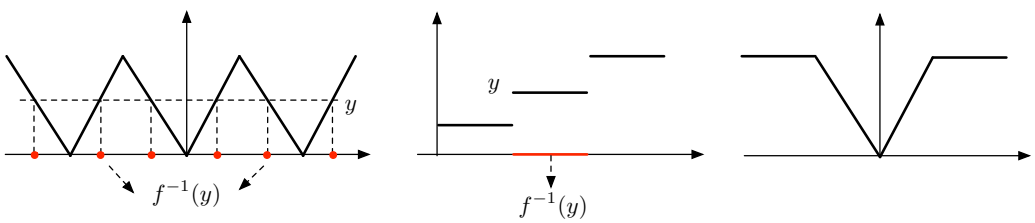

Figure 1: Three types of f. Left: $d(f(Z)) = 1$. Center: $d(f(Z)) = 0$. Right: $0 < d(f(Z)) < 1$.

3 GLM-EP algorithm and performance analysis

In this section, we introduce an expectation propagation (EP) [33, 35] type algorithm, referred to as GLM-EP, for solving our nonlinear inverse problem and derive its state evolution (SE). We then study the impact of the spectrum of the sensing matrix on the performance of this algorithm.

3.1 Summary of GLM-EP

The GLM-EP algorithm is summarized below. We use superscripts to represent iteration indices, and subscripts 'l' and 'r' to distinguish different variables.

Initialization: $\boldsymbol{z}_r^{-1} = \boldsymbol{0}$, $v_r^{-1} = 1$. For $t = 0, \ldots$, excute the following steps iteratively:

$$\boldsymbol{z}_l^t = \frac{1}{1 - \left\langle \eta_z'(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, v_r^{t-1}) \right\rangle} \cdot \left(\eta_z(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, v_r^{t-1}) - \left\langle \eta_z'(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, v_r^{t-1}) \right\rangle \cdot \boldsymbol{z}_r^{t-1} \right), \tag{6a}$$

$$v_l^t = v_r^{t-1} \cdot \frac{\left\langle \eta_z'(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, v_r^{t-1}) \right\rangle}{1 - \left\langle \eta_z'(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, v_r^{t-1}) \right\rangle}, \tag{6b}$$

$$\boldsymbol{R}^t \triangleq \boldsymbol{A} \left(v_l^t \boldsymbol{I} + \boldsymbol{A}^\mathsf{T} \boldsymbol{A} \right)^{-1} \boldsymbol{A}^\mathsf{T}, \tag{6c}$$

$$\boldsymbol{z}_r^t = \frac{1}{1 - \frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t)} \cdot \left(\boldsymbol{R}^t - \frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t) \cdot \boldsymbol{I} \right) \cdot \boldsymbol{z}_l^t, \tag{6d}$$

$$v_r^t = v_l^t \cdot \frac{\frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t)}{1 - \frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t)}, \tag{6e}$$

where η_z is defined by

$$\eta_z(z_r, y, v) \triangleq \frac{\int_{f^{-1}(y)} u \cdot \mathcal{N}(u; z_r, v) du}{\int_{f^{-1}(y)} \mathcal{N}(u; z_r, v) du}, \tag{6f}$$

and η_z' denotes the derivative of η_z with respect to the first argument. When $f^{-1}(y)$ is a discrete set, the integration in the above formula is simply replaced by a summation.

Output: $\hat{\boldsymbol{x}}_{\text{out}}^t = v_l^t(\boldsymbol{I} + v_l^t \boldsymbol{A}^\mathsf{T}\boldsymbol{A})^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{z}_l^t$.

In the above descriptions of the algorithm, we adopted the convention commonly used in the AMP literature: $\eta_z(\boldsymbol{z}_r, \boldsymbol{y}, v)$ denotes a vector with elements obtained by applying the scalar function η_z to the corresponding elements of $\boldsymbol{z}_r$ and $\boldsymbol{y}$, and $\langle\cdot\rangle$ denotes the empirical mean of a vector.

3.2 Asymptotic analysis

The asymptotic performance of GLM-EP could be described by two scalar sequences $\{V_l^t, V_r^t\}_{t\geq 0}$, defined recursively by

$$V_l^t = \left(\frac{1}{\mathsf{mmse}_z\left(V_r^{t-1}\right)} - \frac{1}{V_r^{t-1}}\right)^{-1} \triangleq \phi(V_r^{t-1}), \tag{7a}$$

$$V_r^t = \left(\frac{1}{\frac{1}{\delta}\cdot\mathbb{E}\left[\frac{V_l^t\lambda}{V_l^t+\lambda}\right]} - \frac{1}{V_l^t}\right)^{-1} \triangleq \Phi(V_l^t), \tag{7b}$$

where $V_r^{-1} = 1$, $\mathsf{mmse}_z(V_r) \triangleq \mathsf{mmse}\left(Z, V_r^{-1} - 1 | f(Z)\right)$, and the expectation in (7b) is w.r.t. the limiting eigenvalue distribution of $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$. (Recall that $\mathsf{mmse}(Z, \mathsf{snr}|U)$ denotes a conditional MMSE; see (2)). Equations (7a) and (7b) are known as the state evolution (SE) for the GLM-EP.

Roughly speaking, the deterministic sequences $\{V_l^t, V_r^t\}_{t\geq 0}$ are expected to be accurate predictions of $\{v_l^t, v_r^t\}_{t\geq 0}$ (which are generated by GLM-EP) asymptotically. We will formalize this claim below. Further, we will show that the per coordinate MSE of $\boldsymbol{x}_{\text{out}}^t$ (see Lemma 2 below) is characterized by

$$\mathsf{MSE}_\lambda(V_l^t) \triangleq \mathbb{E}\left[\frac{V_l^t}{V_l^t + \lambda}\right]. \tag{7c}$$

The subscript emphasizes the fact that the MSE depends on the limiting eigenvalue distribution P_λ.

Lemma 2 below gives a formal statement of the accuracy of SE, and its proof is mainly based on that of [21, Theorem 1]. Note that [21] requires the continuity of f and η_z. While this assumption holds for some problems, e.g. phase retrieval problem, it is violated for some other problems, e.g. one-bit quantization. Hence, we construct a new algorithm, called GLM-EP-app hereafter, which satisfies the requirements of [21]. This allows us to use SE for predicting the performance of this algorithm. Please note that if f and η_z are Lipschitz continuous, then we can use GLM-EP as well. GLM-EP-app uses the following iterations:

$$\boldsymbol{z}_l^t = C_t \cdot \left(\tilde{\eta}_z(\boldsymbol{z}_r^{t-1}, \boldsymbol{y}, V_r^{t-1}) - \mathbb{E}\left[\tilde{\eta}_z'(Z_r^{t-1}, Y, V_r^{t-1})\right]\cdot \boldsymbol{z}_r^{t-1}\right), \tag{8a}$$

$$\boldsymbol{z}_r^t = \frac{1}{1 - \frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t)} \cdot \left(\boldsymbol{R}^t - \frac{1}{m}\mathsf{Tr}(\boldsymbol{R}^t)\cdot \boldsymbol{I}\right)\cdot \boldsymbol{z}_l^t, \tag{8b}$$

where $\tilde{\eta}$ is a function for which $\mathbb{E}\left[\tilde{\eta}_z'(Z_r^{t-1}, Y, V_r^{t-1})\right]$ exists, $\boldsymbol{R}^t \triangleq \boldsymbol{A}\left(V_l^t\boldsymbol{I} + \boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\mathsf{T}$, and C_t is a sequence of fixed numbers. The choices we choose for $\tilde{\eta}$ and C_t (to make them close enough to GLM-EP) is discussed in the proof of Lemma 2. Finally, similar to GLM-EP the output of GLM-EP-app is given by

$$\hat{\boldsymbol{x}}_{\text{out}}^t = V_l^t(\boldsymbol{I} + V_l^t\boldsymbol{A}^\mathsf{T}\boldsymbol{A})^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{z}_l^t.$$

Lemma 2 shows that the performance of GLM-EP-app could be arbitrarily close to the SE prediction. The details of the proof can be found in Supplementary Material.

Lemma 2. *Suppose Assumptions (A.1)-(A.3) hold. Additionally, assume $f : \mathbb{R} \mapsto \mathcal{Y}$ to be Lipschitz continuous. Let $\{V_l^t, V_r^t\}_{t\geq 0}$ be generated according to* (7)*. There exists $\tilde{\eta}_z$ and $\{C_t\}_{t\geq 0}$ such that $\hat{\boldsymbol{x}}_{\text{out}}^t$ of GLM-EP-app satisfies*

$$\mathsf{MSE}_\lambda(V_l^t) - \epsilon \leq \frac{1}{m}\left\|\hat{\boldsymbol{x}}_{\text{out}}^t - \boldsymbol{x}\right\|^2 < \mathsf{MSE}_\lambda(V_l^t) + \epsilon, \quad \forall \epsilon > 0, \tag{9}$$

almost surely as $m, n \to \infty$ with $m/n \to \delta \in (1, \infty)$, where MSE_λ is defined in (7c)*.*

According to Lemma 2, the asymptotic MSE of GLM-EP-app in the large system limit as $t \to \infty$ can be obtained from the limiting value of V_r^t (or V_l^t). Because this quantity is of particular importance to us we will characterize it in the following lemma.

Lemma 3 (MSE performance)**.** *Suppose $\delta > 1$. Define $V_r^\star$ by*

$$V_r^\star \triangleq \inf\left\{v \in [0,1] \,:\, P(v_r) > 0, \forall v_r \in [v,1]\right\}. \tag{10}$$

where

$$P(v_r) \triangleq \mathbb{E}\left[\frac{\phi(v_r)}{\phi(v_r)+\lambda}\right] - \underbrace{\left[1-\delta\left(1-\frac{\mathsf{mmse}_z(v_r)}{v_r}\right)\right]}_{g(v_r)}. \tag{11}$$

Let $\{V_l^t, V_r^t\}_{t\geq 0}$ be sequences generated according to (7) *with $V_r^{-1} = 1$. We have*

$$\lim_{t\to\infty} V_r^t = V_r^\star.$$

Further, the final MSE is given by $\mathsf{MSE}_\lambda^\star \triangleq \mathsf{MSE}_\lambda(\phi(V_r^\star))$, where ϕ is defined in (7a)*.*

The proof of this lemma can be found in Supplementary Material. A direct consequence of Lemma 3 is the perfect reconstruction condition stated in Lemma 4 below.

Lemma 4 (Perfect reconstruction condition)**.** *Let $\{V_l^t, V_r^t\}_{t\geq 0}$ be a sequence generated through* (7) *with $V_r^{-1} = 1$. Then, the predicted final MSE $\mathsf{MSE}_\lambda^\star$ is zero if and only if*

$$P(v_r) > 0, \quad \forall v_r \in (0,1], \tag{12}$$

where $P(v_r)$ is defined in (11)*. Furthermore, a necessary condition for* (12) *is $\delta > \delta_{\text{opt}}^{\mathsf{p}}$ where $\delta_{\text{opt}}^{\mathsf{p}}$ is defined in* (3)*.*

The proofs can be found in the Supplementary Material.

4 Impact of sensing matrix spectrum

In this section, we use Lemmas 3 and 4 to study the impact of the sensing matrix on the MSE performance of GLM-EP-app. Before presenting our detailed analysis, we first discuss the so-called Lorenz order that compares the "spikiness" of different distributions.

4.1 A measure of spikiness of distributions

A natural tool to compare the spikiness of the distributions of two non-negative random variables is Lorenz partial order [1]. Lorenz order is well-known in economics to characterize the wealth inequality of different populations. Lorenz order is closely related to majorization, a tool that has been extensively studied for transceiver design in communication systems [36].

Definition 3 (Lorenz order [1])**.** *Consider a nonnegative random variable with cumulative density function $F(x)$. Let $F^{-1}(y)$ be the quantile function defined by*

$$F^{-1}(y) = \sup\{x : F(x) \leq y\}, \quad 0 < y < 1. \tag{13}$$

The Lorenz curve corresponding to $F(x)$ is defined by

$$L(u) = \frac{\int_0^u F^{-1}(y)dy}{\int_0^1 F^{-1}(y)dy}, 0 \leq u \leq 1.$$

Let X and Y be two nonnegative random variables, and $L_x(u)$ and $L_y(u)$ be the corresponding Lorenz curves. We say that X is less spiky than Y in the Lorenz sense, denoted as $X \preceq_L Y$, if $L_X(u) \geq L_Y(u)$ for every $u \in [0,1]$. Conversely, $X \succeq_L Y$ if $L_X(u) \leq L_Y(u)$ for every $u \in [0,1]$.

An important property of Lorenz ordering is the following.

Lemma 5 ([1])**.** *Suppose $X \geq 0$, $Y \geq 0$ and $\mathbb{E}[X] = \mathbb{E}[Y]$. We have $X \preceq_L Y$ if and only if $\mathbb{E}[h(X)] \leq \mathbb{E}[h(Y)]$ for every continuous convex function $h : \mathbb{R}_+ \to \mathbb{R}$.*

4.2 Impact on MSE

Let $\lambda_1 \sim P_{\lambda_1}$ and $\lambda_2 \sim P_{\lambda_2}$ be two limiting eigenvalue distributions of $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$. Let $V^\star_{\lambda_1}$ and $V^\star_{\lambda_2}$ denote the corresponding limiting values of V_r^t (as $t \to \infty$) in (7) (proving that the iterations (7) converge to a fixed point is straightforward). The associated MSEs, denoted as $\mathsf{MSE}^\star_{\lambda_1}$ and $\mathsf{MSE}^\star_{\lambda_2}$, can be compared according to the following lemma. .

Lemma 6. *Let $\delta > 1$. Suppose P_{λ_1} is more spiky than P_{λ_2} in the Lorenz sense, i.e., $\lambda_1 \succeq_L \lambda_2$. Define*

$$G(v_r) \triangleq \max\big(g(v_r),\, 0\big), \quad \forall v_r \in [0,1], \tag{14}$$

where $g(\cdot)$ is defined in (11)*. We have*

- *If $G(v_r)$ is increasing on $v_r \in [0,1]$, then $\mathsf{MSE}^\star_{\lambda_1} \leq \mathsf{MSE}^\star_{\lambda_2}$;*
- *If $G(v_r)$ is decreasing on $v_r \in [0,1]$, then $\mathsf{MSE}^\star_{\lambda_1} \geq \mathsf{MSE}^\star_{\lambda_2}$;*
- *If $G(v_r)$ is not monotonic, then the comparison of $\mathsf{MSE}^\star_{\lambda_1}$ and $\mathsf{MSE}^\star_{\lambda_2}$ is not definite.*

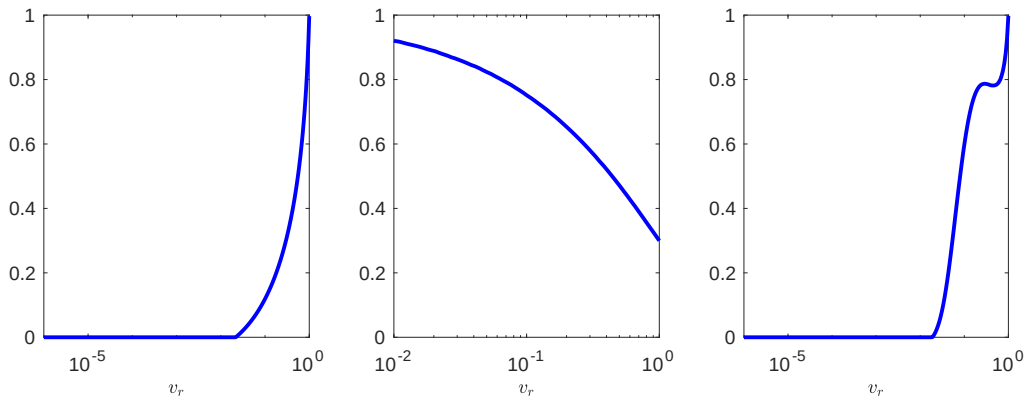

Figure 2: Plots of $G(v_r)$ for three choices of f.

The proof of this lemma can be found in the Supplementary Material. Lemma 6 shows that the impact of the spectrum on the final MSE performance of GLM-EP-app depends on the monotonicity of the function $G(v_r)$ (which further depends on f). For a given f, the function $G(v_r)$ can be numerically computed and its monotonicity can be easily checked. Below are three examples of f, corresponding to each of the cases discussed in Lemma 6.

Example 1: It can be shown that that $G(v_r)$ of the following f is is monotonically increasing:

$$f(z) = |z|.$$

See left panel of Figure 2. Hence, spiky spectrums are beneficial for MSE performance.

Example 2: The $G(v_r)$ of the following function is deacreasing:

$$f(z) = \mathrm{sign}(z).$$

See the middle panel of Figure 2. In this case, flatter spectrums are better.

Example 3: Consider the following function

$$f(z) = \begin{cases} |z|, & \text{if } |z| < 1 \\ |z| - 1, & \text{if } |z| \geq 1. \end{cases} \tag{15}$$

In this case, $G(v_r)$ is not monotonic (See right panel of Figure 2), and other features of the spectrum affect the performance of GLM-EP-app.

4.3 Impact of spectrum on perfect recovery threshold

We have shown that the impact of the spikiness of the spectrum on the MSE performance is related to the monotonicity of the function $G(v_r)$ which depends on the nonlinear function f. In this section, we will show that if our goal is to *minimize the number of measurements required for perfect reconstruction*, then more spiky spectrum benefit GMM-EP-app *for all* f. Furthermore, the information theoretic lower bound for δ can be reached (as close as we wish) if the spectrum of $\boldsymbol{A}$ is spiky enough. Theorem 3, whose proof can be found in the Supplementary Material, summarizes the above discussions.

Theorem 3. *For a given nonlinearity f and eigenvalue distribution P_λ, let $\delta_\lambda^{\mathsf{p}}$ be the minimum δ required for perfectly recovering the signal, i.e.,*

$$\delta_\lambda^{\mathsf{p}} \triangleq \inf\left\{\delta \,:\, \mathsf{MSE}_\lambda^\star = 0\right\}, \tag{16}$$

where $\mathsf{MSE}_\lambda^\star$ is defined in Lemma 3. Let λ_1 and λ_2 represent two limiting eigenvalue distributions and $\delta_{\lambda_1}^{\mathsf{p}}$ and $\delta_{\lambda_2}^{\mathsf{p}}$ the corresponding thresholds for perfect reconstruction. We have

(i) If $\lambda_1 \succeq_L \lambda_2$, then $\delta_{\lambda_1}^{\mathsf{p}} \leq \delta_{\lambda_2}^{\mathsf{p}}$;

(ii) For any f satisfying $\mathsf{mmse}_z(1) < 1$, there exists a distribution of λ for which $\delta_\lambda^{\mathsf{p}}$ is arbitrarily close to $\delta_{\text{opt}}^{\mathsf{p}} \triangleq 1/d(f(Z))$, where $Z \sim \mathcal{N}(0,1)$.

4.4 Noise Sensitivity Analysis

Up to now, we only studied the performance of GLM-EP-app in the noiseless setting. In practice, it is also important to guarantee that the reconstruction performance does not significantly worsen due to the presence of a small amount of measurement noise. We consider the noisy model in (4). GLM-EP-app remains unchanged except that η_z is replaced by a posterior mean estimator that takes the noise effect into consideration.

The following lemme analyzes the MSE performance of GLM-EP-app in the high SNR regime, and shows that its reconstruction is stable. The proof of Lemma 7 and other details about GLM-EP-app in the noisy setting are provided in the Supplementary Material.

Lemma 7. *Assume $d(f(Z)) \neq 0$. Let $\delta > \delta_\lambda^{\mathsf{p}}$, where $\delta_\lambda^{\mathsf{p}}$ is defined in Theorem 3. Let $\mathsf{MSE}_\lambda^\star(\sigma_w^2) \triangleq \lim_{t\to\infty} \mathsf{MSE}_\lambda(V_l^t)$ be the MSE in the noisy setting. As $\sigma_w^2 \to 0$, we have*

$$\mathsf{MSE}_\lambda^\star(\sigma_w^2) = C(\delta, f)\mathbb{E}\left[\lambda^{-1}\right]\sigma_w^2 \cdot (1 + o(1)),$$

where $0 < C(\delta, f) < \infty$ is a constant depending only on δ and f.

This lemma confirms that as long as $\delta > \delta_\lambda^{\mathsf{p}}$, GLM-EP-app can offer stable recovery. However, the minimum mean square error in this case depends on another feature of the spectrum, namely $\mathbb{E}\left[\lambda^{-1}\right]$. The optimal sensing mechanism should be designed by considering both features based on the expected noise level in the system.

5 Conclusion

We studied the impact of the spectrum of the sensing matrix on the performance of the expectation propagation (EP) algorithm in recovering signals from the nonlinear model $\boldsymbol{y} = f(\boldsymbol{A}\boldsymbol{x})$. We defined a notion of spikiness of the distributions and showed that depending on $f(\cdot)$, the spikiness of the distribution can help or hurt the performance of EP. We also showed that spiky sensing matrices can always reduce the number of observations required for the exact recovery of $\boldsymbol{x}$ from $\boldsymbol{y}$.

Our results can serve as the first step toward the optimal design of sensing matrices. However, there are several directions that require further investigation before one can apply our results to real-world applications: (i) Because the structure of the signal is often used in recovery algorithms, the role of the structure should be studied more carefully when we deal with spiky sensing matrices. (ii) While we discussed the high-signal-to-noise ratio regime in the paper, some applications have low signal-to-noise ratios. The impact of the spectrum of the sensing matrix in such cases requires more careful considerations.

References

[1] Barry C Arnold and José María Sarabia. *Majorization and the Lorenz order with applications in applied mathematics and economics*. Springer, 2018.

[2] Benjamin Aubin, Antoine Maillard, Jean Barbier, Florent Krzakala, Nicolas Macris, and Lenka Zdeborová. The committee machine: Computational to statistical gaps in learning a two-layers neural network. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124023, 2019.

[3] Sohail Bahmani and Justin Romberg. Phase retrieval meets statistical learning theory: A flexible convex relaxation. In *Artificial Intelligence and Statistics*, pages 252–260, 2017.

[4] Milad Bakhshizadeh, Arian Maleki, and Shirin Jalali. Using black-box compression algorithms for phase retrieval. *IEEE Transactions on Information Theory*, 66(12):7978–8001, 2020. doi: 10.1109/TIT.2020.3016183.

[5] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

[6] M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, Feb. 2011.

[7] M. Bayati and A. Montanari. The LASSO risk for Gaussian matrices. *IEEE Transactions on Information Theory*, 58(4):1997–2017, April 2012.

[8] Ali Bereyhi, Saba Asaad, Ralf R Müller, and Symeon Chatzinotas. RLS precoding for massive MIMO systems with nonlinear front-end. *arXiv preprint arXiv:1905.05227*, 2019.

[9] Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pages 16–21. IEEE, 2008.

[10] Zhiqi Bu, Jason M Klusowski, Cynthia Rush, and Weijie J Su. Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. *IEEE Transactions on Information Theory*, 67(1): 506–537, 2020.

[11] E. J. Candes, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015.

[12] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, Nov. 2013.

[13] Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Conference on Learning Theory*, pages 1078–1141. PMLR, 2020.

[14] Yuxin Chen and E. J. Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Communications on Pure and Applied Mathematics*, 70:822–883, May 2017.

[15] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[16] David L Donoho, Arian Maleki, and Andrea Montanari. The noise-sensitivity phase transition in compressed sensing. *IEEE Transactions on Information Theory*, 57(10):6920–6941, 2011.

[17] Rishabh Dudeja, Milad Bakhshizadeh, Junjie Ma, and Arian Maleki. Analysis of spectral methods for phase retrieval with random orthogonal matrices. *IEEE Transactions on Information Theory*, 66(8):5182–5203, 2020.

[18] Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *arXiv preprint arXiv:2008.11892*, 2020.

[19] Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J Samworth. A unifying tutorial on approximate message passing. *arXiv preprint arXiv:2105.02180*, 2021.

[20] Alyson Fletcher, Mojtaba Sahraee-Ardakan, Sundeep Rangan, and Philip Schniter. Expectation consistent approximate inference: Generalizations and convergence. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 190–194. IEEE, 2016.

[21] Alyson K Fletcher and Sundeep Rangan. Inference in deep networks in high dimensions. *arXiv preprint arXiv:1706.06549*, 2017.

[22] Bing Gao, Xinwei Sun, Yang Wang, and Zhiqiang Xu. Perturbed amplitude flow for phase retrieval. *IEEE Transaction on Signal Processing*, 68:5427–5440, 2020.

[23] Martin Genzel and Peter Jung. Recovering structured data from superimposed non-linear measurements. *IEEE Transactions on Information Theory*, 66(1):453–477, 2020. doi: 10.1109/TIT.2019.2932426.

[24] T. Goldstein and C. Studer. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, 64(4):2675–2689, April 2018.

[25] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. 51(4):1261–1282, 2005.

[26] Hengtao He, Chao-Kai Wen, and Shi Jin. Generalized expectation consistent signal recovery for nonlinear measurements. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pages 2333–2337. IEEE, 2017.

[27] Yue M Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, 9(3):507–541, 2020.

[28] Junjie Ma and Li Ping. Orthogonal AMP. *IEEE Access*, 5:2020–2033, 2017.

[29] Junjie Ma, Xiaojun Yuan, and Li Ping. On the performance of turbo signal recovery with partial dft sensing matrices. *IEEE Signal Processing Letters*, 22(10):1580–1584, 2015.

[30] Junjie Ma, Ji Xu, and Arian Maleki. Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization. *IEEE Transactions on Information Theory*, 65(6), June 2019.

[31] Junjie Ma, Rishabh Dudeja, Ji Xu, Arian Maleki, and Xiaodong Wang. Spectral method for phase retrieval: an expectation propagation perspective. *IEEE Transactions on Information Theory*, 67(2):1332–1355, 2021.

[32] Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Phase retrieval in high dimensions: Statistical and computational phase transitions. *Advances in Neural Information Processing Systems*, 33, 2020.

[33] Thomas P Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[34] Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. *arXiv preprint arXiv:1708.05932*, 2017.

[35] Manfred Opper and Ole Winther. Expectation consistent approximate inference. *Journal of Machine Learning Research*, 6(Dec):2177–2204, 2005.

[36] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas. Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization. *IEEE Transactions on Signal Processing*, 51(9): 2381–2401, Sep. 2003.

[37] Yaniv Plan and Roman Vershynin. The generalized LASSO with non-linear observations. *IEEE Transactions on Information Theory*, 62(3):1528–1537, 2016.

[38] S. Rangan. Generalized approximate message passing for estimation with random linear mixing. In *IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172, July 2011.

[39] S. Rangan, P. Schniter, and A. K. Fletcher. Vector approximate message passing. *IEEE Transactions on Information Theory*, 65(10):6664–6684, Oct 2019.

[40] Alfréd Rényi. On the dimension and entropy of probability distributions. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(1-2):193–215, 1959.

[41] E. Riegler and G. Tauböck. Almost lossless analog compression without phase information. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 999–1003, June 2015.

[42] Cynthia Rush and Ramji Venkataramanan. Finite sample analysis of approximate message passing algorithms. *IEEE Transactions on Information Theory*, 64(11):7264–7286, 2018.

[43] Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

[44] Takashi Shinzato and Yoshiyuki Kabashima. Perceptron capacity revisited: classification ability for correlated patterns. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324013, 2008.

[45] Pragya Sur and Emmanuel J Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525, 2019.

[46] K. Takeuchi. Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. *IEEE Transactions on Information Theory*, pages 1–1, 2019.

[47] Keigo Takeuchi. A unified framework of state evolution for message-passing algorithms. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 151–155. IEEE, 2019.

[48] M. Vehkaperä, Y. Kabashima, and S. Chatterjee. Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing. *IEEE Transactions on Information Theory*, 62(4):2100–2124, April 2016.

[49] G. Wang, G. B. Giannakis, and Y. C. Eldar. Solving systems of random quadratic equations via truncated amplitude flow. *IEEE Transactions on Information Theory*, 64(2):773–794, Feb 2018.

[50] Shuaiwen Wang, Haolei Weng, Arian Maleki, et al. Which bridge estimator is the best for variable selection? *Annals of Statistics*, 48(5):2791–2823, 2020.

[51] Y-C Wang and Z-Q Luo. Optimized iterative clipping and filtering for PAPR reduction of OFDM signals. *IEEE Transactions on Communications*, 59(1):33–37, 2010.

[52] Haolei Weng, Arian Maleki, Le Zheng, et al. Overcoming the limitations of phase transition by higher order analysis of regularization techniques. *Annals of Statistics*, 46(6A):3099–3129, 2018.

[53] Y. Wu and S. Verdú. Réenyi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Transactions on Information Theory*, 56(8):3721–3748, Aug 2010.

[54] Y. Wu and S. Verdú. Optimal phase transitions in compressed sensing. *IEEE Transactions on Information Theory*, 58(10):6241–6263, Oct 2012.

[55] Huishuai Zhang and Yingbin Liang. Reshaped wirtinger flow for solving quadratic system of equations. In *Advances in Neural Information Processing Systems*, pages 2622–2630, 2016.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes] To be included in Supplementary Material
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] Plan to release the URL to the code in a future arXiv version of the paper.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]