

# Fundamental Limits of PhaseMax for Phase Retrieval: A Replica Analysis

Oussama Dhifallah and Yue M. Lu

John A. Paulson School of Engineering and Applied Sciences

Harvard University, Cambridge, MA 02138, USA

Email: oussama\_dhifallah@g.harvard.edu, yuelu@seas.harvard.edu

**Abstract**—We consider a recently proposed convex formulation, known as the PhaseMax method, for solving the phase retrieval problem. Using the replica method from statistical mechanics, we analyze the performance of PhaseMax in the high-dimensional limit. Our analysis predicts the *exact* asymptotic performance of PhaseMax. In particular, we show that a sharp phase transition phenomenon takes place, with a simple analytical formula characterizing the phase transition boundary. This result shows that the oversampling ratio required by existing performance bounds in the literature can be further reduced. Numerical results confirm the validity of our replica analysis, showing that the theoretical predictions are in excellent agreement with the actual performance of the algorithm, even for moderate signal dimensions.

## I. INTRODUCTION

Let  $\xi \in \mathbb{R}^n$  be an unknown signal, and  $\{\mathbf{a}_i\}_{1 \leq i \leq m}$  be a collection of sensing vectors. Given the measurements

$$y_i = \left| \mathbf{a}_i^T \xi \right|, \quad (1)$$

we are interested in reconstructing  $\xi$  up to a global sign change. This is the real-valued version of the classical phase retrieval problem [1], [2], which has attracted much renewed interests in the signal processing community in recent years (see, *e.g.*, [3]–[8]). The main challenge of the phase retrieval problem comes from the nonconvex nature of the constraints (1). Recently, a simple yet very effective convex relaxation was independently proposed by two groups of authors [9], [10]. Following [10], we shall refer to it as the PhaseMax method, which seeks to estimate  $\xi$  via a linear programming problem:

$$\begin{aligned} \hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \quad & \mathbf{x}_{\text{init}}^T \mathbf{x} \\ \text{s.t.} \quad & \left| \mathbf{a}_i^T \mathbf{x} \right| \leq y_i, \quad \text{for } 1 \leq i \leq m. \end{aligned} \quad (2)$$

Here, the nonconvex equality constraints in (1) have been relaxed to convex inequality constraints. The vector  $\mathbf{x}_{\text{init}}$  is an initial guess of  $\xi$ . In practice,  $\mathbf{x}_{\text{init}}$  can be obtained if we have additional prior knowledge about  $\xi$  (*e.g.*, nonnegativity) or by using a simple spectral method [6], [7], [11].

The performance of the PhaseMax method has been investigated in [9], [10] (see also [12]), where the authors provide *sufficient conditions* for PhaseMax to successfully recover the target vector  $\xi$ . In this paper, we present an *exact performance analysis* of the method in the high-dimensional ( $n \rightarrow \infty$ ) limit. In particular, we show that a sharp phase transition phenomenon takes place, with a simple analytical formula characterizing the phase transition boundary.

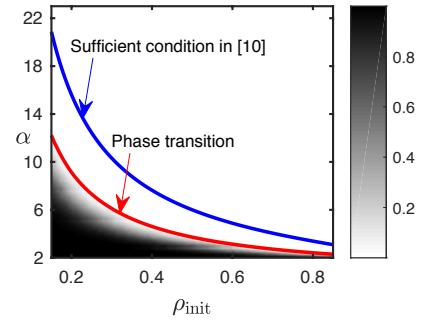


Fig. 1. The normalized mean squared error (NMSE) of the PhaseMax method: theory versus simulations. The signal dimension is set to  $n = 1000$ , and the results are averaged over 50 independent trials. The red curve shows the predicted phase transition boundary given in (4). Above the threshold, PhaseMax recovers the signal perfectly; below the threshold,  $\text{NMSE} > 0$ . The blue curve shows the sufficient condition (5) derived in [10].

We shall quantify the performance of PhaseMax in terms of the normalized mean squared error (NMSE), defined as  $\text{NMSE}_n \stackrel{\text{def}}{=} \min \{ \|\xi - \hat{\mathbf{x}}\|_2^2, \|\xi + \hat{\mathbf{x}}\|_2^2 \} / \|\xi\|_2^2$ . The NMSE depends on two parameters: the oversampling ratio  $\alpha \stackrel{\text{def}}{=} m/n$ , and the quality of the initial guess  $\mathbf{x}_{\text{init}}$ , measured via the input cosine similarity

$$\rho_{\text{init}} \stackrel{\text{def}}{=} \frac{|\mathbf{x}_{\text{init}}^T \xi|}{\|\mathbf{x}_{\text{init}}\|_2 \|\xi\|_2}. \quad (3)$$

Taking values between 0 and 1, the parameter  $\rho_{\text{init}}$  assesses the degree of alignment between the true signal vector  $\xi$  and the initial guess  $\mathbf{x}_{\text{init}}$ .

As the main contribution of our work, we derive the following asymptotically exact characterization of PhaseMax, under the assumption that the sensing vectors are drawn from the normal distribution:

$$\text{NMSE}_n \xrightarrow{n \rightarrow \infty} \begin{cases} 0, & \text{if } \frac{\pi/\alpha}{\tan(\pi/\alpha)} > 1 - \rho_{\text{init}}^2; \\ s(\rho_{\text{init}}, \alpha), & \text{otherwise,} \end{cases} \quad (4)$$

where  $s(\rho_{\text{init}}, \alpha)$  is a positive function that can be computed by solving a fixed point equation (see (13), (14) and (16) in Section II-C.) The above expression characterizes the fundamental limits of PhaseMax: for any fixed input cosine similarity  $\rho_{\text{init}}$ , there is a critical threshold  $\alpha_c(\rho_{\text{init}})$  such that PhaseMax perfectly recovers  $\xi$  if  $\alpha > \alpha_c(\rho_{\text{init}})$ , and that it fails to recover  $\xi$  if  $\alpha < \alpha_c(\rho_{\text{init}})$ .

Figure 1 illustrates our asymptotic characterization and compares it with results from numerical simulations. Specifically,

the red curve in the figure shows the phase transition boundary  $\alpha_c(\rho_{\text{init}})$  as a function of the input cosine similarity  $\rho_{\text{init}}$ , which can be seen to have excellent agreement with the actual performance of the algorithm. In [10], the authors show that PhaseMax is successful with high probability if

$$\alpha > \frac{2\pi}{\pi - 2 \arccos(\rho_{\text{init}})}. \quad (5)$$

This sufficient condition is plotted as the blue curve in Figure 1. We note that our theoretical prediction significantly reduces the required oversampling ratio as given in (5) for any considered quality of the initial guess vector.

Our analysis is based on the powerful replica method [13] from statistical mechanics. Although certain key steps of the replica method have not yet been mathematically proven, the method has been successful in the analysis of a wide-range of high-dimensional inference problems in signal and information processing (see, *e.g.*, [14]–[16]). Some of its sharp predictions have later been proven through alternative mathematical approaches (*e.g.*, [17]). In this work, we use the replica method to derive our asymptotic predictions, and corroborate these analytical results—rigorously speaking, conjectures—via numerical simulations.

The rest of this paper is organized as follows. After precisely laying out the various technical assumptions, we present the main results of this work in Section II. Additional numerical results are provided in Section III to validate our theoretical predictions. Section IV concludes the paper. For readers interested in our replica calculations, we present some of our key derivations in the appendix, and leave the full technical details to a follow-up paper.

## II. MAIN RESULTS

### A. Technical Assumptions

In what follows, we first state the assumptions under which we derive our analytical predictions.

- (A.1) The sensing vectors  $\{\mathbf{a}_i\}_{1 \leq i \leq m}$  are independent random vectors whose entries are *i.i.d.* standard normal random variables.
- (A.2) The number of measurements  $m = m(n)$  with  $\alpha_n = m(n)/n \rightarrow \alpha > 0$  as  $n \rightarrow \infty$ .
- (A.3) Both the target vector  $\xi$  and the initial guess  $\mathbf{x}_{\text{init}}$  are independent from the sensing vectors.
- (A.4) The initial guess vector  $\mathbf{x}_{\text{init}}$  has a positive correlation with the target signal vector  $\xi$ , *i.e.*  $\mathbf{x}_{\text{init}}^T \xi \geq 0$ .
- (A.5)  $\|\xi\|_2 = \|\mathbf{x}_{\text{init}}\|_2 = \sqrt{n}$ .

Note that the last two assumptions can be made without loss of generality, since  $\xi$  and  $-\xi$  are both valid targets and thanks to the scale invariant nature of the convex optimization problem in (2), respectively.

### B. The Boltzmann Distribution

The first step of our replica analysis is to “soften” the optimization problem (2) via a probability distribution. To that end, we introduce the following function

$$\mathcal{H}(\mathbf{x}) \stackrel{\text{def}}{=} -\mathbf{x}_{\text{init}}^T \mathbf{x} - \sum_{i=1}^m \log [U(|\mathbf{a}_i^T \xi| - |\mathbf{a}_i^T \mathbf{x}|)], \quad (6)$$

where  $U(x)$  represents the unit-step function, *i.e.*,  $U(x) = 1$  if  $x \geq 0$  and  $U(x) = 0$  otherwise. Clearly, the convex optimization problem (2) is equivalent to minimizing the function  $\mathcal{H}$  over the variable  $\mathbf{x}$ . Now consider the following probability distribution

$$\begin{aligned} \mathbb{P}_\beta(\mathbf{x}) &= \frac{1}{Z_n(\beta)} \exp(-\beta \mathcal{H}(\mathbf{x})) \\ &= \frac{1}{Z_n(\beta)} \exp(\beta \mathbf{x}_{\text{init}}^T \mathbf{x}) \prod_{i=1}^m U(|\mathbf{a}_i^T \xi| - |\mathbf{a}_i^T \mathbf{x}|), \end{aligned} \quad (7)$$

where  $\beta > 0$  is a fixed parameter and  $Z_n(\beta)$  denotes a normalizing constant. We note that the probability distribution  $\mathbb{P}_\beta(\mathbf{x})$  (often referred to as the *Boltzmann distribution* in the literature) can be interpreted as a softened version of (2). In particular, when  $\beta = 0$ , the distribution is uniform over all  $\mathbf{x}$  satisfying the constraints  $|\mathbf{a}_i^T \mathbf{x}| \leq y_i$  for  $1 \leq i \leq m$ . As we increase  $\beta$ , however, the distribution will become more concentrated around the optimal solution  $\hat{\mathbf{x}}$  of (2). In fact, if  $\hat{\mathbf{x}}$  is the unique solution of (2), then the distribution  $\mathbb{P}_\beta(\mathbf{x})$  will converge to a singular distribution  $\delta(\mathbf{x} - \hat{\mathbf{x}})$  as  $\beta \rightarrow \infty$ . Therefore, to characterize the performance of the PhaseMax method, one only needs to examine the behavior of the Boltzmann distribution  $\mathbb{P}_\beta(\mathbf{x})$  for sufficiently large values of  $\beta$ .

Of particular interest to us are the following two asymptotic moments of  $\mathbb{P}_\beta(\mathbf{x})$ :

$$\vartheta_\beta = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}\{\mathbf{x}^T \xi\} \quad (8)$$

$$q_\beta = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}\{\|\mathbf{x}\|_2^2\} \quad (9)$$

where  $\mathbb{E}\{\cdot\}$  denotes the expectation over the Boltzmann distribution and over the sensing vectors  $\{\mathbf{a}_i\}_i$ . Moreover, define

$$\vartheta^* = \lim_{\beta \rightarrow \infty} \vartheta_\beta \quad \text{and} \quad q^* = \lim_{\beta \rightarrow \infty} q_\beta. \quad (10)$$

Since the Boltzmann distribution  $\mathbb{P}_\beta(\mathbf{x})$  will be concentrated on the global optimal solution  $\hat{\mathbf{x}}$  of (2) as  $\beta \rightarrow \infty$ , the value of  $\vartheta^*$  reveals the normalized inner product between the optimal solution  $\hat{\mathbf{x}}$  and the true signal vector  $\xi$ , *i.e.*,  $\vartheta^* = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}\{\hat{\mathbf{x}}^T \xi\}$ . Similarly, the value of  $q^*$  is equal to the normalized squared norm of  $\hat{\mathbf{x}}$ , *i.e.*,  $q^* = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}\{\|\hat{\mathbf{x}}\|_2^2\}$ . It follows that the asymptotic NMSE can be computed as

$$\begin{aligned} \text{NMSE} &= \lim_{n \rightarrow +\infty} \frac{\mathbb{E} \min\{\|\xi - \hat{\mathbf{x}}\|_2^2, \|\xi + \hat{\mathbf{x}}\|_2^2\}}{\|\xi\|_2^2} \\ &= q^* - 2|\vartheta^*| + 1, \end{aligned} \quad (11)$$

where in reaching (11) we have used the assumption that  $\|\xi\|_2 = \sqrt{n}$ . Thus, the task of analyzing the asymptotic performance of PhaseMax boils down to calculating the values of  $q^*$  and  $\vartheta^*$ , which we do next by using the replica method.

### C. Asymptotic Predictions via the Replica Method

Much of the information about the Boltzmann distribution  $\mathbb{P}_\beta(\mathbf{x})$ , in particular, the moments as in (8) and (9), can be obtained from its log-partition function, defined as

$$f(\beta) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \frac{-1}{\beta n} \mathbb{E}\{\log Z_n(\beta)\}.$$

The challenge here is to compute the partition function  $Z_n(\beta)$ , which involves a high-dimensional integration. And this is where the replica method [13] comes in. Using this method, we can calculate  $f(\beta)$  for all  $\beta > 0$ . In particular, its limit as  $\beta \rightarrow \infty$  can be derived as

$$f(\infty) = \text{extr}_{\vartheta, q, \chi} \left[ \frac{\alpha}{2\pi\chi} \left\{ -2\sqrt{q - \vartheta^2} + (1 + q + 2\vartheta)A_1 + (1 + q - 2\vartheta)A_2 - 4\vartheta A_3 \right\} - \frac{1}{2}\hat{Q}q + \frac{1}{2}\hat{\chi}\chi + \hat{\vartheta}\vartheta - \frac{\hat{\vartheta}^2 + 1 + 2\hat{\vartheta}\rho_{\text{init}}}{2\hat{Q}} - \frac{\hat{\chi}}{2\hat{Q}} \right], \quad (12)$$

where  $A_1 = \arctan\left(\frac{q+\vartheta}{\sqrt{q-\vartheta^2}}\right)$ ,  $A_2 = \arctan\left(\frac{q-\vartheta}{\sqrt{q-\vartheta^2}}\right)$  and  $A_3 = \arctan\left(\frac{\vartheta}{\sqrt{q-\vartheta^2}}\right)$  are three functions, and  $\text{extr}_{\mathcal{X}}\{g(X)\}$  denotes the extremization of a function  $g$  over a set of variables  $\mathcal{X}$ . Here,  $\mathcal{X}$  includes six scalar variables  $\vartheta$ ,  $q$ ,  $\chi$ ,  $\hat{\vartheta}$ ,  $\hat{q}$  and  $\hat{\chi}$ . To streamline our presentation, we postpone the details of our replica calculations leading to (12) to the appendix.

To solve the extremization problem in (12), we set the gradient with respect to the variables to zero, which leads to a set of nonlinear saddle point equations. After some further simplifications, we can eliminate the variables  $\vartheta$ ,  $\chi$ ,  $\hat{\vartheta}$ ,  $\hat{q}$  and  $\hat{\chi}$  and just need to study a simple fixed-point equation:

$$g = h(q) \stackrel{\text{def}}{=} 1 + \frac{\pi}{2\alpha c} [w^2(q) - \rho_{\text{init}}^{-2}(\vartheta - w(q))^2 - 1], \quad (13)$$

where  $c = \tan(\pi/\alpha)/2$  is a constant,

$$\vartheta = \sqrt{q - c^2(1 - q)^2} \quad (14)$$

and

$$w(q) = 1 - \frac{2\alpha}{\pi} \arctan\left(\frac{c|1-q|}{1 + \sqrt{q - c^2(1 - q)^2}}\right). \quad (15)$$

The solution to the above equations then gives us the key parameters of interest  $\vartheta^*$  and  $q^*$  as defined in (10), from which we can compute the asymptotic NMSE by using (11).

We observe that  $q = 1$  is always a fixed point of (13). However, for any fixed  $\rho_{\text{init}}$  and when we reduce the oversampling ratio  $\alpha$  to below a threshold, a second fixed point emerges and the original solution  $q = 1$  becomes unstable. This is indeed the origin of the phase transition. To locate the phase transition boundary, we study the stability of the solution  $q = 1$ . Specifically, by the definition of  $h(q)$ , we can verify that  $\frac{dh}{dq} \Big|_{q=1} \equiv 1$  and

$$\frac{d^2h}{dq^2} \Big|_{q=1} = \frac{\alpha c}{2\pi} - \frac{(\pi/2 - \alpha c)^2}{2\pi\alpha c \rho_{\text{init}}^2} - 1/4.$$

Thus, the solution  $q = 1$  becomes unstable (*i.e.* a phase transition happens) when

$$\rho_{\text{init}}^2 < 1 - \frac{\pi/\alpha}{\tan(\pi/\alpha)},$$

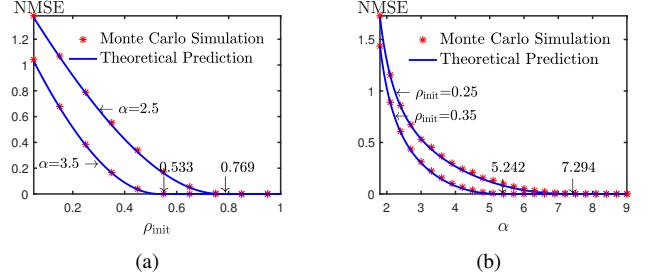


Fig. 2. Analytical predictions v.s. numerical simulations. (a) The NMSE as a function of the input cosine similarity, for two different values of the oversampling ratio; (b) The NMSE as a function of the oversampling ratio, for two different values of  $\rho_{\text{init}}$ . In both cases, the theoretical results obtained by the replica analysis can accurately predict the actual performance of the PhaseMax method.

which is exactly when  $\frac{d^2h}{dq^2} \Big|_{q=1}$  changes its sign. Finally, the function  $s(\rho_{\text{init}}, \alpha)$  in (4) can be obtained as

$$s(\rho_{\text{init}}, \alpha) = q^* - 2|\vartheta^*| + 1, \quad (16)$$

where  $q^*$  is the stable solution of (13) and  $\vartheta^*$  is given by (14).

### III. NUMERICAL RESULTS

In this section, we present additional numerical results to verify our analytical predictions found through the replica method. In all of our experiments, we solve the convex optimization problem (2) using the approach presented in [18] where the signal dimension is set to  $n = 1000$ . The results are also averaged over 50 independent Monte Carlo trials.

Our first simulation example, shown in Figure 2(a), studies the performance of our analytical prediction of the NMSE [see (16)] as a function of the input cosine similarity  $\rho_{\text{init}}$  for two different values of the oversampling ratio:  $\alpha = 2.5$  and  $\alpha = 3.5$ , respectively. As seen from the figure, the theoretical prediction obtained by the replica method is in excellent agreement with the experimental results obtained by numerically solving the convex optimization problem (2). The results also validate our theoretical prediction of the phase transition points: the critical input cosine similarity corresponding to each value of  $\alpha$  is  $\rho_{\text{init}}(\alpha = 2.5) = 0.769$  and  $\rho_{\text{init}}(\alpha = 3.5) = 0.533$ , respectively.

A different example is shown in Figure 2(b), where we examine the performance of our analytical prediction of the NMSE as a function of the oversampling ratio  $\alpha$  for two different values of the input cosine similarity:  $\rho_{\text{init}} = 0.25$  and  $\rho_{\text{init}} = 0.35$ . Again, as seen from the figure, our theoretical results can accurately predict the actual performance of the algorithm.

### IV. CONCLUSION

We presented in this paper an exact characterization of the performance of the PhaseMax method for phase retrieval. Our replica analysis leads to an analytical formula for the asymptotic NMSE of the estimate given by PhaseMax in the high-dimensional limit. It also reveals a sharp phase transition phenomenon: for PhaseMax to succeed, the oversampling ratio must be above a critical threshold, given as a function of the input cosine similarity. Simulation results confirm the validity

of our theoretical predictions. They also show that our theoretical results significantly reduce the required oversampling ratio given by an existing sufficient condition in the literature.

## APPENDIX

This appendix provides a sketch of our derivations leading to (12). To start, we write the partition function  $Z_n(\beta)$  as

$$Z_n(\beta) = \int \exp \left( \beta \mathbf{x}_{\text{init}}^T \mathbf{x} \right) \prod_{i=1}^m U \left\{ |\mathbf{a}_i^T \boldsymbol{\xi}| - |\mathbf{a}_i^T \mathbf{x}| \right\} d\mathbf{x}.$$

Using the replica trick [13], the *free energy density* for any given parameter  $\beta$  can be expressed as follows

$$\begin{aligned} f(\beta) &= \lim_{n \rightarrow \infty} \frac{-1}{\beta n} \mathbb{E} \left\{ \log(Z_n(\beta)) \right\} \\ &= \lim_{n \rightarrow \infty} \lim_{\ell \rightarrow 0} \frac{-1}{\beta \ell n} \log \left\{ \mathbb{E}(Z_n^\ell(\beta)) \right\}, \end{aligned} \quad (17)$$

where the expectation is over the random measurement vectors  $\{\mathbf{a}_i, 1 \leq i \leq m\}$ . To determine the expression of the free energy density, we first need to compute an analytical expression of  $\mathbb{E}(Z_n^\ell(\beta))$ . To this end, we write the expression of  $\mathbb{E}(Z_n^\ell(\beta))$  as follows

$$\begin{aligned} \mathbb{E}(Z_n^\ell(\beta)) &= \int \exp \left( \sum_{a=1}^{\ell} \beta \mathbf{x}_{\text{init}}^T \mathbf{x}_a \right) \times \\ &\left[ \mathbb{E} \left\{ \prod_{a=1}^{\ell} U \left\{ \frac{|\mathbf{a}^T \boldsymbol{\xi}|}{\sqrt{n}} - \frac{|\mathbf{a}^T \mathbf{x}_a|}{\sqrt{n}} \right\} \right\} \right]^m \prod_{a=1}^{\ell} d\mathbf{x}_a, \end{aligned} \quad (18)$$

where  $\mathbf{x}_a$  denotes the  $a$ th replica signal and where the expectation is over the random vector  $\mathbf{a}$  which is normally distributed with zero mean and covariance matrix  $\mathbf{I}_n$ .

Define the following variables  $\vartheta_a = n^{-1} \boldsymbol{\xi}^T \mathbf{x}_a$ ,  $r_a = n^{-1} \mathbf{x}_a^T \mathbf{x}_a$  and  $q_{ab} = n^{-1} \mathbf{x}_a^T \mathbf{x}_b$ , for all  $a, b \in \{1, \dots, \ell\}$ . The expression of  $\mathbb{E}(Z_n^\ell(\beta))$  given in (18) can be rewritten as

$$\begin{aligned} \mathbb{E}(Z_n^\ell(\beta)) &= \int \exp \left( n \left\{ \alpha \log [\mathcal{G}(\{\vartheta_a, r_a, q_{ab}\})] \right. \right. \\ &\quad \left. \left. + \mathcal{I}(\{\vartheta_a, r_a, q_{ab}\}) \right\} \right) \prod_{a=1}^{\ell} d\{r_a\} d\{\vartheta_a\} \prod_{a < b} d\{q_{ab}\}, \end{aligned} \quad (19)$$

where  $\alpha = \frac{m}{n}$  denotes the oversampling ratio and where the function  $\mathcal{G}$  can be expressed as follows

$$\mathcal{G}(\{\vartheta_a, r_a, q_{ab}\}) = \mathbb{E} \left\{ \prod_{a=1}^{\ell} U \left\{ |u_0| - |u_a| \right\} \right\}, \quad (20)$$

with  $u_0 = \mathbf{a}^T \boldsymbol{\xi} / \sqrt{n}$  and  $u_a = \mathbf{a}^T \mathbf{x}_a / \sqrt{n}$ , for all  $a$ . Furthermore, using a result in large deviation theory known as the Gartner-Ellis theorem [19], the rate function  $\mathcal{I}$  can be expressed as the Fenchel-Legendre transform of a cumulant generating

function. Specifically, the rate function  $\mathcal{I}$  can be expressed as follows

$$\begin{aligned} \mathcal{I}(\{\vartheta_a, r_a, q_{ab}\}) &= \underset{\tilde{\vartheta}_a, \tilde{r}_a, \tilde{q}_{ab}}{\text{extr}} \left( - \sum_{a=1}^{\ell} \tilde{\vartheta}_a \vartheta_a - \sum_{a=1}^{\ell} \tilde{r}_a r_a \right. \\ &\quad \left. - \sum_{1 \leq a < b \leq \ell} \tilde{q}_{ab} q_{ab} + \lambda(\{\tilde{\vartheta}_a, \tilde{r}_a, \tilde{q}_{ab}\}) \right), \end{aligned} \quad (21)$$

where the function  $\lambda$  represents the cumulant generating function and is given by

$$\begin{aligned} \lambda(\{\tilde{\vartheta}_a, \tilde{r}_a, \tilde{q}_{ab}\}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left[ \int \exp \left( \sum_{a=1}^{\ell} \tilde{\vartheta}_a \boldsymbol{\xi}^T \mathbf{x}_a \right. \right. \\ &\quad \left. \left. + \sum_{a=1}^{\ell} \tilde{r}_a \mathbf{x}_a^T \mathbf{x}_a + \sum_{a < b} \tilde{q}_{ab} \mathbf{x}_a^T \mathbf{x}_b + \sum_{a=1}^{\ell} \beta \mathbf{x}_{\text{init}}^T \mathbf{x}_a \right) \prod_{a=1}^{\ell} d\mathbf{x}_a \right]. \end{aligned}$$

We use the replica symmetry (RS) ansatz where it is assumed that, when the dimension  $n$  is sufficiently large, the integration in (19) is dominated by the configurations satisfying the following particular property:  $\vartheta_a = \vartheta$ ,  $\tilde{\vartheta}_a = \tilde{\vartheta}$ ,  $r_a = r$ ,  $\tilde{r}_a = \tilde{r}$ ,  $q_{ab} = q$  and  $\tilde{q}_{ab} = \tilde{q}$ , for all  $a, b \in \{1, 2, \dots, \ell\}$ . This leads to the following representation of the random variables  $u_0$  and  $\{u_a, 1 \leq a \leq \ell\}$ :

$$\begin{aligned} u_0 &= \sqrt{1 - \vartheta^2/q} s_0 + \sqrt{\vartheta^2/q} t \\ u_a &= \sqrt{r - q} s_a + \sqrt{q} t \end{aligned} \quad (22)$$

where  $s_0$ ,  $t$  and  $\{s_a, 1 \leq a \leq n\}$  are i.i.d. Gaussian random variables with zero mean and unit variance. Using the introduced representation of the random variables  $u_0$  and  $u_a$ , the free energy density can be rewritten as follows

$$\begin{aligned} f(\beta) &= \underset{\vartheta, q, r}{\text{extr}} \left\{ - \frac{1}{2\beta} \tilde{q} q + \frac{1}{\beta} \tilde{r} r + \frac{1}{\beta} \tilde{\vartheta} \vartheta + \frac{1}{2\beta} \log(\tilde{q} - 2\tilde{r}) \right. \\ &\quad \left. - \frac{\alpha}{\beta} \mathbb{E}_{s_0, t} \left\{ \log \left[ \Phi \left( \frac{|u_0| - \sqrt{q} t}{\sqrt{r - q}} \right) - \Phi \left( \frac{-|u_0| - \sqrt{q} t}{\sqrt{r - q}} \right) \right] \right\} \right. \\ &\quad \left. - \frac{\tilde{\vartheta}^2 + \beta^2 + 2\tilde{\vartheta}\beta\rho_{\text{init}}}{2\beta(\tilde{q} - 2\tilde{r})} - \frac{\tilde{q}}{2\beta(\tilde{q} - 2\tilde{r})} \right\}, \end{aligned} \quad (23)$$

where the expectation is over the random variables  $s_0$  and  $t$  and where  $\Phi$  denotes the cumulative distribution function of the standard normal distribution. Recall that our objective is to study the behavior of the free energy density as the parameter  $\beta$  tends to infinity. When  $\beta$  goes to  $+\infty$ , the only non-trivial solution of the extremization problem given in (23) occurs when  $\chi = \beta(r - q)$  is finite. Assume that  $\hat{\chi} = \beta^{-2}\tilde{q}$ ,  $\hat{\vartheta} = \beta^{-1}\tilde{\vartheta}$  and  $\hat{Q} = \beta^{-1}(\tilde{q} - 2\tilde{r})$ . Under the RS assumption and using the Laplace method, the free energy density, as  $\beta$  tends to infinity, can then be expressed as in (12).

## REFERENCES

- [1] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, p. 237, 1972.
- [2] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

- [3] E. J. Candes, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.
- [4] K. Jaganathan, S. Oymak, and B. Hassibi, “Sparse phase retrieval: Convex algorithms and limitations,” in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1022–1026.
- [5] I. Waldspurger, A. d’Aspremont, and S. Mallat, “Phase recovery, maxcut and complex semidefinite programming,” *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.
- [6] P. Netrapalli, P. Jain, and S. Sanghavi, “Phase retrieval using alternating minimization,” in *Advances in Neural Information Processing Systems*, 2013, pp. 2796–2804.
- [7] E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *Information Theory, IEEE Transactions on*, vol. 61, no. 4, pp. 1985–2007, 2015.
- [8] G. Wang, G. B. Giannakis, and Y. C. Eldar, “Solving Systems of Random Quadratic Equations via Truncated Amplitude Flow,” *arXiv:1605.08285*, May 2016.
- [9] S. Bahmani and J. Romberg, “Phase Retrieval Meets Statistical Learning Theory: A Flexible Convex Relaxation,” *CoRR*, vol. abs/1610.04210, 2016. [Online]. Available: <http://arxiv.org/abs/1610.04210>
- [10] T. Goldstein and C. Studer, “PhaseMax: Convex Phase Retrieval via Basis Pursuit,” *CoRR*, vol. abs/1610.07531, 2016. [Online]. Available: <http://arxiv.org/abs/1610.07531>
- [11] Y. M. Lu and G. Li, “Phase transitions of spectral initialization for high-dimensional nonconvex estimation,” *arXiv:1702.06435 [cs.IT]*, 2017. [Online]. Available: <https://arxiv.org/abs/1702.06435>
- [12] P. Hand and V. Voroninski, “Corruption Robust Phase Retrieval via Linear Programming,” *arXiv:1612.03547 [cs, math]*, Dec. 2016.
- [13] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, ser. World Scientific Lecture Notes in Physics. World Scientific, Nov. 1986, vol. 9.
- [14] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, Nov 2002.
- [15] Y. Kabashima, T. Wadayama, and T. Tanaka, “A typical reconstruction limit for compressed sensing based on  $\ell_p$ -norm minimization,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 09, p. L09003, 2009.
- [16] S. Rangan, A. K. Fletcher, and V. K. Goyal, “Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing,” *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1902–1923, 2012.
- [17] M. Talagrand, *Mean Field Models for Spin Glasses*. Springer, 2010, vol. 1 and 2.
- [18] T. Goldstein, C. Studer, and R. Baraniuk, “A field guide to forward-backward splitting with a FASTA implementation,” *arXiv eprint*, vol. abs/1411.3406, 2014. [Online]. Available: <http://arxiv.org/abs/1411.3406>
- [19] H. Touchette, “The large deviation approach to statistical mechanics,” *Physics Reports*, vol. 478, no. 1, pp. 1 – 69, 2009.