Convex Phase Retrieval without Lifting via PhaseMax

Tom Goldstein [* 1] **Christoph Studer** [* 2]

Abstract

Semidefinite relaxation methods transform a variety of non-convex optimization problems into convex problems, but square the number of variables. We study a new type of convex relaxation for phase retrieval problems, called *PhaseMax*, that convexifies the underlying problem without lifting. The resulting problem formulation can be solved using standard convex optimization routines, while still working in the original, low-dimensional variable space. We prove, using a random spherical distribution measurement model, that PhaseMax succeeds with high probability for a sufficiently large number of measurements. We compare our approach to other phase retrieval methods and demonstrate that our theory accurately predicts the success of PhaseMax.

1. Introduction

Semidefinite relaxation is the technique of replacing a broad range of non-convex optimization problems involving a vector of (possibly discrete) variables with convex problems involving matrices. The relaxed convex problem is then solved to global optimality, and the solution is used to extract a global minimizer of the original non-convex problem. Unfortunately, convexity comes at a steep cost: semidefinite relaxation squares the dimensionality of the problem, resulting in formulations that are convex but computationally intractable in many situations. In fact, the increase in dimensionality, which is called *lifting*, often prevents the use of this technique in machine learning and computer vision applications involving thousands to millions of variables.

This article studies convex relaxation for *phase retrieval*, a canonical non-convex problem that can be solved via semidefinite relaxation. We present a relaxation approach that convexifies the problem *without lifting*, i.e., it solves the problem in its original, low-dimensional space.

---

[*]Equal contribution [1]University of Maryland [2]Cornell University. Correspondence to: Tom Goldstein <tomg at cs.umd.edu>.

2. Phase Retrieval Problems

Phase retrieval deals with the recovery of an n-dimensional signal $\mathbf{x}^0 \in \mathcal{H}^n$, with $\mathcal{H}$ either $\mathbb{R}$ or $\mathbb{C}$, from $m \geq n$ magnitude measurements of the form (Candès et al., 2013)

$$b_i = |\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|, \quad i = 1, 2, \ldots, m, \tag{1}$$

where $\mathbf{a}_i \in \mathcal{H}^n$, and $i = 1, 2, \ldots, m$ are the (known) measurement vectors. Because of the nonlinearity caused by measuring the magnitude of the linear terms in (1), the phase retrieval problem is non-convex.

A variety of algorithms exist for finding $\mathbf{x}^0$ in (1). Classical methods, such as the Gerchberg-Saxton and Feinup algorithms, search for a solution via alternating least-squares steps that are efficient when the measurement ensemble $\{\mathbf{a}_i\}$ forms a tight frame (e.g., a collection of Fourier matrices).

More recently, there has been significant interest in convex methods for phase retrieval that provably find global solutions without the danger of getting trapped in local minima. These methods, including PhaseLift (Candès et al., 2013) and its dual formulation PhaseCut (Waldspurger et al., 2015), rely on semidefinite relaxation and replace the unknown n-dimensional vector $\mathbf{x}$ with a (much larger) $n \times n$ matrix of unknowns that is recovered using semi-definite programming. While such methods come with strong theoretical guarantees and do not get trapped in local minima, they require lifting (semidefinite relaxation squares the number of unknowns), which makes this approach intractable for real-world image processing applications involving thousands or millions of variables. As a result, there has been a flurry of interest in global convergence properties of non-convex solvers that operate in the original feature space, such as the methods in (Netrapalli et al., 2013; Schniter & Rangan, 2015; Candès et al., 2015; Chen & Candès, 2015; Wang et al., 2016). While these methods come with theoretical guarantees, non-convexity makes it difficult to combine them with commonly used regularizers, and typically precludes the use of sophisticated optimization methods that can handle constraints, non-differentiable terms, etc.

We note that the PhaseMax formulation has also been proposed by (Bahmani & Romberg, 2016) and was recently analyzed by (Hand & Voroninski, 2016). We discuss and compare these results in Section 8.2.

3. A Word on Notation

Scalars are lower-case letters, and vector quantities are boldface, lower-case letters. Matrices are boldface upper-case letters, and sets are written using upper-case script font. We denote the inner product between the vectors $\mathbf{x}, \mathbf{y} \in \mathcal{H}^n$ as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$, and $\mathbf{x}^T$ is the transpose in the real case or the Hermitian transpose in the complex case. We denote the real part of the inner product as $\langle \mathbf{x}, \mathbf{y} \rangle_\Re = \Re(\mathbf{x}^T \mathbf{y})$ and the imaginary part as $\langle \mathbf{x}, \mathbf{y} \rangle_\Im$ so that $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_\Re + j\langle \mathbf{x}, \mathbf{y} \rangle_\Im$ with $j^2 = -1$. The non-negative reals are denoted $\mathbb{R}_0^+$.

We denote the unit sphere embedded in $\mathbb{R}^n$ as $\mathcal{S}_\mathbb{R}^{n-1}$, and use $\mathcal{S}_\mathbb{C}^{n-1}$ for the sphere in complex space. To address both of these cases at once, we will often write

$$\mathcal{S}_\mathcal{H}^{n-1} = \{\mathbf{x} \in \mathcal{H}^n \,|\, \|\mathbf{x}\|_2 = 1\},$$

where $\mathcal{H} \in \{\mathbb{R}, \mathbb{C}\}$ is either the real or complex numbers.

4. Proposed Relaxation

We propose PhaseMax, a formulation of the phase retrieval problem that avoids lifting. Put simply, PhaseMax relaxes the non-convex equality constraints $|\langle \mathbf{a}_i, \mathbf{x} \rangle| = b_i$ into convex inequality constraints of the form $|\langle \mathbf{a}_i, \mathbf{x} \rangle| \leq b_i$. The resulting convex problem can then be solved using linear programming in the real case or using second-order cone programming in the complex case.

At first glance, the proposed relaxation seems too simplistic to recover $\mathbf{x}^0$. Indeed, the relaxed system of inequalities has a trivial solution: the all-zeros vector. To obtain a meaningful solution, we need to force the solution vector to lie on the boundary of the constraint set, where the inequality constraints hold with equality. To force the solution to lie on the boundary, we rely on some intelligent "guess" $\hat{\mathbf{x}} \in \mathcal{H}^n$ of the solution to (1); we will discuss methods for producing such guesses in Section 8.1. We then find the feasible point that lies as far in the direction of $\hat{\mathbf{x}} \in \mathcal{H}^n$ as possible. This results in the convex optimization problem

$$(\text{PhaseMax}) \begin{cases} \underset{\mathbf{x} \in \mathcal{H}^n}{\text{maximize}} & \langle \mathbf{x}, \hat{\mathbf{x}} \rangle_\Re \\ \text{subject to} & |\langle \mathbf{a}_i, \mathbf{x} \rangle| \leq b_i, \, i = 1, \ldots, m. \end{cases}$$

The key idea behind PhaseMax is that the objective forces the solution vector to lie along the boundary of the constraint set, where the constraints are active. Evidently, if all of the constraints are active at this solution, then we have recovered a solution to the original non-convex problem (1).

Quite surprisingly, the PhaseMax relaxation provably recovers the true solution to (1) in many situations. In particular, we have the following main result:

Theorem 1. *Consider the case of recovering a signal* $\mathbf{x} \in \mathcal{H}^n$ *from* m *measurements of the form (1) with measure-*

ment vectors $\mathbf{a}_i$, $i = 1, 2, \ldots, m$, *sampled independently and uniformly from the unit sphere. Let*

$$\text{angle}(\mathbf{x}^0, \hat{\mathbf{x}}) = \arccos\left(\frac{\langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle_\Re}{\|\mathbf{x}^0\|_2 \|\hat{\mathbf{x}}\|_2}\right)$$

be the angle between the true vector $\mathbf{x}^0$ *and the "guess"* $\hat{\mathbf{x}}$, *and define the constant*

$$\alpha = 1 - \frac{2}{\pi} \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$$

that measures the accuracy of our guess. If $\mathcal{H} = \mathbb{C}$, *then whenever* $\alpha m > 4n - 1$, *the probability that PhaseMax recovers the true signal* $\mathbf{x}^0$ *is at least*

$$1 - \exp\left(-\frac{(\alpha m - 4n)^2}{4m}\right). \tag{2}$$

Similar results for the real case are derived below.

5. Formulation using Geometric Probability

Conditions for which PhaseMax enables exact recovery can be formulated as a classical problem from geometric probability involving random hemispheres, or "caps." We begin with a few simple definitions. Recall that we use $\mathcal{S}_\mathcal{H}^{n-1}$ to denote the unit sphere embedded in $\mathcal{H}^n$. Given a vector $\mathbf{a} \in \mathcal{S}_\mathcal{H}^{n-1}$, the *hemisphere cap* centered at $\mathbf{a}$ is

$$\mathcal{C}_\mathcal{H}(\mathbf{a}) = \{\boldsymbol{\delta} \in \mathcal{S}_\mathcal{H}^{n-1} \,|\, \langle \mathbf{a}, \boldsymbol{\delta} \rangle_\Re \geq 0\}. \tag{3}$$

This cap contains all vectors that form an acute angle with $\mathbf{a}$.

We also need the concept of *aligned vectors*. A complex vector $\mathbf{a} \in \mathcal{S}_\mathcal{H}^{n-1}$ is said to be aligned with $\mathbf{x} \in \mathcal{S}_\mathcal{H}^{n-1}$ if $\langle \mathbf{x}, \mathbf{a} \rangle \in \mathbb{R}_0^+$. In words, two vectors are aligned if their inner product is real valued and non-negative. Given a vector $\mathbf{x}$ and a measurement vector $\mathbf{a}$, we have $|\langle \mathbf{a}_i, \mathbf{x} \rangle| = |\langle \omega \mathbf{a}_i, \mathbf{x}^0 \rangle|$ for any unit-magnitude $\omega \in \mathbb{C}$. For this reason, we will often consider the set $\{\tilde{\mathbf{a}}_i\} = \{\text{sign}(\langle \mathbf{a}_i, \mathbf{x}^0 \rangle) \mathbf{a}_i\}$ of measurement vectors aligned with x^0 without loss of generality. Note that replacing $\{\mathbf{a}_i\}$ with $\{\tilde{\mathbf{a}}_i\}$ in PhaseMax does not change the feasible set, and thus does not impact the recovered solution. Finally, observe that a solution $\mathbf{x}^\star$ to (PM) must be aligned with $\hat{\mathbf{x}}$. This is because $\langle \omega \mathbf{x}^\star, \hat{\mathbf{x}} \rangle = |\langle \mathbf{x}^\star, \hat{\mathbf{x}} \rangle|$ when $\omega = \text{sign}(\langle \mathbf{x}^\star, \hat{\mathbf{x}} \rangle)$. It follows that $\mathbf{x}^\star$ cannot be optimal unless $\text{sign}(\langle \mathbf{x}^\star, \hat{\mathbf{x}} \rangle) = 1$.

We are now ready to formulate the following exact recovery condition for PhaseMax.

Theorem 2. *Consider the recovery of a vector* $\mathbf{x}^0$ *using PhaseMax with guess* $\hat{\mathbf{x}}$. *Assume, without loss of generality, that* $\hat{\mathbf{x}}$ *and the measurement ensemble* $\{\mathbf{a}_i\}$ *are aligned with* $\mathbf{x}^0$. *Let* $\mathcal{D} = \{\boldsymbol{\delta} \,|\, \langle \boldsymbol{\delta}, \hat{\mathbf{x}} \rangle \in \mathbb{R}_0^+\}$ *be the set of unit vectors aligned with* $\hat{\mathbf{x}}$. *Then,* $\mathbf{x}^0$ *is the unique solution of PhaseMax if*

$$\mathcal{D} \subset \bigcup_{i=1}^m \mathcal{C}_\mathcal{H}(\mathbf{a}_i). \tag{4}$$

Proof. Suppose the conditions of this theorem hold, and let $\mathbf{x}^\star$ be a solution to (PM). Since $\mathbf{x}^0$ is in the feasible set, $\mathbf{x}^\star$ must produce an objective value at least as large as $\mathbf{x}^0$, and so $\langle \mathbf{x}^\star, \hat{\mathbf{x}} \rangle \geq \langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle$. We know that $\mathbf{x}^\star$ is aligned with $\hat{\mathbf{x}}$. Since $\mathbf{x}^0$ is assumed to be aligned with $\hat{\mathbf{x}}$, the vector $\boldsymbol{\Delta} = \mathbf{x}^\star - \mathbf{x}^0$ is also aligned with $\hat{\mathbf{x}}$, and satisfies

$$\langle \boldsymbol{\Delta}, \hat{\mathbf{x}} \rangle = \langle \mathbf{x}^\star, \hat{\mathbf{x}} \rangle - \langle \mathbf{x}^0, \hat{\mathbf{x}} \rangle \geq 0.$$

Since $\mathbf{x}^\star$ is a feasible solution for (PM), we have

$$|\langle \mathbf{a}_i, \mathbf{x}^0 + \boldsymbol{\Delta} \rangle|^2 = |\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|^2 +$$
$$2[\langle \mathbf{a}_i, \mathbf{x}^0 \rangle^* \langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle]_\Re + |\langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle|^2 \leq b_i^2, \ \forall i. \quad (5)$$

Now, recall that $|\langle \mathbf{a}_i, \mathbf{x}^0 \rangle|^2 = b_i^2$, and $\mathbf{a}_i$ is aligned with $\mathbf{x}^0$. Hence, we get

$$\langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle_\Re \leq -\frac{1}{2}|\langle \mathbf{a}_i, \boldsymbol{\Delta} \rangle|^2 \leq 0, \ \forall i. \quad (6)$$

If $\|\boldsymbol{\Delta}\|_2 > 0$, then we see from (6) that the unit-length vector $\boldsymbol{\delta} = \boldsymbol{\Delta}/\|\boldsymbol{\Delta}\|_2$ satisfies $\boldsymbol{\delta} \notin \mathcal{C}(\mathbf{a}_i), \forall i$, which contradicts the covering condition (4). It follows that $\|\boldsymbol{\Delta}\|_2 = 0$ and $\mathbf{x}^\star = \mathbf{x}^0$. $\qquad \square$

6. Sphere Covering Results

Theorem 2 states that exact recovery of $\mathbf{x}^0$ occurs when the measurement ensemble $\{\mathbf{a}_i\}$ is aligned with $\mathbf{x}^0$, and the set $\mathcal{D}$ is covered by the caps $\{\mathcal{C}_\mathcal{H}(\mathbf{a}_i)\}$. We now study the probability that this condition holds. To do so, we need a few elementary sphere covering results. Our proof builds on the following simple result.

Lemma 1. *Suppose we slice the sphere $\mathcal{S}_\mathbb{R}^{n-1} \subset \mathbb{R}^n$ using k planes through the origin. Then we divide the sphere into as most*

$$r(n, k) = 2 \sum_{i=0}^{n-1} \binom{k-1}{i} \quad (7)$$

regions.

Proof. The proof is by induction. As a base case, we have $r(n, 1) = 2$ and $r(2, k) = 2k$ for $k \geq 1$. Now suppose we have a sphere $\mathcal{S}_\mathbb{R}^{n-1}$ in n dimensions sliced by $k - 1$ planes into $r(n, k-1)$ "original" regions. Consider the effect of adding a kth plane, p_k. The number of new regions created is equal to the number of original regions that are intersected by p_k. To count the number of new regions, we project the $k - 1$ original normal vectors into p_k, and count the minimal number of regions formed inside p_k by the resulting projected planes, which is at most $r(n-1, k-1)$. Adding this to the number of original regions yields

$$r(n, k) = r(n, k-1) + r(n-1, k-1).$$

We leave it to the reader to verify that (7) satisfies this recurrence relation. $\qquad \square$

Lemma 1 appears to have been first proved by (Schläfli, 1953), and simple induction arguments can be found in (Wendel, 1962; Gilbert, 1965; Füredi, 1986).

Before we attack the problem of when (4) holds, we begin with a simpler question: "how often is a sphere covered by caps with centers chosen at random from the sphere?" This question has been studied in detail by Gilbert (Gilbert, 1965) in the case $n = 3$. For our purposes, we need to study the covering probability in the more general case when caps are only chosen from a *subset* of the sphere, and the sphere can have arbitrarily high dimension. We show below that calculating this probability is easy when the caps are chosen from a *symmetric* subset of the sphere. We say that the set $\mathcal{A}$ is symmetric if, for all $\mathbf{x} \in \mathcal{A}$, we also have $-\mathbf{x} \in \mathcal{A}$. A probability measure defined over $\mathcal{A}$ is symmetric if the measure of a is the same as $-a$ for every $a \subset \mathcal{A}$.

Lemma 2. *Consider some non-empty symmetric set $\mathcal{A} \subset \mathcal{S}_\mathbb{R}^{n-1}$. Choose some set of $m_\mathcal{A}$ measurements $\{\mathbf{a}_i\}_{i=1}^{m_\mathcal{A}}$ at random from $\mathcal{A}$ using a symmetric measure. Then, the caps $\{\mathcal{C}_\mathbb{R}(\mathbf{a}_i)\}$ cover the sphere $\mathcal{S}_\mathbb{R}^{n-1}$ with probability*

$$p_{cover}(m_\mathcal{A}, n) \geq 1 - \frac{1}{2^{m_\mathcal{A}-1}} \sum_{k=0}^{n-1} \binom{m_\mathcal{A}-1}{k}.$$

This is the probability of getting n or more heads when flipping $m_\mathcal{A} - 1$ fair coins.

Proof. Let $\{\mathbf{a}'_i\}$ be a collection of m i.i.d. vectors sampled from $\mathcal{A}$ using a symmetric measure. Define $\mathbf{a}_i = c_i \mathbf{a}'_i$, where $\{c_i\}$ are i.i.d. Bernoulli variables that take value $+1$ or -1 with probability $\frac{1}{2}$. Clearly, the random vectors $\{\mathbf{a}_i\}$ are i.i.d. with the same distribution as $\{\mathbf{a}'_i\}$.

Consider the set

$$\bigcap_i \mathcal{C}_\mathbb{R}(-\mathbf{a}_i) = \bigcap_i \mathcal{C}_\mathbb{R}(-c_i \mathbf{a}'_i), \quad (8)$$

which contains all points *not* covered by the caps $\{\mathcal{C}_\mathbb{R}(\mathbf{a}_i)\}$. The caps $\{\mathcal{C}_\mathbb{R}(\mathbf{a}_i)\}$ cover the sphere whenever the intersection (8) is empty. There are $2^{m_\mathcal{A}}$ such intersections that can be formed, one for each choice of the sequence $\{c_i\}$. Now, from Lemma 1, we know that the $m_\mathcal{A}$ random planes $\{\{\mathbf{x} \mid \langle \mathbf{a}_i, \mathbf{x} \rangle = 0\}\}$ divide the sphere into at most $r(n, m_\mathcal{A})$ non-empty regions. Each of these regions corresponds to the intersection (8) for one possible choice of $\{c_i\}$. Therefore, of the $2^{m_\mathcal{A}}$ possible intersections, at most $r(n, m_\mathcal{A})$ of them are non-empty. Since each intersection is equally likely, the probability of covering the sphere is at least

$$p_{\text{cover}}(m_\mathcal{A}, n) \geq 1 - \frac{r(n, m_\mathcal{A})}{2^{m_\mathcal{A}}}.$$

$\qquad \square$

Remark: It can be shown that the bound in Lemma 1 is tight/exact when the slicing planes are chosen from a continuous probability distribution. Similarly, the bound in Lemma 2 is exact when the set $\{\mathbf{a}_i\}$ is sampled from a continuous distribution over $\mathcal{A}$.

We are now ready to present our main result. Exact recovery theorems for PhaseMax will follow immediately from the following geometric theorem. The result considers the case where the measurement vectors are drawn from only one hemisphere.

Lemma 3. *Consider two vectors* $\mathbf{x}, \mathbf{y} \subset \mathcal{S}_{\mathbb{R}}^{n-1}$, *and the caps* $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$ *and* $\mathcal{C}_{\mathbb{R}}(\mathbf{y})$. *Let* $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}, \mathbf{y})$ *be a measure of the similarity between the vectors* $\mathbf{x}$ *and* $\mathbf{y}$. *Draw some collection* $\{\mathbf{a}_i \in \mathcal{C}_{\mathbb{R}}(\mathbf{x})\}_{i=1}^m$ *of* m *vectors uniformly from* $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$ *so that* $\alpha(m-1) > 2n$. *Then,*

$$\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcup_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)$$

holds with probability at least

$$p_{cover}(m, n; \mathbf{x}, \mathbf{y}) \geq 1 - \exp\left(-\frac{(\alpha m - \alpha - 2n)^2}{2m - 2}\right).$$

Proof. To simplify notation, we assume $\mathbf{y} = [1, 0, \ldots, 0]^T$. Because of rotational symmetries this does not change the generality of our proof. Consider the reflection of $\mathbf{x}$ over $\mathbf{y}$ given by $\tilde{\mathbf{x}} = [x_1, -x_2, \ldots, -x_n]^T$. Suppose we have some collection $\{\mathbf{a}_i\}$ independently and uniformly distributed on the entire sphere. Consider the collection of vectors

$$\mathbf{a}_i' = \begin{cases} \mathbf{a}_i, & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle \geq 0 \quad (9) \\ \mathbf{a}_i - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \mathbf{y}, & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle < 0, \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle < 0 \quad (10) \\ -\mathbf{a}_i, & \text{if } \langle \mathbf{a}_i, \mathbf{x} \rangle < 0, \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle \geq 0. \quad (11) \end{cases}$$

The mapping $\mathbf{a}_i \to \mathbf{a}_i'$ maps the half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle < 0\}$ onto the half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle > 0\}$ using a combination of reflections and translations (see Figure 1). This makes the mapping $\mathbf{a}_i \to \mathbf{a}_i'$ onto and (piecewise) isometric, and so $\{\mathbf{a}_i'\}$ will be uniformly distributed over the half sphere $\{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x}^0 \rangle > 0\}$ whenever $\{\mathbf{a}_i\}$ is independently and uniformly distributed over the entire sphere.

Consider the "hourglass" shaped, symmetric set

$$\mathcal{A} = \{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle \geq 0, \langle \mathbf{a}, \tilde{\mathbf{x}} \rangle \geq 0\}$$
$$\cup \{\mathbf{a} \mid \langle \mathbf{a}, \mathbf{x} \rangle \leq 0, \langle \mathbf{a}, \tilde{\mathbf{x}} \rangle \leq 0\}.$$

We claim that $\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcup_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$ whenever

$$\mathcal{S}_{\mathbb{R}}^{n-1} \subset \bigcup_{\mathbf{a}_i \in \mathcal{A}} \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i). \quad (12)$$

In words, if the caps defined by the subset of $\{\mathbf{a}_i\}$ in $\mathcal{A}$ cover the entire sphere, then the caps $\{\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')\}$ (which have
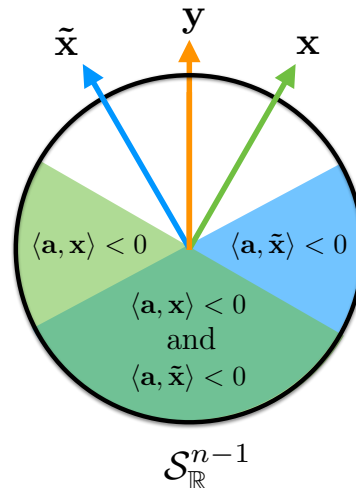

Figure 1. An illustration showing the construction of $\mathbf{x}$, $\mathbf{y}$, and $\tilde{\mathbf{x}}$ in the proof of Lemma 3.

centers in $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$) not only cover $\mathcal{C}_{\mathbb{R}}(\mathbf{x})$, but also cover the nearby cap $\mathcal{C}_{\mathbb{R}}(\mathbf{y})$. To justify this claim, suppose that (12) holds. Choose some $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(\mathbf{y})$. This point is covered by some cap $\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i)$ with $\mathbf{a}_i \in \mathcal{A}$. If $\langle \mathbf{a}_i, \mathbf{x} \rangle \geq 0$, then $\mathbf{a}_i = \mathbf{a}_i'$ and $\boldsymbol{\delta}$ is covered by $\mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$. If $\langle \mathbf{a}_i, \mathbf{x} \rangle < 0$, then

$$\langle \boldsymbol{\delta}, \mathbf{a}_i' \rangle = \langle \boldsymbol{\delta}, \mathbf{a}_i - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \mathbf{y} \rangle$$
$$= \langle \boldsymbol{\delta}, \mathbf{a}_i \rangle - 2\langle \mathbf{a}_i, \mathbf{y} \rangle \langle \boldsymbol{\delta}, \mathbf{y} \rangle \geq \langle \boldsymbol{\delta}, \mathbf{a}_i \rangle \geq 0.$$

Note we have used the fact that $\langle \boldsymbol{\delta}, \mathbf{y} \rangle$ is real-valued and non-negative because $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(y)$. We have also used $\langle \mathbf{a}_i, \mathbf{y} \rangle = [\mathbf{a}_i]_1 = \frac{1}{2}(\langle \mathbf{a}_i, \mathbf{x} \rangle + \langle \mathbf{a}_i, \tilde{\mathbf{x}} \rangle) < 0$, which follows from the definition of $\tilde{\mathbf{x}}$ and the definition of $\mathcal{A}$. Since $\langle \boldsymbol{\delta}, \mathbf{a}_i' \rangle \geq 0$, we have $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$, which proves our claim.

We can now see that the probability that $\mathcal{C}_{\mathbb{R}}(\mathbf{y}) \subset \bigcup_i \mathcal{C}_{\mathbb{R}}(\mathbf{a}_i')$ is at least as high as the probability that (12) holds. Let $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}})$ denote the probability of covering $\mathcal{C}(\mathbf{y})$ conditioned on the number $m_{\mathcal{A}}$ of points lying in $\mathcal{A}$. From Lemma 2, we know that $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}}) \geq p_{\text{cover}}(m_{\mathcal{A}}, n)$. As noted in Lemma 2, this is the chance of turning up n or more heads when flipping $m_{\mathcal{A}} - 1$ fair coins, which is one coin for every measurement $\mathbf{a}_i$ in $\mathcal{A}$. This probability $p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y})$ is

$$p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y}) = \mathbb{E}_{m_{\mathcal{A}}}[p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y} \mid m_{\mathcal{A}})]$$
$$\geq \mathbb{E}_{m_{\mathcal{A}}}[p_{\text{cover}}(m_{\mathcal{A}}, n)].$$

Let's evaluate this expectation. The region $\mathcal{A}$ is defined by two planes that intersect at an angle of $\beta = \operatorname{angle}(\mathbf{x}, \tilde{\mathbf{x}}) = 2\operatorname{angle}(\mathbf{x}, \mathbf{y})$. The probability of a random point $\mathbf{a}_i$ lying in $\mathcal{A}$ is given by $\alpha = \frac{2\pi - 2\beta}{2\pi} = 1 - \frac{2\beta}{\pi}$, which is the fraction of the unit sphere that lies either above or below both planes. The probability of a measurement $\mathbf{a}_i$ contributing to the heads count is half the probability of it lying in $\mathcal{A}$, or $\frac{1}{2}\alpha$.

The probability of turning up n or more heads is therefore given by

$$p_{\text{cover}}(m, n; \mathbf{x}, \mathbf{y}) = \tag{13}$$
$$1 - \sum_{k=0}^{n-1} \left(\frac{1}{2}\alpha\right)^k \left(1 - \frac{1}{2}\alpha\right)^{m-k-1} \binom{m-1}{k}.$$

Our final result is obtained by applying Hoeffding's inequality to (13). □

7. Recovery Guarantees for PhaseMax

Using the covering result of Lemma 3, together with the covering formulation of the exact recovery condition (4), proving exact recovery of PhaseMax is now rather straightforward. We begin with the real-valued case.

Theorem 3. *Consider the case of recovering a real-valued signal $\mathbf{x}^0 \in \mathbb{R}^n$ from m measurements of the form (1) with i.i.d. and uniform vectors $\{\mathbf{a}_i \in \mathcal{S}_{\mathbb{R}}^{n-1}\}$. PhaseMax recovers the true signal $\mathbf{x}^0$, with probability at least*

$$p_{\mathbb{R}}(m, n) \geq 1 - \exp\left(\frac{-(\alpha m - \alpha - 2n)^2}{2m - 2}\right),$$

where $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$ and $\alpha(m-1) > 2n$.

Proof. Consider the set of m independent and uniformly sampled measurements $\{\mathbf{a}_i \in \mathcal{S}_{\mathbb{R}}^{n-1}\}_{i=1}^m$. The aligned vectors $\{\tilde{\mathbf{a}}_i = \operatorname{phase}(\langle \mathbf{a}_i, \mathbf{x}^0\rangle)\mathbf{a}_i\}$ are uniformly distributed over the half sphere $\mathcal{C}_{\mathbb{R}}(\mathbf{x}^0)$. Exact reconstruction happens when the condition in Lemma 2 holds. To bound this probability, we invoke Lemma 3 with $\mathbf{x} = \mathbf{x}^0$ and $\mathbf{y} = \hat{\mathbf{x}}$. □

We have an analogous result in the complex case.

Theorem 4. *Consider the case of recovering a complex-valued signal $\mathbf{x}^0 \in \mathbb{R}^n$ from m measurements of the form (1) with i.i.d. and uniform vectors $\{\mathbf{a}_i \in \mathcal{S}_{\mathbb{R}}^{n-1}\}$. PhaseMax recovers the true signal $\mathbf{x}^0$, with probability at least*

$$p_{\mathbb{C}}(m, n) \geq 1 - \exp\left(\frac{-(\alpha m - 4n)^2}{4m}\right),$$

where $\alpha = 1 - \frac{2}{\pi} \operatorname{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$ and $\alpha m > 4n + 1$.

Proof. Let $\{\tilde{\mathbf{a}}_i\} = \{\operatorname{phase}(\langle \mathbf{a}_i, \mathbf{x}^0\rangle)\mathbf{a}_i\}$ be aligned measurement vectors. Define the half sphere of aligned ascent directions

$$\mathcal{D}_{\mathbb{C}} = \{\boldsymbol{\delta} \in \mathcal{S}_{\mathbb{C}}^{n-1} | \langle \boldsymbol{\delta}, \hat{\mathbf{x}}\rangle \in \mathbb{R}_0^+\}.$$

By Lemma 2, PhaseMax recovers $\mathbf{x}^0$ if

$$\mathcal{D}_{\mathbb{C}} \subset \bigcup_i \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i). \tag{14}$$

Let us bound the probability of this event. Consider the set $\mathcal{A} = \{\boldsymbol{\delta} \,|\, \langle \boldsymbol{\delta}, \mathbf{x}^0\rangle_{\Im} = 0\}$. We now claim that (14) holds whenever

$$\mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A} \subset \bigcup_i \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i). \tag{15}$$

To prove this claim, consider some $\boldsymbol{\delta} \in \mathcal{D}_{\mathbb{C}}$. To keep notation light, we will assume without loss of generality that $\|\mathbf{x}^0\|_2 = 1$. Form the vector $\boldsymbol{\delta}' = \boldsymbol{\delta} + j\langle \boldsymbol{\delta}, \mathbf{x}^0\rangle_{\Im} \mathbf{x}^0$, which is the projection of $\boldsymbol{\delta}$ onto $\mathcal{A}$. Suppose now that (15) holds. Since $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A}$, there is some i with $\boldsymbol{\delta}' \in \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i)$. But then

$$0 \leq \langle \boldsymbol{\delta}', \tilde{\mathbf{a}}_i\rangle_{\Re} = \langle \boldsymbol{\delta}, \tilde{\mathbf{a}}_i\rangle_{\Re} + \langle j\langle \boldsymbol{\delta}, \mathbf{x}^0\rangle_{\Im} \mathbf{x}^0, \tilde{\mathbf{a}}_i\rangle_{\Re} \tag{16}$$
$$= \langle \boldsymbol{\delta}, \tilde{\mathbf{a}}_i\rangle_{\Re}. \tag{17}$$

We have used the fact that $\langle j\langle \boldsymbol{\delta}, \mathbf{x}^0\rangle_{\Im} \mathbf{x}^0, \tilde{\mathbf{a}}_i\rangle = j\langle \boldsymbol{\delta}, \mathbf{x}^0\rangle_{\Im}\langle \tilde{\mathbf{a}}_i, \mathbf{x}^0\rangle$, which is imaginary valued and thus has no real component. We see that $\boldsymbol{\delta} \in \mathcal{C}_{\mathbb{C}}(\tilde{\mathbf{a}}_i)$, and the claim is proved.

We now know that exact reconstruction happens whenever condition (15) holds. Note that the sphere $S_{\mathbb{C}}^{n-1}$ is isomorphic to $S_{\mathbb{R}}^{2n-1}$, and the set $\mathcal{A}$ is isomorphic to the sphere $S_{\mathbb{R}}^{2n-2}$. The aligned vectors $\{\tilde{\mathbf{a}}_i\}$ are uniformly distributed over a half sphere in $\mathcal{C}_{\mathbb{C}}(\mathbf{x}^0) \cap \mathcal{A}$, which is isomorphic to the upper half sphere in $S_{\mathbb{R}}^{2n-2}$. The probability of these vectors covering the cap $\mathcal{C}_{\mathbb{C}}(\hat{\mathbf{x}}) \cap \mathcal{A}$ is thus given by $p_{\text{cover}}(m, 2n - 1; \mathbf{x}^0, \hat{\mathbf{x}})$ from Lemma 3, which is at least

$$1 - \exp\left(\frac{-(\alpha m - \alpha - 4n + 2)^2}{4m - 4}\right).$$

We can simplify this by using the fact that $\alpha < 1$. We also throw away the small constants in the numerator and denominator, which weakens the bound very slightly but tidies up the result. □

8. Experiments and Discussion

8.1. Initialization Methods

A variety of methods exist for generating the initial guess $\mathbf{x}^0$. The simplest approach is to use a random vector; see (Goldstein & Studer, 2016) for a corresponding analysis and exact recovery proof in this case. A more powerful method is the truncated spectral initializer (Chen & Candès, 2015), which is a refinement of the method put forward in (Netrapalli et al., 2013). A detailed analysis of such initializers is provided in (Lu & Li, 2017). As proved in Prop. 8 of (Chen & Candès, 2015), for any $\delta < \sqrt{2}$, there is a constant c_0 such that, with probability exceeding $1 - \exp(-c_0 m)$, a unit-length version of the approximation vector $\hat{\mathbf{x}}$ computed by the truncated spectral initializer satisfies $1 - \frac{\delta^2}{2} \leq |\langle \mathbf{x}^0, \hat{\mathbf{x}}\rangle|$, provided that $m > c_1 n$ for

Table 1. Comparison of theoretical recovery guarantees for noiseless phase retrieval.

Algorithm	Sample complexity	Lower bound on $p_\mathbb{C}(m,n)$
PhaseMax	$m > (4n-1)/\alpha + 1$	$1 - e^{-(\alpha(m-1)-4n+2)^2/(2m-2)}$
PhaseLift (Candès & Li, 2014)	$m \geq c_0 n$	$1 - c_1 e^{-c_2 m}$
TWF (Candès & Li, 2014)	$m \geq c_0 n$	$1 - c_1 e^{-c_2 m}$
TAF (Wang et al., 2016)	$m \geq c_0 n$	$1 - (m+5)e^{-n/2} - c_1 e^{-c_2 m} - 1/n^2$
(Bahmani & Romberg, 2016)	$m > \frac{32}{\sin^4(\alpha)} \log\left(\frac{8e}{\sin^4(\alpha)}\right)n$	$1 - 8e^{-\sin^4(\alpha)\left(M - \frac{32}{\sin^4(\alpha)} \log\left(\frac{8e}{\sin^4(\alpha)}\right)N\right)/16}$
(Hand & Voroninski, 2016)[a]	$m > c_0 n$	$1 - 6e^{-c_2 m}$

_____

[a]The theory only guarantees recovery for real-valued measurements.

some constant $c_1 > 0$. This implies that the approximation accuracy satisfies

$$\alpha = 1 - \frac{2}{\pi} \, \text{angle}(\mathbf{x}^0, \hat{\mathbf{x}})$$
$$\geq 1 - \frac{2}{\pi} \arccos\left(1 - \frac{\delta^2}{2}\right) > 0 \qquad (18)$$

with high probability.

This motivates the following process that recovers $\mathbf{x}^0$ using a number of measurements that grows linearly with n. First, choose some $\delta < \sqrt{2}$ and calculate α from (18). Next, generate a spectral initializer $\hat{\mathbf{x}}$ using c_1 random Gaussian measurements. This initializer has accuracy α with high probability. Finally, using $5n/\alpha$ (the constant in this expression must be larger than 4 to guarantee recovery with high probability) additional random Gaussian measurements, recover the vector $\mathbf{x}^0$ using PhaseMax. This recovery step succeeds with high probability when n is large. This process recovers $\mathbf{x}^0$ using $(5/\alpha + c_1)n$ measurements, which is linear in n.

A simpler recovery approach only samples $\max\{5/\alpha, c_1\}n$ measurements, and then uses the same set of measurements for both the initialization and recovery. This process works well in practice, and is used in the experiments below. Technically, our results assume the measurements $\{\mathbf{a}_i\}$ are independent of $\hat{\mathbf{x}}$, and so our theory does not formally guarantee convergence in this case. For an analysis that considers the case where $\{\mathbf{a}_i\}$ and $\hat{\mathbf{x}}$ are dependent, we refer to the analysis in (Bahmani & Romberg, 2016).

8.2. Comparison to Other Phase Retrieval Methods

We compare PhaseMax with other recovery methods, including lifted convex methods and non-convex approaches. Table 1 lists the sample complexity (measurements needed to achieve exact recovery with 0.5 probability) of various phase retrieval methods as a function of the number of unknowns n. We also list the probability of reconstruction

from m measurements. We see that PhaseMax requires the same sample complexity ($O(n)$) as compared to PhaseLift, TWF, and TAW, when used together with the truncated spectral initializer proposed in (Candès & Li, 2014).

The recovery bounds for all other methods require unspecified constants (c_i in Table 1) that are generally extremely large and require a lower bound on the initialization accuracy. In contrast, the bounds for PhaseMax contain no unspecified constants, explicitly depend on the approximation factor α, and our new analytical approach yields extremely sharp bounds (see below for the details).

We also compare in Table 1 to a different analysis of the PhaseMax formulation by (Bahmani & Romberg, 2016) that appeared shortly before our own, and to a later analysis by (Hand & Voroninski, 2016). By using methods from machine learning theory, (Bahmani & Romberg, 2016) produce exact reconstruction bounds that, for a specified value of α, are uniform with respect to the initialization $\hat{\mathbf{x}}$, and thus guarantee exact signal recovery in the case that $\hat{\mathbf{x}}$ is dependent on the measurement vectors. The analysis presented here is weaker in the sense that is does not have this uniformity property, but stronger in the sense that it produces tighter bounds without unspecified constants. The bounds by (Hand & Voroninski, 2016) require unspecified constants, but the authors show that PhaseMax can be analyzed using standard concentration of measure arguments.

8.3. Tightness of Guarantees

By using strong concentration bounds for the unit sphere, our analysis produces sharp recovery guarantees that lie close to behavior observed in practice. In Figure 2, we use random Gaussian test problems and the accelerated gradient-based solver described in (Goldstein et al., 2014) to plot the empirical and theoretical probabilities of exact signal recovery for $n = 100$ and $n = 500$ measurements while varying the accuracy $\beta = \text{angle}(\hat{\mathbf{x}}, \mathbf{x}^0)$ of the initial guess. We declared exact recovery when the relative error of the
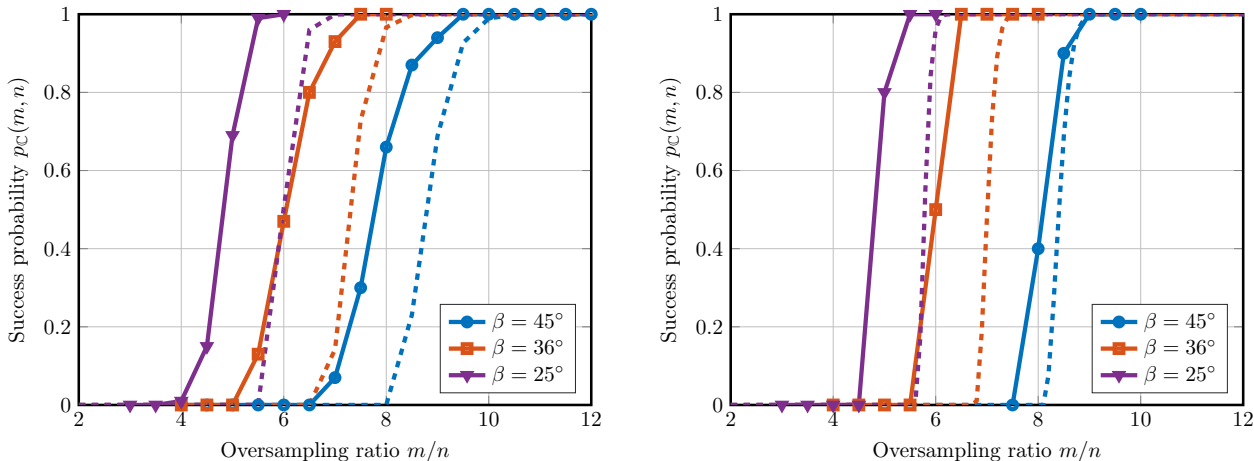
Figure 2. Comparison between the empirical success probability (solid lines) and our theoretical lower bound (dashed lines) for varying angles β between the true signal and the approximation vector. Our theoretical results accurately characterize the empirical success probability of PhaseMax. Furthermore, PhaseMax exhibits a sharp phase transition for larger dimensions.

recovered signal fell below 10^{-5}.

Our theoretical bounds tend to agree fairly closely with observations, and generally require fewer than 50% more measurements than is needed in practice. We also observe a sharp phase transition between inaccurate and accurate recovery, as predicted by our theory.

8.4. Performance Limits of PhaseMax

To compare PhaseMax to other phase retrieval methods, we observe the accuracy of signal reconstruction as a function of the number of measurements. We emphasize that this comparison is only done in the random Gaussian problem setting, and results may differ with different types of signal, measurement, and noise models. The sole purpose of this experiment is to explore the efficacy and limits of PhaseMax, and to test the tightness of the predicted recovery guarantees. For an extensive comparison between existing methods, see (Waldspurger et al., 2015; Jaganathan et al., 2015)). We compare the Gerchberg-Saxton algorithm (Gerchberg & Saxton, 1972), the Fienup algorithm (Fienup, 1982), the truncated Wirtinger flow (Chen & Candès, 2015), and PhaseMax. All methods were initialized using the truncated spectral initializer (Chen & Candès, 2015).

We also run simulations using the semidefinite relaxation method PhaseLift (Candès et al., 2013) implemented using a proximal gradient solver. PhaseLift, and its equivalent dual formulation PhaseCut (Waldspurger et al., 2015), is the only convex alternative to PhaseMax. However, unlike Phase-Max, PhaseLift/PhaseCut "lifts" the problem to a higher dimension and squares the number of unknowns.

Figure 3 reveals that PhaseMax requires larger oversampling ratios m/n to enable faithful signal recovery compared to non-convex phase-retrieval algorithms that operate in the original signal dimension. This is because the truncated spectral initializer requires oversampling ratios of about six or higher to yield sufficiently accurate approximation vectors $\hat{\mathbf{x}}$ that enable PhaseMax to succeed. While PhaseMax does not achieve exact reconstruction with the lowest number of measurements, it is convex, operates in the original signal dimension, can be implemented via solvers for Basis Pursuit, and comes with *sharp* performance guarantees that do not sweep constants under the rug (cf. Figure 2).

The convexity of PhaseMax enables a natural extension to sparse phase retrieval (Jaganathan et al., 2013; Shechtman et al., 2014) or other signal priors (e.g., total variation, group sparsity, or bounded infinity norm) that can be formulated with convex functions. Such non-differentiable priors cannot be efficiently minimized using simple gradient descent methods (which form the basis of Wirtinger or amplitude flow, and many other methods), but can potentially be solved using standard convex solvers when combined with the PhaseMax formulation.

9. Conclusions

We have proposed a convex relaxation for phase retrieval problems called PhaseMax that does not require lifting. Using methods from geometric probability, we have provided tight bounds on the probability of correct signal recovery.

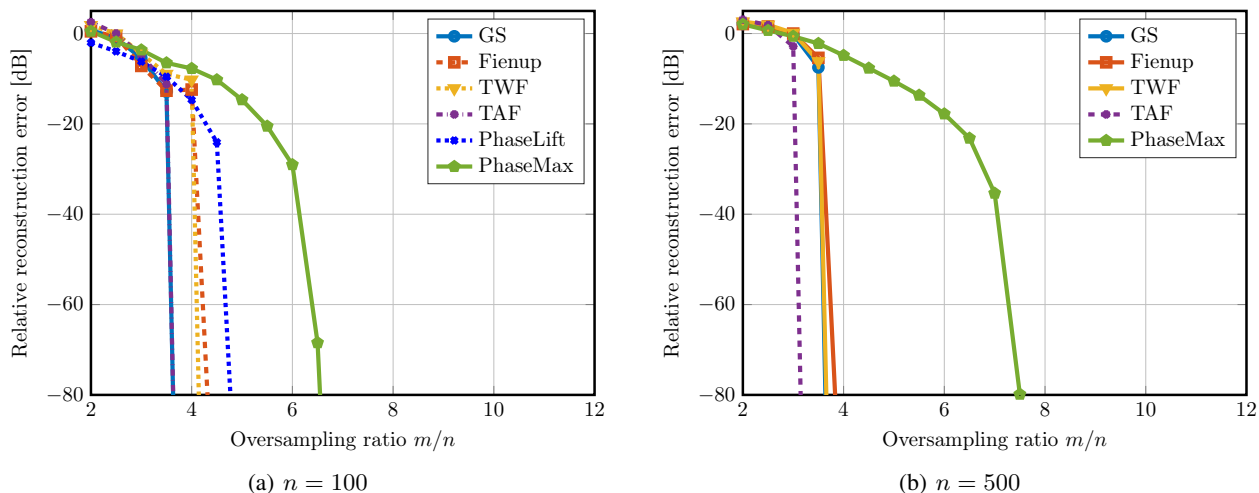The proposed problem and its analysis also represents a

(a) $n = 100$

(b) $n = 500$

Figure 3. Comparison of the relative reconstruction error. We use the truncated spectral initializer for Gerchberg-Saxton (GS), Fienup, truncated Wirtinger flow (TWF), truncated amplitude flow (TAF), and PhaseMax. PhaseMax does not achieve exact recovery for the lowest number of measurements among the considered methods, but is convex, operates in the original dimension, and comes with sharp performance guarantees. PhaseLift only terminates in reasonable computation time for $n = 100$.

radical departure, both in theory and algorithms, from conventional methods for convex or semidefinite relaxation. By providing a convex relaxation for phase retrieval in the native parameter space, our approach opens the door for using a broad range of convex optimization routines, regularizers, and priors to solve phase retrieval or related problems in machine learning, computer vision, or signal processing.

Finally, the new analytical methods used in this paper have recently been used to prove tight reconstruction bounds for bi-convex problems outside the field of phase retrieval (Aghasi et al., 2017), and may be broadly applicable to a wide range of signal processing problems.

Acknowledgments

References

Aghasi, A., Ahmed, A., and Hand, P. Branchhull: Convex bilinear inversion from the entrywise product of signals with known signs. *arXiv preprint :1702.04342*, Feb. 2017.

Bahmani, S. and Romberg, J. Phase retrieval meets statisti-cal learning theory: A flexible convex relaxation. *arXiv preprint: 1610.04210*, Oct. 2016.

Candès, E. J. and Li, X. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Found. Comput. Math.*, 14(5):1017–1026, Oct. 2014.

Candès, E. J., Strohmer, T., and Voroninski, V. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pure Appl. Math.*, 66(8):1241–1274, 2013.

Candès, J. E., Li, X., and Soltanolkotabi, M. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory*, 61(4):1985–2007, Feb. 2015.

Chen, Y. and Candès, E. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Adv. Neural Inf. Process. Syst.*, pp. 739–747, 2015.

Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, Aug. 1982.

Füredi, Z. Random polytopes in the d-dimensional cube. *Disc. Comput. Geom.*, 1(4):315–319, Dec. 1986.

Gerchberg, R. W. and Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, Aug. 1972.

Gilbert, E. N. The probability of covering a sphere with n circular caps. *Biometrika*, 52(3/4):323–330, Dec. 1965.

Goldstein, T. and Studer, C. PhaseMax: convex phase retrieval via basis pursuit. *arXiv preprint: 1610.07531*, Oct. 2016.

Goldstein, T., Studer, C., and Baraniuk, R. A field guide to forward-backward splitting with a FASTA implementation. *arXiv preprint: 1411.3406*, Feb. 2014.

Hand, P. and Voroninski, V. An elementary proof of convex phase retrieval in the natural parameter space via the linear program PhaseMax. *arXiv preprint: 1611.03935*, Nov. 2016.

Jaganathan, K., Oymak, S., and Hassibi, B. Sparse phase retrieval: Convex algorithms and limitations. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1022–1026, Jul. 2013.

Jaganathan, K., Eldar, Y. C., and Hassibi, B. Phase retrieval: An overview of recent developments. *arXiv:1510.07713*, Oct. 2015.

Lu, Y. M. and Li, G. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *arXiv preprint: 1702.06435*, 2017.

Netrapalli, P., Jain, P., and Sanghavi, S. Phase retrieval using alternating minimization. In *Adv. Neural Inf. Process. Syst.*, pp. 2796–2804, 2013.

Schläfli, L. *Gesammelte Mathematische Abhandlungen I*. Springer Basel, 1953.

Schniter, P. and Rangan, S. Compressive phase retrieval via generalized approximate message passing. *IEEE Trans. Sig. Process.*, 63(4):1043–1055, Feb. 2015.

Shechtman, Y., Beck, A., and Eldar, Y. C. GESPAR: efficient phase retrieval of sparse signals. *IEEE Trans. Sig. Process.*, 62(4):928–938, Jan. 2014.

Studer, C., Goldstein, T., Yin, W., and Baraniuk, R. G. Democratic representations. *arXiv preprint: 1401.3420*, Jan. 2014.

Waldspurger, I., d'Aspremont, A., and Mallat, S. Phase recovery, maxcut and complex semidefinite programming. *Math. Prog.*, 149(1-2):47–81, Feb. 2015.

Wang, G., Giannakis, G. B., and Eldar, Y. C. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv: 1605.08285*, Jul. 2016.

Wendel, J. G. A problem in geometric probability. *Math. Scand.*, 11:109–111, 1962.