

Phase Retrieval via Linear Programming: Fundamental Limits and Algorithmic Improvements

Oussama Dhifallah¹, Christos Thrampoulidis² and Yue M. Lu¹

¹Harvard University, Cambridge, MA 02138, USA

Email: oussama_dhifallah@g.harvard.edu, yuelu@seas.harvard.edu

²Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Email: cthrampo@mit.edu

Abstract—A recently proposed convex formulation of the phase retrieval problem estimates the unknown signal by solving a simple linear program. This new scheme, known as PhaseMax, is computationally efficient compared to standard convex relaxation methods based on lifting techniques. In this paper, we present an exact performance analysis of PhaseMax under Gaussian measurements in the large system limit. In contrast to previously known performance bounds in the literature, our results are asymptotically exact and they also reveal a sharp phase transition phenomenon. Furthermore, the geometrical insights gained from our analysis led us to a novel nonconvex formulation of the phase retrieval problem and an accompanying iterative algorithm based on successive linearization and maximization over a polytope. This new algorithm, which we call PhaseLamp, has provably superior recovery performance over the original PhaseMax method.

I. INTRODUCTION

A. Background

We consider the real-valued phase retrieval problem, which seeks to recover an unknown signal vector $\xi \in \mathbb{R}^n$ from m magnitude measurements $\{y_i, 1 \leq i \leq m\}$ of the form:

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

where $\{\mathbf{a}_i \in \mathbb{R}^n, 1 \leq i \leq m\}$ is a set of (known) sensing vectors. In order to recover ξ , we must resolve the uncertainty due to the missing phase (or sign) information. This is a classical problem [1], [2], with many applications in applied physics and engineering. Over the past decade, it has attracted significant attention in the optimization and signal processing communities, with a particular effort towards establishing rigorous recovery guarantees for either already existing or newly proposed solution methods. See [3]–[8] and many references therein.

Among the most well-established methods are those based on semidefinite relaxation (e.g., [3], [9].) Such convex optimization methods operate by lifting the original n -dimensional natural parameter space to a higher dimensional matrix space. Unfortunately, the increase in the dimensionality introduces challenges in computational complexity and memory requirement for the resulting algorithms. Subsequent works [6]–[8], [10] suggest going around this issue by developing nonconvex formulations of the phase retrieval problem and solution algorithms that start with a careful spectral initialization [6], [10], [11], which is then iteratively refined by a gradient-descent-like scheme of low computational complexity.

More recently, an alternative convex formulation of the phase retrieval problem in the original n -dimensional parameter space was independently proposed by two groups of authors [12], [13]. The resulting method, referred to as PhaseMax in [13], relaxes the *nonconvex equality* constraints in (1) to *convex inequality* constraints, and solves the following linear program:

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

Here, \mathbf{x}_{init} represents an initial guess (or “anchor vector”) that is correlated with the target vector ξ .

Despite its simple formulation, PhaseMax has strong theoretical performance guarantees. Existing analysis [12]–[14] shows that PhaseMax achieves exact signal recovery from a nearly optimal number of *random measurements*. Specifically, in the case when the sensing vectors are drawn from the Gaussian distribution, the required number of measurements for perfect reconstruction is shown to be linear with respect to the underlying dimension, *i.e.*, $m = cn$ for some constant c that depends on the quality of the initial vector \mathbf{x}_{init} . The analysis in [12]–[14] gives various upper bounds on the constant c . The exact value of c , namely the sharp *phase transition threshold*, is predicted in a recent work [15] by a subset of the authors of the current paper, but the analysis in [15] uses the non-rigorous replica method from statistical physics.

B. Contributions

Our main contributions in this paper are two-fold.

1. *Exact recovery guarantees.* We present an exact performance analysis of the PhaseMax method for the (real-valued) phase retrieval problem with Gaussian sensing vectors in the large system limit. When $m, n \rightarrow \infty$ at a proportional ratio $\alpha = m/n$, we *rigorously* establish the exact phase transition threshold. Furthermore, in the regime where perfect recovery is not feasible, we derive asymptotically exact formulas for the normalized mean squared error (NMSE), defined as

$$\text{NMSE}_n \stackrel{\text{def}}{=} \min\{\|\xi - \hat{\mathbf{x}}\|_2^2, \|\xi + \hat{\mathbf{x}}\|_2^2\} / \|\xi\|_2^2.$$

Our formulas reveal the precise dependence of the NMSE on the oversampling ratio α and on the quality of the initial guess

\mathbf{x}_{init} as measured via the input cosine similarity

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

Our main results can be summarized by the following asymptotic characterization of the NMSE:

$$\text{NMSE}_n \xrightarrow{n \rightarrow \infty} \begin{cases} 0 & , \text{if } \rho_{\text{init}} > \rho_c(\alpha), \\ f(\rho_{\text{init}}, \alpha) > 0 & , \text{otherwise,} \end{cases} \quad (4)$$

where

$$\rho_c(\alpha) \stackrel{\text{def}}{=} \sqrt{1 - \frac{\pi/\alpha}{\tan(\pi/\alpha)}}, \quad (5)$$

and $f(\rho_{\text{init}}, \alpha)$ is explicitly determined by solving a one-dimensional deterministic fixed point equation [see (8) and Theorem 1.]

Our analysis builds upon the recently developed convex Gaussian min-max theorem (CGMT) [16], [17], which involves a tight version of a classical Gaussian comparison inequality [18]. The CGMT framework has been successfully applied to derive precise performance guarantees for structured signal recovery under (noisy) linear Gaussian measurements, *e.g.*, [16], [17], [19], [20]. In [21], the CGMT is used to study signal recovery from a class of non-linear measurements. However, this excludes magnitude-only or quadratic measurements that are relevant for the phase retrieval problem considered here.

The precise nature of our results serves to tighten up the previously known performance bounds of PhaseMax [12], [13], [22]. They also exactly match and thus rigorously verify the predictions in [15] obtained from the non-rigorous replica method from statistical physics. In fact, to the best of our knowledge, this is the first *exact* performance analysis of any of the existing solution methods for the phase retrieval problem.

2. From precise analysis to algorithmic improvements. As the second contribution of this paper, we propose a new nonconvex formulation and an efficient iterative algorithm for the phase retrieval problem. Our new formulation is inspired by PhaseMax, the key idea of which is to relax the nonconvex equality constraints in (1) to convex *inequality* constraints. The intersections of all these inequality constraints form a high-dimensional (random) polytope. Our analysis of the PhaseMax method provides useful insights on the exact high-dimensional geometry of that random polytope. These insights then lead us to a novel nonconvex formulation of the phase retrieval problem, as follows:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \max_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_2^2 \\ \text{s.t.} \quad & \left| \mathbf{a}_i^T \mathbf{x} \right| \leq y_i, \quad \text{for } 1 \leq i \leq m. \end{aligned} \quad (6)$$

Note that (6) is indeed a nonconvex problem, as we aim to *maximize* a convex function over a convex domain. We devise an efficient iterative method, which we call *PhaseLamp*, to solve (6). The name comes from the fact that the algorithm is based on the idea of successive linearization and maximization over a polytope, where in each step we solve a PhaseMax

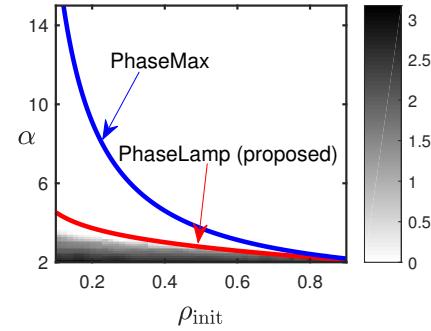


Fig. 1. The NMSE of the PhaseLamp method: theory versus simulations. The signal dimension is set to $n = 1000$, and the results are averaged over 10 independent trials. The red curve shows the sufficient condition, as given in (7), for PhaseLamp to successfully recover the target signal. This is compared against the blue curve, which shows the phase transition boundary of the original PhaseMax method as given in (5).

problem with the initialization given by the estimate from the previous iteration.

We prove that the proposed PhaseLamp method has (strictly) superior recovery performance over PhaseMax. Specifically, we show that a *sufficient condition* for PhaseLamp to perfectly recover the target signal $\boldsymbol{\xi}$ is

$$\rho_{\text{init}} > \rho_s(\alpha), \quad (7)$$

where $\rho_s(\alpha)$ is determined explicitly by solving a one-dimensional deterministic fixed point equation (see (17) and Theorem 3.) In particular, $\rho_s(\alpha)$ is strictly smaller than $\rho_c(\alpha)$ as defined in (5). This is illustrated through a numerical example shown in Figure 1. We can see that the proposed PhaseLamp method significantly improves the recovery performance of the PhaseMax method, especially in the more challenging, and arguably the more practically relevant regime of small input cosine similarities ρ_{init} . Moreover, although (7) is only a sufficient condition, it nevertheless provides a good estimate of the actual performance of the algorithm.

II. PRECISE PERFORMANCE ANALYSIS OF PHASEMAX

A. Technical Assumptions

The asymptotic predictions derived in this paper are based on the following assumptions.

- (A.1) The sensing vectors $\{\mathbf{a}_i\}_{1 \leq i \leq m}$ are known and drawn independently from a Gaussian distribution with zero mean and covariance matrix \mathbf{I}_n .
- (A.2) $m = m(n)$ with $\alpha_n = m(n)/n \rightarrow \alpha > 0$ as $n \rightarrow \infty$.
- (A.3) Both the target signal vector $\boldsymbol{\xi}$ and the initial guess vector \mathbf{x}_{init} are independent from the sensing vectors $\{\mathbf{a}_i\}_{1 \leq i \leq m}$.

For convenience, we shall also assume that the initial guess vector \mathbf{x}_{init} has a *positive* correlation with the target signal vector $\boldsymbol{\xi}$, *i.e.*, $\boldsymbol{\xi}^T \mathbf{x}_{\text{init}} > 0$. This can be made without loss of generality since the vectors $\boldsymbol{\xi}$ and $-\boldsymbol{\xi}$ are both valid target vectors.

B. Fundamental Limits of PhaseMax

In this section, we characterize the asymptotic NMSE of the PhaseMax method under the stated assumptions. In particular,

our results point out necessary and sufficient conditions on the oversampling ratio α and on the cosine similarity ρ_{init} for perfect recovery.

In order to state our results we need a few definitions. For any fixed cosine similarity ρ_{init} and fixed oversampling ratio $\alpha > 2$, define s^* as follows:

$$s^* \stackrel{\text{def}}{=} \arg \max_{0 \leq s \leq 1} \rho_{\text{init}} s + \sqrt{(1 - \rho_{\text{init}}^2) g_\alpha(s)} \quad (8)$$

where the function $g_\alpha(s) : [-1, 1] \rightarrow (0, \infty)$ (parametrized by α) is given by

$$g_\alpha(s) \stackrel{\text{def}}{=} -1 - s^2 + \frac{2\alpha r_\alpha(s)}{\pi} + \frac{2\alpha s}{\pi} \tan\left(\frac{s}{r_\alpha(s) + c_\alpha}\right), \quad (9)$$

with $c_\alpha = 1/\tan(\pi/\alpha)$ and

$$r_\alpha(s) \stackrel{\text{def}}{=} \sqrt{c_\alpha^2 + 1 - s^2} - c_\alpha. \quad (10)$$

Moreover, define

$$r^* \stackrel{\text{def}}{=} r_\alpha(s^*). \quad (11)$$

Theorem 1 (NMSE of PhaseMax): For any fixed input cosine similarity $\rho_{\text{init}} > 0$ and any fixed oversampling ratio $\alpha > 2$, let s^*, r^* be defined as in (8) and (11), respectively. Then, under the assumptions in Section II-A, the NMSE of the PhaseMax method converges in probability as follows:

$$\text{NMSE}_n \xrightarrow{n \rightarrow \infty} 1 + (s^*)^2 + (r^*)^2 - 2|s^*|. \quad (12)$$

The proof of Theorem 1 is based on the CGMT [16], [17]. To streamline our presentation, we postpone a sketch of the proof to the appendix. Theorem 1 accurately predicts the NMSE of PhaseMax in the large system limit. The prediction is expressed in terms of s^* , the solution to the one-dimensional deterministic maximization problem in (8). It can be shown that this optimization problem is concave and that s^* can be uniquely determined by a fixed point equation.

Theorem 2 (Phase transition of PhaseMax): For any fixed cosine similarity ρ_{init} and any fixed oversampling ratio $\alpha > 2$, the PhaseMax method perfectly recovers the target signal (in the sense that $\text{NMSE}_n \xrightarrow{n \rightarrow \infty} 0$, in probability) if and only if

$$\rho_{\text{init}} > \sqrt{1 - \frac{\pi/\alpha}{\tan(\pi/\alpha)}} \stackrel{\text{def}}{=} \rho_c(\alpha). \quad (13)$$

A sketch of the proof of Theorem 2 can be found at the end of the appendix. The theorem establishes a precise phase transition behavior on the performance of PhaseMax: for any fixed oversampling ratio $\alpha > 2$, there is a critical cosine similarity $\rho_c(\alpha)$ such that the PhaseMax method perfectly recovers the target signal vector ξ if and only if $\rho_{\text{init}} > \rho_c(\alpha)$.

C. Numerical Simulations

In this section, we present simulation results that verify the validity of our predictions given in Theorems 1 and 2. We solve the convex optimization problem (2) using the technique introduced in [23]. The signal dimension is set to $n = 1000$.

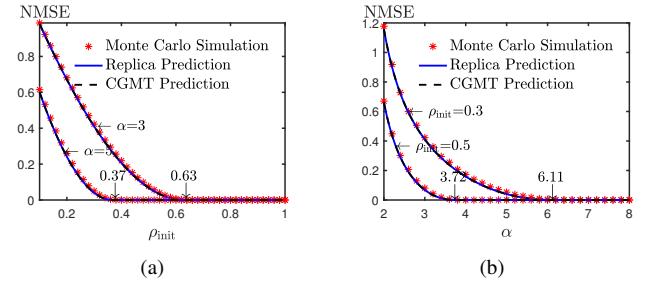


Fig. 2. Asymptotic predictions v.s. numerical simulations. (a) The NMSE of the PhaseMax method as a function of ρ_{init} , for two different values of α ; (b) The NMSE of the PhaseMax method as a function of α , for two different values of ρ_{init} . The results are averaged over 50 independent Monte Carlo trials. The CGMT expressions in this paper verify the non-rigorous replica predictions derived in [15]. The asymptotic formulas are also in excellent agreement with the actual performance the PhaseMax method, for $n = 1000$.

In the first example, we investigate the performance of our asymptotic predictions for PhaseMax given in (12). Specifically, we compare the asymptotic predictions against simulation results for different values of the input cosine similarity ρ_{init} and the oversampling ratio α . Figure 2(a) illustrates the NMSE of PhaseMax as a function of the input cosine similarity given in (3), for two different values of the oversampling ratio. It can be noticed that our asymptotic predictions of the PhaseMax performance obtained using the CGMT perfectly match the asymptotic predictions derived using the non-rigorous replica method [15]. The asymptotic predictions are also in excellent agreement with the actual performance of the PhaseMax method in finite dimensions. Figure 2(b) further shows that the critical cosine similarity for the considered oversampling ratio α is given by $\rho_{\text{init}}(\alpha = 3) \approx 0.63$ and $\rho_{\text{init}}(\alpha = 5) \approx 0.37$, respectively, which perfectly matches the theoretical predictions given in Theorem 2.

Figure 2(b) provides an additional example, where we plot the NMSE of the PhaseMax method as a function of the oversampling ratio, for two different values of the input cosine similarity. Again, the asymptotic performance obtained in our analysis perfectly matches the actual performance of the algorithm.

III. ALGORITHMIC IMPROVEMENTS

A. PhaseLamp: A New Algorithm

This section proposes an efficient iterative algorithm to solve the norm maximization problem formulated in (6). The optimization problem (6) consists of maximizing a convex function over a convex feasibility set. Hence, it is nonconcave where the cost function can be written as a difference of concave functions. To solve this problem, we propose the following scheme, named PhaseLamp, based on the idea of successive *linearization and maximization over a polytope*:

$$\begin{aligned} \mathbf{x}_{k+1} = \arg \max_{\mathbf{x}} \quad & \mathbf{x}_k^T \mathbf{x} \\ \text{s.t.} \quad & \left| \mathbf{a}_i^T \mathbf{x} \right| \leq y_i, \text{ for } 1 \leq i \leq m. \end{aligned} \quad (14)$$

In essence, at each iteration we approximate (*i.e.*, linearize) the cost function of (6) via

$$\|\mathbf{x}\|_2^2 \approx \mathbf{x}_k^T \mathbf{x},$$

where \mathbf{x}_k is the estimate obtained from the previous iteration. The proposed algorithm starts from an initial guess $\mathbf{x}_0 = \mathbf{x}_{\text{init}}$ of the target vector ξ . In practice, we terminate the proposed iterative algorithm when the number of iterations exceeds a pre-specified number I_{max} or when $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2 \leq \epsilon$ for some fixed threshold $\epsilon > 0$.

There are several ways to interpret the proposed PhaseLamp algorithm. First, it can be viewed as an iterative and *bootstrapped* version of the PhaseMax method (2) where at each iteration the previous optimal solution is used as an (improved) initial guess of the target signal vector ξ . Second, PhaseLamp is a special case of a minorize-maximization (MM) algorithm [24]. To see this, we note from the convexity of the cost function $\|\mathbf{x}\|_2^2$ that

$$\|\mathbf{x}\|_2^2 \geq \mathbf{x}_k^T \mathbf{x}_k + 2\mathbf{x}_k^T (\mathbf{x} - \mathbf{x}_k), \forall \mathbf{x}_k, \mathbf{x} \in \mathbb{R}^n. \quad (15)$$

The PhaseLamp procedure consists of iteratively maximizing the lower bound in (15) over the convex feasibility set in (6). One particular property of the MM procedure is that it guarantees that the optimal objective value of the optimization problem (14) is nondecreasing, *i.e.*,

$$\|\mathbf{x}_{k+1}\|_2^2 \geq \|\mathbf{x}_k\|_2^2, \forall k \geq 1. \quad (16)$$

Due to the nonconcavity of the maximization problem (6), the proposed iterative algorithm is not guaranteed to converge to the global optimal solution of (6). One particular property of the fixed points of the optimization problem (14) is that it is an extreme point of the feasibility set given in (6).

B. Performance Guarantees for PhaseLamp

Using the analysis strategy that leads to Theorems 1 and 2, we are further able to derive a sufficient condition for PhaseLamp to perfectly recover the target signal ξ .

Again, we first need a few definitions. One can show that, for any $\alpha > 2$, the equation

$$\begin{aligned} \theta \cos^2 \theta + (1 + 3 \sin^2 \theta) \operatorname{atan} \left(\frac{\sin \theta \cos \theta}{1 + \sin^2 \theta} \right) \\ = 2 \sin \theta \cos \theta + \left(\frac{\pi}{\alpha} \right) \sin^2 \theta \cos^2 \theta, \end{aligned} \quad (17)$$

has a unique solution in the interval $\theta \in (0, \pi/2)$. We denote that solution by θ_α^* . Let

$$\hat{s}_\alpha \stackrel{\text{def}}{=} \frac{\tan(\theta_\alpha^*)}{\sqrt{1 + c_\alpha^2 + \tan(\theta_\alpha^*)^2 + c_\alpha^2}}, \quad (18)$$

where $c_\alpha = 1/\tan(\pi/\alpha)$, and

$$\ell_\alpha \stackrel{\text{def}}{=} \frac{\hat{s}_\alpha - \frac{\alpha}{\pi} \operatorname{atan} \left(\frac{\hat{s}_\alpha}{\sqrt{c_\alpha^2 + 1 - \hat{s}_\alpha^2}} \right)}{\sqrt{g_\alpha(\hat{s}_\alpha)}}, \quad (19)$$

where g_α is the function defined in (9). We are now ready to state the main theorem of this section.

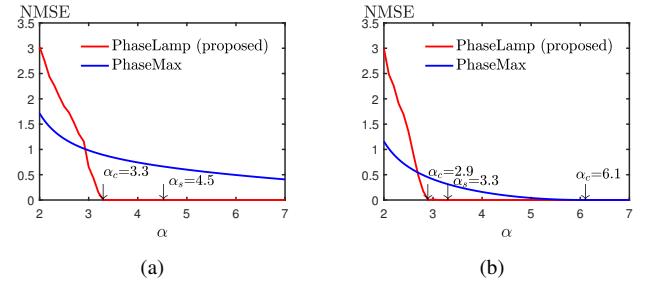


Fig. 3. Sufficient condition for successful recovery of the PhaseLamp method. The NMSE as a function of the oversampling ratio α , for (a) $\rho_{\text{init}} = 0.1$; (b) $\rho_{\text{init}} = 0.3$. The maximum number of iterations is set to $I_{\text{max}} = 20$ and $\epsilon = 10^{-4}$. The results are averaged over 10 independent Monte Carlo trials. In both cases, the PhaseLamp method outperforms the PhaseMax method. Moreover, the sufficient condition for perfect recovery (20) denoted by α_s provides a good estimate of the actual recovery performance of the PhaseLamp method.

Theorem 3 (Sufficient condition for perfect recovery): PhaseLamp perfectly recovers the unknown signal, *i.e.*, it holds in probability that $\text{NMSE}_n \xrightarrow{n \rightarrow \infty} 0$, if

$$\rho_{\text{init}} > \frac{\ell_\alpha}{\sqrt{\ell_\alpha^2 + 1}} =: \rho_s(\alpha). \quad (20)$$

The proof of Theorem 3 is based on CGMT and the properties of the fixed points of the optimization problem (14). Due to space constraint, we defer the proof of this theorem to the long version of the current paper. Unlike the asymptotically exact characterization given in Theorem 2, the condition given in Theorem 3 is sufficient but not necessary. However, as shown in Figure 1 and the additional simulation results given in the next section, the condition in (20) provides a reasonably tight bound on the actual performance of the PhaseLamp algorithm.

C. Numerical Results

We present some numerical results to illustrate the performance of the proposed PhaseLamp method and our theoretical predictions given in Theorem 3. In our experiments, the signal dimension is set to $n = 1000$. Figure 3 plots the NMSE as a function of the oversampling ratio for two different values of the input cosine similarity ($\rho_{\text{init}} = 0.1$ and $\rho_{\text{init}} = 0.3$, respectively.)

We observe that the proposed PhaseLamp method indeed outperforms the original PhaseMax method, and the amount of improvement is greater when the input cosine similarity is smaller. Specifically, for $\rho_{\text{init}} = 0.1$, the empirical minimum sampling ratio for PhaseLamp to perfectly recover ξ is at $\alpha \approx 3.3$, whereas PhaseMax requires $\alpha \approx 18.2$. [The latter point is not shown in Figure 3(a).] Moreover, Figure 3 also shows that the sufficient condition developed in Theorem 3 for PhaseLamp provides a good estimate of the actual performance of the proposed algorithm. For example, Figure 3(b) demonstrates that the actual critical oversampling ratio of PhaseLamp is at $\alpha_c \approx 2.9$, whereas the sufficient oversampling ratio as given in Theorem 3 is $\alpha_s \approx 3.3$.

IV. CONCLUSION

We presented in this paper an asymptotically exact characterization of the performance of the PhaseMax method for phase retrieval. Specifically, our analysis reveals a sharp phase transition behavior in the performance of the method as one varies the oversampling ratio and the input cosine similarity. Our analysis is based on the CGMT, and the results match previous predictions derived from the non-rigorous replica method. Moreover, we also presented a new nonconvex formulation of the phase retrieval problem and PhaseLamp, an iterative algorithm based on linearization and maximization over a polytope. We provided a sufficient condition for PhaseLamp to perfectly retrieve the target vector. Simulation results confirm the validity of our theoretical predictions. They also show that the proposed iterative algorithm significantly improves the recovery performance of the original PhaseMax method.

APPENDIX

In this appendix we provide proof sketches for Theorems 1 and 2.

A. Notation

To simplify the exposition we assume onwards that $\xi = e_1$, the first vector of the canonical basis of \mathbb{R}^n , and that $\|x_{\text{init}}\|_2 = 1$. This assumption can be made without loss of generality due to rotational invariance of the Gaussian distribution and since the optimization problems (2) and (6) are scale invariant.

Further, we stack the sensing vectors $\{\mathbf{a}_i^T, 1 \leq i \leq m\}$ to form the sensing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Let \mathbf{q} denote the first column of \mathbf{A} and $\mathbf{G} \in \mathbb{R}^{m \times (n-1)}$ the remaining part, *i.e.*,

$$\mathbf{A} = [\mathbf{q} \ \mathbf{G}].$$

Similarly, define η_1 and $\tilde{\eta}$ such that

$$\mathbf{x}_{\text{init}} = [\eta_1 \ \tilde{\eta}^T]^T.$$

Finally, for a vector \mathbf{c} , we let $|\mathbf{c}|$ and $\text{sign}(\mathbf{c})$ to denote its component-wise absolute value and sign, respectively. Also, we let $\min(\mathbf{c})$ return the minimum value in the vector, and, $\mathbf{z} = \mathbf{c} \wedge \mathbf{0}$ be a vector such that $\mathbf{z}_i = \min(c_i, 0)$.

B. Convex Gaussian Min-Max Theorem (CGMT)

The proof follows the CGMT framework [16], [17]. For ease of reference we summarize here the essential ideas of the framework; please see [16, Section 6] for the formal statement of the theorem and further details. The CGMT associates with a primary optimization (PO) problem a simplified auxiliary optimization (AO) problem from which we can tightly infer properties of the original (PO), including the optimal cost and the optimal solution. The two problems are of the following form:

$$\Phi(\mathbf{C}) = \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \mathbf{u}^T \mathbf{C} \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (21)$$

and

$$\phi(\mathbf{g}, \mathbf{h}) = \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \|\mathbf{w}\|_2 \mathbf{g}^T \mathbf{u} + \|\mathbf{u}\|_2 \mathbf{h}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u}), \quad (22)$$

respectively. Above, $\mathbf{C} \in \mathbb{R}^{m \times n}$, $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{h} \in \mathbb{R}^n$, the sets $\mathcal{S}_w \subset \mathbb{R}^n$ and $\mathcal{S}_u \subset \mathbb{R}^m$ are compact, and, $\psi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$. We think of the two problems in (21) and (22) as random optimization problems, in which \mathbf{C} , \mathbf{g} and \mathbf{h} all have i.i.d standard normal entries. According to the CGMT [16, Theorem 6.1], if the sets \mathcal{S}_w and \mathcal{S}_u are convex and ψ is continuous *convex-concave* on $\mathcal{S}_w \times \mathcal{S}_u$, then, for any $\mu \in \mathbb{R}$ and $t > 0$, it holds

$$\mathbb{P}(|\Phi(\mathbf{C}) - \mu| > t) \leq 2\mathbb{P}(|\phi(\mathbf{g}, \mathbf{h}) - \mu| > t). \quad (23)$$

In words, concentration of the optimal cost of the AO problem around μ implies concentration of the optimal cost of the corresponding PO problem around the same value μ . Moreover, starting from (23) and under strict convexity conditions, the CGMT shows that concentration of the optimal solution of the AO problem implies concentration of the optimal solution of the PO to the same value. For example, if minimizers of (22) satisfy $\|\mathbf{w}^*(\mathbf{g}, \mathbf{h})\|_2 \rightarrow \zeta^*$ for some $\zeta^* > 0$, then, the same holds true for the minimizers of (21): $\|\mathbf{w}^*(\mathbf{C})\|_2 \rightarrow \zeta^*$ [16, Theorem 6.1(iii)]. Thus, one can analyze the AO to infer corresponding properties of the PO, the premise being of course that the former is simpler to handle than the latter.

C. CGMT for the PhaseMax Method

We apply the CGMT to characterize the asymptotic NMSE of the PhaseMax optimization in (2) as in (12) and (13). In this section, we write the PhaseMax optimization in the form of a PO as in (21), which in turn leads to a corresponding AO optimization problem. For these problems, we can show that the conditions of the CGMT on convexity and compactness are satisfied.

First, we appropriately write the linear program in (2) as a minmax program. Start with its dual:

$$\min_{\substack{\boldsymbol{\lambda} \geq 0 \\ \boldsymbol{\mu} \geq 0}} \boldsymbol{\mu}^T |\mathbf{A} \boldsymbol{\xi}| + \boldsymbol{\lambda}^T |\mathbf{A} \boldsymbol{\xi}| \quad \text{s.t.} \quad \mathbf{A}^T \boldsymbol{\mu} - \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{x}_{\text{init}}. \quad (24)$$

Since $m > n$ (recall: $\alpha > 2$) both the primal and the dual are bounded feasible with probability one, and strong duality holds [25]. Therefore, (24) is equivalent to the following

$$\min_{\substack{\boldsymbol{\lambda} \in \mathcal{S}_\lambda \\ \boldsymbol{\mu} \in \mathcal{S}_\mu}} \max_{\mathbf{x}} \mathbf{x}_{\text{init}}^T \mathbf{x} + \boldsymbol{\mu}^T (|\mathbf{A} \boldsymbol{\xi}| - \mathbf{A} \mathbf{x}) + \boldsymbol{\lambda}^T (|\mathbf{A} \boldsymbol{\xi}| + \mathbf{A} \mathbf{x}), \quad (25)$$

for $\mathcal{S}_\lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^m \text{ s.t. } \|\boldsymbol{\lambda}\|_\infty \leq \Gamma, \boldsymbol{\lambda} \geq 0\}$, $\mathcal{S}_\mu = \{\boldsymbol{\mu} \in \mathbb{R}^m \text{ s.t. } \|\boldsymbol{\mu}\|_\infty \leq \Gamma, \boldsymbol{\mu} \geq 0\}$, and Γ a sufficiently large positive constant. Note that the feasibility sets \mathcal{S}_λ and \mathcal{S}_μ are convex and compact in \mathbb{R}^m , the feasibility set of the variable \mathbf{x} is convex and the cost function in (25) is linear with respect to each optimization variable. Thus, the order of min-max can be flipped to obtain the following equivalent,

$$\max_{\mathbf{x}} \min_{\substack{\boldsymbol{\lambda} \in \mathcal{S}_\lambda \\ \boldsymbol{\mu} \in \mathcal{S}_\mu}} \mathbf{x}_{\text{init}}^T \mathbf{x} + \boldsymbol{\mu}^T (|\mathbf{A} \boldsymbol{\xi}| - \mathbf{A} \mathbf{x}) + \boldsymbol{\lambda}^T (|\mathbf{A} \boldsymbol{\xi}| + \mathbf{A} \mathbf{x}). \quad (26)$$

From strict duality, this is also equivalent to (2). Furthermore, as already discussed, the optimality set of (2) is bounded.

Naturally then there is a sufficiently large constant $B > 0$ and sets $\mathcal{S}_{x_1} = \{x_1 \in \mathbb{R} \text{ s.t. } |x_1| \leq B\}$ and $\mathcal{S}_{\tilde{x}} = \{\tilde{x} \in \mathbb{R}^{n-1} \text{ s.t. } \|\tilde{x}\|_\infty \leq B\}$ such that (26) can be formulated as

$$\begin{aligned} \max_{\substack{x_1 \in \mathcal{S}_{x_1} \\ \tilde{x} \in \mathcal{S}_{\tilde{x}}}} \min_{\substack{\lambda \in \mathcal{S}_\lambda \\ \mu \in \mathcal{S}_\mu}} & \eta_1 x_1 + \tilde{\eta}^T \tilde{x} + (\lambda - \mu)^T q x_1 + (\lambda - \mu)^T G \tilde{x} \\ & + (\lambda + \mu)^T |q| \end{aligned} \quad (27)$$

At this point, observe that (27) is in the desired form of a PO as in (21) with $G \in \mathbb{R}^{m \times (n-1)}$ having i.i.d standard normal entries and the function ψ , defined as

$$\psi(\mathbf{x}, (\lambda, \mu)) = \eta_1 x_1 + \tilde{\eta}^T \tilde{x} + (\lambda - \mu)^T q x_1 + (\lambda + \mu)^T |q|.$$

Further note that the constraint sets are convex compact and ψ is concave-convex on $\mathcal{S}_x \times (\mathcal{S}_\lambda \times \mathcal{S}_\mu)$, where $\mathbf{x} = [x_1 \ \tilde{x}^T]^T$ and $\mathcal{S}_x = \mathcal{S}_{x_1} \times \mathcal{S}_{\tilde{x}}$.

We are now ready to formulate the corresponding AO problem:

$$\begin{aligned} \max_{\substack{x_1 \in \mathcal{S}_{x_1} \\ \tilde{x} \in \mathcal{S}_{\tilde{x}}}} \min_{\substack{\lambda \in \mathcal{S}_\lambda \\ \mu \in \mathcal{S}_\mu}} & \|\tilde{x}\|_2 g^T (\lambda - \mu) + \|\lambda - \mu\|_2 h^T \tilde{x} + \eta_1 x_1 \\ & + \tilde{\eta}^T \tilde{x} + (\lambda - \mu)^T q x_1 + (\lambda + \mu)^T |q|. \end{aligned} \quad (28)$$

Following the CGMT framework we proceed onwards with analyzing (28).

D. Analysis of the Auxiliary Optimization Problem

1) *Simplifying the AO*: Consider the following change of variables: $\mathbf{v} = \lambda - \mu$ and $\mathbf{b} = \lambda + \mu$. To respect the nonnegativity of λ and μ , it must be that $\mathbf{b} \geq |\mathbf{v}|$. In fact, it can be checked that the optimal solution for \mathbf{b} is $\mathbf{b} = |\mathbf{v}|$. Thus, the optimization problem (28) can be reduced to the following:

$$\begin{aligned} \max_{\substack{x_1 \in \mathcal{S}_{x_1} \\ \tilde{x} \in \mathcal{S}_{\tilde{x}}}} \min_{\mathbf{v} \in \mathcal{S}_v} & (\|\tilde{x}\|_2 g + x_1 q)^T \mathbf{v} + \|\mathbf{v}\|_2 h^T \tilde{x} + \eta_1 x_1 \\ & + \tilde{\eta}^T \tilde{x} + |\mathbf{v}|^T |q|. \end{aligned}$$

Above, the set \mathcal{S}_v is defined as $\mathcal{S}_v = \{\mathbf{v} \in \mathbb{R}^m \text{ s.t. } \|\mathbf{v}\|_\infty \leq \Delta\}$, where Δ is a sufficiently large positive constant.

Next, observe that if we fix $|\mathbf{v}|$, then the optimal \mathbf{v} satisfies $\text{sign}(\mathbf{v}) = -\text{sign}(\|\tilde{x}\|_2 g + q x_1)$ which simplifies the optimization to the following

$$\begin{aligned} \max_{\substack{x_1 \in \mathcal{S}_{x_1} \\ \tilde{x} \in \mathcal{S}_{\tilde{x}}}} \min_{\Delta \geq |\mathbf{v}| \geq 0} & [|\mathbf{q}| - \|\tilde{x}\|_2 g + q x_1]^T \mathbf{v} + \|\mathbf{v}\|_2 h^T \tilde{x} \\ & + \eta_1 x_1 + \tilde{\eta}^T \tilde{x}. \end{aligned}$$

Next, in the optimization above one can fix the norm of \mathbf{v} and optimize over its direction. Omitting some details, the optimization becomes

$$\begin{aligned} \max_{\substack{x_1 \in \mathcal{S}_{x_1} \\ \tilde{x} \in \mathcal{S}_{\tilde{x}}}} & \eta_1 x_1 + \tilde{\eta}^T \tilde{x} \\ \text{s.t. } & h^T \tilde{x} + h(|\mathbf{q}| - \|\tilde{x}\|_2 g + q x_1) \geq 0, \end{aligned} \quad (29)$$

where we defined the function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$h(\mathbf{c}) = \begin{cases} -\|\mathbf{c} \wedge \mathbf{0}\|_2 & , \text{if } \min(\mathbf{c}) \leq 0, \\ \min(\mathbf{c}) & , \text{otherwise.} \end{cases}$$

The final step in simplifying the AO problem is as follows. For fixed value of x_1 (say $x_1 = s > 0$), and for fixed norm of \tilde{x} (say, $\|\tilde{x}\|_2 = r$), we optimize over the direction of \tilde{x} . It can be shown that this optimization further reduces (29) to the following two-dimensional optimization problem:

$$\begin{aligned} \max_{\substack{s \in \mathcal{S}_{x_1} \\ B \geq r \geq 0}} & \eta_1 s + \|\tilde{\eta}\|_2 \sqrt{r^2 - c(s, r)^2} \\ \text{s.t. } & c(s, r) \leq r, \end{aligned} \quad (30)$$

where the function $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is defined as

$$c(s, r) = -\frac{h(|\mathbf{q}| - |r\mathbf{g} + s\mathbf{q}|)}{\|\mathbf{h}\|_2}.$$

2) *Convergence Analysis*: Now that we have simplified the AO to a maximization problem over only two scalar variables as in (30), we are ready to study its asymptotic behavior in the regime $m, n \rightarrow \infty, m/n \rightarrow \alpha$. Specifically, it can be shown that the optimization problem (30) converges to the following *deterministic* optimization problem

$$\begin{aligned} \max_{\substack{|s| < B \\ B \geq r \geq 0}} & \eta_1 s + \|\tilde{\eta}\|_2 \sqrt{r^2 - \alpha c_d(r, s)} \\ \text{s.t. } & c_d(s, r) \leq r^2/\alpha, \end{aligned} \quad (31)$$

where $c_d(s, r) = \mathbb{E}_{q, g} [\min \{|\mathbf{q}| - |r\mathbf{g} + s\mathbf{q}|, 0\}^2]$, which takes the following closed-form:

$$\begin{aligned} c_d(s, r) = \frac{1}{\pi} & \left[((1-s)^2 + r^2) \left(\frac{\pi}{2} - \text{atan} \left(\frac{1-s}{r} \right) \right) \right. \\ & \left. + ((1+s)^2 + r^2) \left(\frac{\pi}{2} - \text{atan} \left(\frac{1+s}{r} \right) \right) - 2r \right]. \end{aligned} \quad (32)$$

The full technical details of obtaining the convergence result in (31) are deferred to the full version of the paper. In short, pointwise convergence of the objective function of (30) to (31) for fixed s and r follows easily from the weak law of large numbers. The corresponding convergence of the optimal costs requires proof of uniform convergence, which follows by pointwise convergence and concavity of the objective function [26, Lemma 7.75]. We call the deterministic two-dimensional optimization problem in (31) as the scalar performance optimization (SPO); according to the CGMT solving the SPO allows us to conclude on the asymptotic performance of the PhaseMax problem (cc. the PO).

3) *Solving the scalar performance optimization*: Recall that the SPO in (31) is the converging limit of the AO in (28). Specifically, the optimization variables s and r in (31) correspond exactly to x_1 and $\|\tilde{x}\|_2$ in (28). From this and uniform convergence discussed previously, the optimal values of s and r are the converging limits of x_1 and of $\|\tilde{x}\|_2$, respectively. In what follows, we solve the SPO problem for

the optimal s and r . First, using the assumption of the theorem that $\alpha > 2$, it can be shown that the feasible set of (31) is nonempty iff $|s| \leq 1$. Second, for fixed $|s| \leq 1$, the problem

$$\max_{r \geq 0} r^2 - \alpha c_d(s, r) \quad (33)$$

is concave and admits a unique solution

$$r^*(s) = \sqrt{\frac{1}{\tan(\frac{\pi}{\alpha})^2} + (1 - s^2)} - \frac{1}{\tan(\frac{\pi}{\alpha})}. \quad (34)$$

At this point, note that we can always find a large enough constant $\tilde{B} > 0$ such that $r^*(s) < \tilde{B}$ for all $|s| < 1$. Therefore, choosing B in (31) such that $B = \tilde{B}$ guarantees that the optimal value of r in (31) is given by (34). Substituting this value back in (31), we can now optimize over s by solving the following:

$$\max_{|s| \leq 1} \eta_1 s + \|\tilde{\eta}\|_2 \sqrt{(r^*(s))^2 - \alpha c_d(r^*(s), s)}. \quad (35)$$

A few algebra manipulations show that (35) is equivalent to (8) in the statement of the theorem. To show the equivalence, further note that η_1 and $\tilde{\eta}$ in (35) are related to the input cosine similarity ρ_{init} , defined in (3), as follows (recall: $\xi = e_1$),

$$\frac{\eta_1}{\|\tilde{\eta}\|_2} = \frac{\rho_{\text{init}}}{\sqrt{1 - \rho_{\text{init}}^2}}. \quad (36)$$

Finally, note that the optimization in (35) [eqv., in (8)] inherits the concavity of (31), *i.e.*, it is a concave program.

4) *Phase transition calculations:* In this section, we compute the phase transition boundary of the PhaseMax method. Our goal is to find necessary and sufficient conditions under which the solution \hat{x} of the PhaseMax is, with high probability, equal to $\xi = e_1$. Mapping this to the SPO in (35) [eqv., see (8)], we seek conditions under which $s^* = 1$ and $r^* = 0$. From concavity, this happens if and only if the derivative of the cost function of the optimization problem (8) at $s = 1$ is nonnegative. Hence, the necessary and sufficient condition for perfect recovery of the PhaseMax method is given by

$$\frac{\rho_{\text{init}}}{\sqrt{1 - \rho_{\text{init}}^2}} > \sqrt{\frac{\alpha}{\pi} \tan\left(\frac{\pi}{\alpha}\right) - 1}, \quad (37)$$

for $\alpha > 2$. Equivalently, the oversampling ratio α and the input cosine similarity given in (3) must satisfy

$$\frac{\pi}{\alpha \tan(\pi/\alpha)} > 1 - \rho_{\text{init}}^2. \quad (38)$$

This then gives us the statement of Theorem 2.

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] R. Balan, B. G. Bodmann, P. G. Casazza, and D. Edidin, "Painless reconstruction from magnitudes of frame coefficients," *Journal of Fourier Analysis and Applications*, vol. 15, no. 4, pp. 488–501, 2009.
- [10] Y. Chen and E. J. Candes, "Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems," *arXiv:1505.05114*, 2015.
- [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] 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>
- [13] 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>
- [14] P. Hand and V. Voroninski, "An Elementary Proof of Convex Phase Retrieval in the Natural Parameter Space via the Linear Program PhaseMax," *arXiv:1611.03935 [cs, math]*, Nov. 2016, arXiv: 1611.03935. [Online]. Available: <http://arxiv.org/abs/1611.03935>
- [15] O. Difallah and Y. M. Lu, "Fundamental Limits of PhaseMax for Phase Retrieval: A Replica Analysis," *Proc. International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017. [Online]. Available: <https://arxiv.org/abs/1708.03355>
- [16] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Precise error analysis of regularized m-estimators in high-dimensions," *CoRR*, vol. abs/1601.06233, 2016. [Online]. Available: <http://arxiv.org/abs/1601.06233>
- [17] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *Proceedings of The 28th Conference on Learning Theory*, vol. 40. Paris, France: PMLR, 03–06 Jul 2015, pp. 1683–1709.
- [18] Y. Gordon, "Some inequalities for Gaussian processes and applications," *Israel Journal of Mathematics*, vol. 50, no. 4, pp. 265–289, Dec 1985.
- [19] M. Stojnic, "A framework to characterize performance of LASSO algorithms," *CoRR*, vol. abs/1303.7291, 2013. [Online]. Available: <http://arxiv.org/abs/1303.7291>
- [20] C. Thrampoulidis, A. Panahi, and B. Hassibi, "Asymptotically exact error analysis for the generalized equation-lasso," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 2021–2025.
- [21] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Lasso with non-linear measurements is equivalent to one with linear measurements," in *Advances in Neural Information Processing Systems*, 2015, pp. 3420–3428.
- [22] P. Hand and V. Voroninski, "An elementary proof of convex phase retrieval in the natural parameter space via the linear program phasemax," *arXiv:1611.03935*, 2016.
- [23] 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>
- [24] K. Lange, *MM Optimization Algorithms*. SIAM, 2016.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.
- [26] K.-J. Miescke and F. Liese, *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer New York, 2008.