PHASELIFTOFF: AN ACCURATE AND STABLE PHASE RETRIEVAL METHOD BASED ON DIFFERENCE OF TRACE AND FROBENIUS NORMS[*]

PENGHANG YIN[†] AND JACK XIN[‡]

Dedicated to George Papanicolaou in honor of his 70th birthday

Abstract. Phase retrieval aims to recover a signal $x \in \mathbb{C}^n$ from its amplitude measurements $|\langle x, a_i \rangle|^2$, $i = 1, 2, \cdots, m$, where a_i's are over-complete basis vectors, with m at least $3n - 2$ to ensure a unique solution up to a constant phase factor. The quadratic measurement becomes linear in terms of the rank-one matrix $X = xx^*$. Phase retrieval is then a rank-one minimization problem subject to linear constraint for which a convex relaxation based on trace-norm minimization (PhaseLift) has been extensively studied recently. At $m = O(n)$, PhaseLift recovers with high probability the rank-one solution. In this paper, we present a precise proxy of the rank-one condition via the difference of trace and Frobenius norms which we call PhaseLiftOff. The associated least squares minimization with this penalty as regularization is equivalent to the rank-one least squares problem under a mild condition on the measurement noise. Stable recovery error estimates are valid at $m = O(n)$ with high probability. Computation of PhaseLiftOff minimization is carried out by a convergent difference of convex functions algorithm. In our numerical example, a_i's are Gaussian distributed. Numerical results show that PhaseLiftOff outperforms PhaseLift and its nonconvex variant (log-determinant regularization), and successfully recovers signals near the theoretical lower limit on the number of measurements without the noise.

Key words. Phase retrieval, phase lift regularized by trace minus Frobenius norms, equivalent minimization problems, Karush-Kuhn-Tucker conditions, difference of convex functions algorithm, convergence.

AMS subject classifications. 94A12, 65K05, 90C22.

1. Introduction

Phase retrieval has been a long standing problem in imaging sciences such as X-ray crystallography, electron microscopy, array imaging, optics, and signal processing, [20, 18, 19, 22, 23] among others. It is concerned with signal recovery when only the amplitude measurements (say of its Fourier transform) are available. Major recent advances have been made for phase retrieval by formulating it as a matrix completion and rank one minimization problem (PhaseLift) which is relaxed and solved as a convex trace (nuclear) norm minimization problem under sufficient measurement conditions [8, 10, 7, 6]; see also [2, 3] for related work. An alternative viable approach makes use of random masks in measurements to achieve uniqueness of solution with high probability [13, 14].

In this paper, we study a nonconvex Lipschitz continuous metric, the difference of trace and Frobenius norms, and show that its minimization characterizes the rank one solution exactly and serves as a new tool to solve the phase retrieval problem. We shall see that it is more accurate than trace norm or the heuristic log-determinant [15, 16], and performs the best when the number of measurements approaches the theoretical lower limit [2]. We shall call our method PhaseLiftOff, where Off is short for subtracting off Frobenius norm from the trace norm in PhaseLift [8, 6].

The phase retrieval problem aims to reconstruct an unknown signal $\hat{x} \in \mathbb{C}^n$ satisfying

m quadratic constraints

$$|\langle a_i, \hat{x} \rangle|^2 = b_i, \quad i = 1, \ldots, m,$$

where the bracket is inner product, $a_i \in \mathbb{C}^n$, and $b_i \in \mathbb{R}$. Letting $X = xx^* \in \mathbb{C}^{n \times n}$ be a rank-1 positive semidefinite matrix ($*$ is conjugate transpose), one can recast quadratic measurements as linear ones about X:

$$|\langle a_i, x \rangle|^2 = a_i^* X a_i, \quad i = 1, \ldots, m.$$

Thus we can define a linear operator $\mathcal{A}$ uniquely determined by the measurement matrix $A = (a_1, \ldots, a_m) \in \mathbb{C}^{n \times m}$:

$$\mathbb{H}^{n \times n} \to \mathbb{R}^m$$
$$X \quad \mapsto \mathrm{diag}(A^* X A),$$

which maps Hermitian matrices into real-valued vectors. Denote $\hat{x}\hat{x}^*$ by $\hat{X}$, and suppose $b = (b_1, \ldots, b_m)^{\mathrm{T}} = \mathcal{A}(\hat{X}) \in \mathbb{R}^m$ is the measurement vector. Then the phase retrieval becomes the feasibility problem, being equivalent to a rank minimization problem:

$$
\begin{array}{ll}
\text{find } X \in \mathbb{C}^{n \times n} & \\
\text{s.t. } \mathcal{A}(X) = b & \quad \min_{X \in \mathbb{C}^{n \times n}} \mathrm{rank}(X) \\
\quad X \succeq 0 & \Leftrightarrow \text{s.t.} \quad \mathcal{A}(X) = b \\
\quad \mathrm{rank}(X) = 1. & \quad\quad\quad\quad X \succeq 0.
\end{array}
\qquad (1.1)
$$

To arrive at the original solution $\hat{x}$ to the phase retrieval problem, one needs to factorize the solution $\hat{X}$ of (1.1) as $\hat{x}\hat{x}^*$. It gives $\hat{x}$ up to multiplication by a constant scalar with unit modulus (a constant phase factor), because if $\hat{x}$ solves the phase retrieval problem, so does $c\hat{x}$, for any $c \in \mathbb{C}$ with $|c| = 1$. At least $3n - 2$ intensity measurements are necessary to guarantee uniqueness (up to a constant phase factor) of the solution to (1.1) [17], whereas $4n - 2$ generic measurements suffice for uniqueness with probability one [2].

Instead of (1.1), Candès *et al.* [6, 8] suggest solving the convex PhaseLift problem, namely minimizing the trace norm as a convex surrogate for the rank functional:

$$\min_{X \in \mathbb{C}^{n \times n}} \mathrm{Tr}(X) \quad \text{s.t.} \quad \mathcal{A}(X) = b, X \succeq 0.$$

It is shown in [7] that if each a_i is Gaussian or uniformly sampled on the sphere, then with high probability, $m = O(n)$ measurements are sufficient to recover the ground truth $\hat{X}$ via PhaseLift. For the noisy case, the following variant is considered in [7]:

$$\min_{X \in \mathbb{C}^{n \times n}} \|\mathcal{A}(X) - b\|_1 \quad \text{s.t.} \quad X \succeq 0.$$

In this case, $b = \mathcal{A}(\hat{X}) + e$ is contaminated by the additive noise $e \in \mathbb{R}^m$. Similarly, $m = O(n)$ measurements guarantee stable recovery in the sense that the solution X^{opt} satisfies $\|X^{\mathrm{opt}} - \hat{X}\|_F = O(\frac{\|e\|_1}{m})$ with probability close to 1. On the computational side, the regularized trace-norm minimization is considered in [6, 8]:

$$\min_{X \in \mathbb{C}^{n \times n}} \frac{1}{2} \|\mathcal{A}(X) - b\|_2^2 + \lambda \mathrm{Tr}(X) \quad \text{s.t.} \quad X \succeq 0. \qquad (1.2)$$

If there is no noise, a tiny value of λ would work well. However, when the measurements are noisy, determining λ requires extra work, such as employing the cross validation technique.

Besides PhaseLift and its nonconvex variant (log-determinant) proposed in [6], related formulations such as feasibility problem or weak PhaseLift [11] and PhaseCut [26] also lead to phase retrieval solutions under certain measurement conditions. PhaseCut is a convex relaxation where trace minimization is in the form $\min_U \text{Tr}(UM)$, where M (resp., U) is a known (resp., unknown) positive semidefinite Hermitian matrix, and $\text{diag}(U) = 1$. The exact recovery (tightness) conditions for PhaseLift and PhaseCut are studied in [26] and references therein.

From the point of view of energy minimization, the phase retrieval problem is simply:

$$\min_{X \in \mathbb{C}^{n \times n}} \|\mathcal{A}(X) - b\|_2^2 \quad \text{s.t.} \quad X \succeq 0, \text{rank}(X) = 1. \tag{1.3}$$

This is a least squares-type model applicable to both noiseless and noisy cases. Our main contribution in this work is to reformulate the phase retrieval problem (1.3) as a nearly equivalent nonconvex optimization problem that can be efficiently solved by the so-called difference of convex functions algorithm (DCA). Specifically, we propose to solve the following regularization problem:

$$\min_{X \in \mathbb{C}^{n \times n}} \varphi(X) := \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \lambda(\text{Tr}(X) - \|X\|_F) \quad \text{s.t.} \quad X \succeq 0. \tag{1.4}$$

Recently the authors of [12, 27] have reported that minimizing the difference of ℓ_1 and ℓ_2 norms would promote sparsity when recovering a sparse vector from linear measurements. The $\ell_1 - \ell_2$ minimization is extremely favorable for the reconstruction of the 1-sparse vector x because $\|x\|_1 - \|x\|_2$ attains the possible minimum value zero at such x. Note that when $X \succeq 0$, $\text{Tr}(X)$ is nothing but the ℓ_1 norm of the vector $\sigma(X)$ formed by X's singular values and $\|X\|_F$ the ℓ_2 norm. Thus (1.4) is basically the counterpart of $\ell_1 - \ell_2$ minimization with nonnegativity constraint discussed in [12]. Similarly, $\text{Tr}(X) - \|X\|_F$ is minimized when $\sigma(X)$ is 1-sparse or equivalently $\text{rank}(X) = 1$.

The rest of the paper is organized as follows. After setting notations and giving preliminaries in Section 2, we establish the equivalence between (1.4) and (1.3) under mild conditions on λ and $\|e\|_2$ in Section 3. In particular, the equivalence holds in the absence of noise. We will see that λ plays a very different role in (1.4) from that in (1.2), as we have much more freedom to choose λ in (1.4). We then introduce the DCA method for solving (1.4) and analyze its convergence in Section 4. The DCA calls for solving a sequence of convex subproblems which we carry out with the alternating direction method of multipliers (ADMM). As an extension, we tailor our method to the task of retrieving real-valued or nonnegative signals. In Section 5, we show numerical results demonstrating the superiority of our method through examples where the columns of A are sampled from Gaussian distribution. The PhaseLiftOff problem (1.4) with the $\text{Tr}(X) - \|X\|_F$ regularization produces far more accurate phase retrieval than either the trace norm or $\log(\det(X + \varepsilon I))$ ($\varepsilon > 0$). We also observe that for a full interval of regularization parameters, the DCA produces robust solutions in the presence of noise. The concluding remarks are given in Section 6.

2. Notations and preliminaries

For any $X, Y \in \mathbb{C}^{n \times n}$, $\langle X, Y \rangle = \text{Tr}(X^*Y)$ is the inner product for matrices, which is a generalization of that for vectors. The Frobenius norm of X is $\|X\|_F = \sqrt{\langle X, X \rangle}$, while $X \circ Y$ denotes the entry-wise product, namely $(X \circ Y)_{ij} = X_{ij}Y_{ij}$, $\forall i, j$. $\text{diag}(X) \in \mathbb{C}^n$ extracts the diagonal elements of X. The spectral norm of X is $\|X\|_2$, while the nuclear norm of X is $\|X\|_*$. We have the following elementary inequalities:

$$\|X\|_2 \leq \|X\|_F \leq \sqrt{\text{rank}(X)}\|X\|_2,$$

and

$$\|X\|_F \leq \|X\|_* \leq \sqrt{\operatorname{rank}(X)}\|X\|_F.$$

For any vector $x \in \mathbb{R}^m$, $\|x\|_1$ and $\|x\|_2$ are the ℓ_1 norm and ℓ_2 norm respectively, while $\operatorname{Diag}(x) \in \mathbb{R}^{m \times m}$ is the diagonal matrix with x on its diagonal.

We assume that $m \geq n$ and that A is of full rank unless otherwise stated, i.e. $\operatorname{rank}(A) = n$. Recall that if $\mathcal{A}(X) := \operatorname{diag}(A^* X A)$ is a linear operator from $\mathbb{H}^{n \times n}$ to $\mathbb{R}^m$, then the adjoint operator $\mathcal{A}^*$ is defined as $\mathcal{A}^*(x) := A \operatorname{Diag}(x) A^* \in \mathbb{H}^{n \times n}$ for all $x \in \mathbb{R}^m$. Furthermore, the norms of $\mathcal{A}$ and $\mathcal{A}^*$ are given by

$$\|\mathcal{A}\| := \sup_{X \in \mathbb{H}^{n \times n} \setminus \{0\}} \frac{\|\mathcal{A}(X)\|_2}{\|X\|_F}, \quad \|\mathcal{A}^*\| := \sup_{x \in \mathbb{R}^m \setminus \{0\}} \frac{\|\mathcal{A}^*(x)\|_F}{\|x\|_2}.$$

Since $(\mathbb{H}^{n \times n}, \langle \cdot, \cdot \rangle)$ and $(\mathbb{R}^m, \langle \cdot, \cdot \rangle)$ are both Hilbert spaces, we have

$$\|\mathcal{A}^*\|^2 = \|\mathcal{A}\|^2 = \|\mathcal{A}\mathcal{A}^*\|. \tag{2.1}$$

The following lemma will be frequently used in the proofs.

LEMMA 2.1. *Suppose $X, Y \in \mathbb{C}^{n \times n}$ and $X, Y \succeq 0$, then*
  1. $\langle X, Y \rangle \geq 0$.
  2. $\langle X, Y \rangle = 0 \Leftrightarrow XY = 0$.
  3. $\|\mathcal{A}(X)\|_2 = 0 \Leftrightarrow X = 0$.

Proof.
  (1) Suppose $Y = U \Sigma U^*$ is the singular value decomposition (SVD), let $Y^{\frac{1}{2}} := U \Sigma^{\frac{1}{2}} U^* \succeq 0$, where the diagonal elements of $\Sigma^{\frac{1}{2}}$ are square roots of the singular values. Then we have $Y = Y^{\frac{1}{2}} Y^{\frac{1}{2}}$ and

$$\langle X, Y \rangle = \operatorname{Tr}(X^* Y) = \operatorname{Tr}(XY) = \operatorname{Tr}(Y^{\frac{1}{2}} X Y^{\frac{1}{2}}) \geq 0.$$

The last inequality holds because $Y^{\frac{1}{2}} X Y^{\frac{1}{2}} \succeq 0$.

  (2) "$\Rightarrow$" Further assume $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$, where $\Sigma_1 \succ 0$, and let $Z = U^* X U \succeq 0$. By (1), we have $\operatorname{Tr}(Y^{\frac{1}{2}} X Y^{\frac{1}{2}}) = \langle X, Y \rangle = 0$, thus $Y^{\frac{1}{2}} X Y^{\frac{1}{2}} = 0$. So

$$0 = \Sigma^{\frac{1}{2}} U^* X U \Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}} Z \Sigma^{\frac{1}{2}} = \begin{pmatrix} \Sigma_1^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{12}^* & Z_{22} \end{pmatrix} \begin{pmatrix} \Sigma_1^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \Sigma_1^{\frac{1}{2}} Z_{11} \Sigma_1^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix},$$

then we have $\Sigma_1^{\frac{1}{2}} Z_{11} \Sigma_1^{\frac{1}{2}} = 0$ and $Z_{11} = 0$. Next we want to show $Z_{12} = 0$. Suppose $Z_{12} \neq 0$, and let us consider $v_c = \begin{pmatrix} c Z_{12} w \\ w \end{pmatrix} \in \mathbb{C}^n$, where w is a fixed vector making $Z_{12} w$ nonzero and $c \in \mathbb{R}$. Then since $Z \succeq 0$, we have

$$0 \leq v_c^* Z v_c = (c w^* Z_{12}^*, w^*) \begin{pmatrix} 0 & Z_{12} \\ Z_{12}^* & Z_{22} \end{pmatrix} \begin{pmatrix} c Z_{12} w \\ w \end{pmatrix} = 2c|Z_{12} w|^2 + w^* Z_{22} w, \quad \forall c \in \mathbb{R}.$$

In above inequality, letting $c \to -\infty$ leads to a contradiction. Therefore $Z_{12} = 0$. A simple computation gives $U^* X U \Sigma = Z \Sigma = 0$, and thus $XY = X U \Sigma U^* = 0$.
  "$\Leftarrow$" If $XY = 0$, then $\langle X, Y \rangle = \operatorname{Tr}(X^* Y) = \operatorname{Tr}(XY) = 0$.

(3) "$\Rightarrow$" Let $X^{\frac{1}{2}} \succeq 0$ such that $X^{\frac{1}{2}} X^{\frac{1}{2}} = X$. Then

$$0 = \|\mathcal{A}(X)\|_2 = \|\mathrm{diag}(A^* X^{\frac{1}{2}} X^{\frac{1}{2}} A)\|_2.$$

So $\mathrm{diag}(A^* X^{\frac{1}{2}} X^{\frac{1}{2}} A) = 0$, and thus $0 = \mathrm{Tr}(A^* X^{\frac{1}{2}} X^{\frac{1}{2}} A) = \|A^* X^{\frac{1}{2}}\|_F^2$. This together with $\mathrm{rank}(A) = n \leq m$ implies $X^{\frac{1}{2}} = 0$.

"$\Leftarrow$" Trivial. $\qquad\square$

Karush-Kuhn-Tucker conditions. Let us consider a first-order stationary point $\tilde{X}$ of the minimization problem

$$\min_{X \in \mathbb{C}^{n \times n}} f(X) \quad \text{s.t.} \quad X \succeq 0.$$

Suppose f is differentiable at $\tilde{X}$. Then there exists $\tilde{\Lambda} \in \mathbb{C}^{n \times n}$ such that the following Karush-Kuhn-Tucker (KKT) optimality conditions hold:

- Stationarity: $\nabla f(\tilde{X}) = \tilde{\Lambda}$.
- Primal feasibility: $\tilde{X} \succeq 0$.
- Dual feasibility: $\tilde{\Lambda} \succeq 0$.
- Complementary slackness: $\tilde{X}\tilde{\Lambda} = 0$.

In order to make better use of the last condition, by Lemma 2.1 (2), we can express it as

- Complementary slackness: $\langle \tilde{X}, \tilde{\Lambda} \rangle = 0$.

3. Exact and stable recovery theory

In this section, we present the PhaseLiftOff theory for exact and stable recovery of complex signals.

3.1. Equivalence. We first develop mild conditions that guarantee the full equivalence between Phase Retrieval (1.3) and PhaseLiftOff (1.4).

THEOREM 3.1. *Let $\mathcal{A}$ be an arbitrary linear operator from $\mathbb{H}^{n \times n}$ to $\mathbb{R}^m$, and let $b = \mathcal{A}(\hat{X}) + e$. If $\|b\|_2 > \|e\|_2$ and $\lambda > \frac{\|\mathcal{A}\|\|e\|_2}{\sqrt{2}-1}$, suppose X^{opt} is a solution (global minimizer) to (1.4). Then $\mathrm{rank}(X^{\mathrm{opt}}) = 1$. Moreover, minimization problems (1.3) and (1.4) are equivalent in the sense that they share the same set of solutions.*

Proof. Let X^{opt} be a solution to (1.4). Since $\varphi(\hat{X}) = \frac{1}{2}\|e\|_2^2 < \frac{1}{2}\|b\|_2^2 = \varphi(0)$, $X^{\mathrm{opt}} \neq 0$. Suppose $\mathrm{rank}(X^{\mathrm{opt}}) = r \geq 1$, and let

$$X^{\mathrm{opt}} = U\Sigma U^* = (U_1, U_2)\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}(U_1, U_2)^* = U_1 \Sigma_1 U_1^*$$

be the SVD, where $U_1 = (u_1, \ldots, u_r) \in \mathbb{C}^{n \times r}$, $U_2 = (u_{r+1}, \ldots, u_n) \in \mathbb{C}^{n \times (n-r)}$, and $\Sigma_1 = \mathrm{Diag}((\sigma_1, \ldots, \sigma_r)) \in \mathbb{R}^{r \times r}$ with X^{opt}'s positive singular values on its diagonal.

Since X^{opt} is a global minimizer, it is also a stationary point. This means KKT conditions must hold at X^{opt}, i.e., there exists $\Lambda \in \mathbb{C}^{n \times n}$ such that

$$\mathcal{A}^*(\mathcal{A}(X^{\mathrm{opt}}) - b) + \lambda(I_n - \frac{X^{\mathrm{opt}}}{\|X^{\mathrm{opt}}\|_F}) = \Lambda,$$

$$X^{\mathrm{opt}} \succeq 0, \ \Lambda \succeq 0, \ \langle X^{\mathrm{opt}}, \Lambda \rangle = 0. \tag{3.1}$$

Rewriting $I_n = UU^* = U_1 U_1^* + U_2 U_2^*$, (3.1) becomes

$$-\mathcal{A}^*(\mathcal{A}(X^{\mathrm{opt}}) - b) = \lambda(I_n - \frac{X^{\mathrm{opt}}}{\|X^{\mathrm{opt}}\|_F}) - \Lambda = \lambda U_1 U_1^* + (\lambda U_2 U_2^* - \Lambda) - \lambda \frac{X^{\mathrm{opt}}}{\|X^{\mathrm{opt}}\|_F}.$$

Taking the Frobenius norm of both sides above, we obtain

$$\|\mathcal{A}^*(\mathcal{A}(X^{\mathrm{opt}}) - b)\|_F = \|\lambda U_1 U_1^* + (\lambda U_2 U_2^* - \Lambda) - \lambda \frac{X^{\mathrm{opt}}}{\|X^{\mathrm{opt}}\|_F}\|_F$$
$$\geq \|\lambda U_1 U_1^* + (\lambda U_2 U_2^* - \Lambda)\|_F - \lambda. \qquad (3.2)$$

Also, we have $0 = \langle X^{\mathrm{opt}}, \Lambda \rangle = \langle U_1 \Sigma_1 U_1^*, \Lambda \rangle = \sum_{i=1}^r \sigma_i \langle u_i u_i^*, \Lambda \rangle$. But $\langle u_i u_i^*, \Lambda \rangle \geq 0$, since $\Lambda \succeq 0$ and $u_i u_i^* \succeq 0$. So $\langle u_i u_i^*, \Lambda \rangle = 0$ for $1 \leq i \leq m$, and $\langle U_1 U_1^*, \Lambda \rangle = \sum_{i=1}^r \langle u_i u_i^*, \Lambda \rangle = 0$. Moreover,

$$\langle U_1 U_1^*, U_2 U_2^* \rangle = \sum_{i=1}^r \sum_{j=r+1}^n \langle u_i u_i^*, u_j u_j^* \rangle = \sum_{i=1}^r \sum_{j=r+1}^n \langle u_j^* u_i, u_j^* u_i \rangle = 0.$$

In a word, $U_1 U_1^*$ is orthogonal to both $U_2 U_2^*$ and Λ. Then from the Pythagorean theorem it follows that

$$\|\lambda U_1 U_1^* + (\lambda U_2 U_2^* - \Lambda)\|_F = \sqrt{\|\lambda U_1 U_1^*\|_F^2 + \|(\lambda U_2 U_2^* - \Lambda)\|_F^2} \geq \lambda \|U_1 U_1^*\|_F,$$

and thus (3.2) reduces to

$$\|\mathcal{A}^*(\mathcal{A}(X^{\mathrm{opt}}) - b)\|_F \geq \lambda \|U_1 U_1^*\|_F - \lambda = \lambda(\sqrt{r} - 1). \qquad (3.3)$$

On the other hand, since

$$\|\mathcal{A}(X^{\mathrm{opt}}) - b\|_2 \leq \sqrt{\|\mathcal{A}(X^{\mathrm{opt}}) - b\|_2^2 + 2\lambda(\mathrm{Tr}(X^{\mathrm{opt}}) - \|X^{\mathrm{opt}}\|_F)} = \sqrt{2\varphi(X^{\mathrm{opt}})},$$

we have

$$\|\mathcal{A}^*(\mathcal{A}(X^{\mathrm{opt}}) - b)\|_F \leq \|\mathcal{A}^*\| \|\mathcal{A}(X^{\mathrm{opt}}) - b\|_2 = \|\mathcal{A}\| \|\mathcal{A}(X^{\mathrm{opt}}) - b\|_2$$
$$\leq \|\mathcal{A}\| \sqrt{2\varphi(X^{\mathrm{opt}})} \leq \|\mathcal{A}\| \sqrt{2\varphi(\hat{X})} = \|\mathcal{A}\| \|e\|_2. \qquad (3.4)$$

Combining (3.3) and (3.4) gives $\lambda(\sqrt{r} - 1) \leq \|\mathcal{A}\| \|e\|_2$, or equivalently

$$r \leq (\frac{\|\mathcal{A}\| \|e\|_2}{\lambda} + 1)^2 < 2.$$

The last inequality above follows from the assumption $\lambda > \frac{\|\mathcal{A}\| \|e\|_2}{\sqrt{2} - 1}$. r is a natural number, so $r = 1$.

Note that $\mathrm{Tr}(X) - \|X\|_F \geq 0$ for $X \succeq 0$ with equality when $\mathrm{rank}(X) = 1$. It is not hard to see the equivalence between (1.3) and (1.4). $\qquad \square$

COROLLARY 3.1. *In the absence of measurement noise, the equivalence between (1.3) and (1.4) holds for all $\lambda > 0$. In this sense, (1.4) is essentially a parameter-free model.*

Theorem 3.1 claims that provided the noise in measurement is smaller than the measurement itself, all λ that exceed an explicit threshold would work equally well for

(1.4) in theory. In contrast, the λ in (1.2) needs to be carefully chosen to balance the fidelity and penalty terms. Particularly in noiseless case, the λ in (1.4) acts like a 'fool-proof' regularization parameter, and $\mathcal{A}(X) = b$ is always exact at the solution $X^{\mathrm{opt}} = \hat{X}$ whenever $\lambda > 0$, whereas a perfect reconstruction via solving (1.2) generally requires a dynamic λ that goes to 0.

REMARK 3.1.   Despite the tremendous room for λ values in view of Theorem 3.1, we should point out that in practice the choice of λ could be more subtle because

- The theoretical lower bound $\frac{\|\mathcal{A}\|\|e\|_2}{\sqrt{2}-1}$ for λ may be too stringent, and a smaller λ could also be feasible.
- Choosing λ too large may reduce the mobility of the energy minimizing iterations due to trapping by local minima.

An efficient algorithm designed for PhaseLiftOff should be as insensitive as possible to the choice of λ when it is large enough.

3.2. Exact and stable recovery under Gaussian measurements.   In the framework of [8, 7], assuming a_i's are i.i.d. complex-valued normally distributed random vectors, we establish the exact recovery and stability results for (1.4). Due to the equivalence between (1.3) and (1.4) under the conditions stated in Theorem 3.1, it suffices to discuss the model (1.3) only. Similar to [7], $m = O(n)$ measurements suffice to ensure exact recovery in the noiseless case or stability in the noisy case with probability close to 1. Although the required number of measurements for (1.3) and that for PhaseLift are both on the minimal order $O(n)$, the scalar factor of the former is actually smaller, and so is the probability of failure.

THEOREM 3.2.   *Suppose column vectors of A are i.i.d. complex-valued normally distributed. Fix $\alpha \in (0,1)$, there are constants $\theta, \gamma > 0$ such that if $m > \theta[\alpha^{-2}\log\alpha^{-1}]n$, for any $\hat{X}$, (1.3) is stable in the sense that its solution X^{opt} satisfies*

$$\|X^{\mathrm{opt}} - \hat{X}\|_F \leq C_\alpha \frac{\|e\|_2}{\sqrt{m}} \tag{3.5}$$

for some constant $C_\alpha := \frac{\sqrt{2}}{(\sqrt{2}-1)(1-\alpha)} > 0$ with probability at least $1 - 3e^{-\gamma m\alpha^2}$. In particular, when $e = 0$, the recovery is exact.

The proof is straightforward with the aid of Lemma 5.1 in [8]:

LEMMA 3.1 ([8]).   *Under the assumption of Theorem 3.2, we have that $\mathcal{A}$ obeys the following property with probability at least $1 - 3e^{-\gamma m\alpha^2}$: for any Hermitian matrix X with $\mathrm{rank}(X) \leq 2$,*

$$\frac{1}{m}\|\mathcal{A}(X)\|_1 \geq 2(\sqrt{2}-1)(1-\alpha)\|X\|_2.$$

Proof of Theorem 3.2.
   Proof. Let $X^{\mathrm{opt}} = \hat{X} + H$, where $\hat{X}$ satisfies $\mathcal{A}(\hat{X}) + e = b$. Then H is Hermitian with $\mathrm{rank}(H) \leq 2$. Since

$$\|e\|_2 = \|\mathcal{A}(\hat{X}) - b\|_2 \geq \|\mathcal{A}(X^{\mathrm{opt}}) - b\|_2 \geq \|\mathcal{A}(X^{\mathrm{opt}} - \hat{X})\|_2 - \|\mathcal{A}(\hat{X}) - b\|_2,$$

we have $\|\mathcal{A}(H)\|_2 \leq 2\|e\|_2$. Invoking Lemma 3.1 above, we further have

$$\frac{1}{\sqrt{m}}\|\mathcal{A}(H)\|_2 \geq \frac{1}{m}\|\mathcal{A}(H)\|_1 \geq 2(\sqrt{2}-1)(1-\alpha)\|H\|_2 \geq \frac{2(\sqrt{2}-1)(1-\alpha)}{\sqrt{2}}\|H\|_F.$$

Therefore,

$$\|X^{\mathrm{opt}} - \hat{X}\|_F = \|H\|_F \leq \frac{\sqrt{2}}{(\sqrt{2}-1)(1-\alpha)} \frac{\|e\|_2}{\sqrt{m}}.$$

The above inequality holds with probability at least $1 - 3e^{-\gamma m \alpha^2}$.   □

3.3. Computation of $\|\mathcal{A}\|$.   The $\|\mathcal{A}\|$ in Theorem 3.1 can be actually computed. To do this, we first prove the following result:

LEMMA 3.2. $\mathcal{A}\mathcal{A}^*(x) = (A^*A \circ \overline{A^*A})x, \forall x \in \mathbb{R}^m$, where the overline denotes the complex conjugate.

Proof.   By the definitions of $\mathcal{A}$ and $\mathcal{A}^*$, $\mathcal{A}\mathcal{A}^*(x) = \mathrm{diag}(A^*A\mathrm{Diag}(x)A^*A)$, so $\forall 1 \leq i \leq m$, the i-th entry of $\mathcal{A}\mathcal{A}^*(x)$ reads

$$(\mathcal{A}\mathcal{A}^*(x))_i = (A^*A\mathrm{Diag}(x)A^*A)_{ii} = \sum_{j=1}^{m} x_j (A^*A)_{ij}(A^*A)_{ji}$$

$$= \sum_{j=1}^{m} x_j (A^*A)_{ij}\overline{(A^*A)_{ij}} = \sum_{j=1}^{m} x_j (A^*A \circ \overline{A^*A})_{ij}$$

$$= ((A^*A \circ \overline{A^*A})x)_i.$$

□

Hence, from Lemma 3.2 and (2.1) it follows that

$$\|\mathcal{A}\| = \sqrt{\|\mathcal{A}\mathcal{A}^*\|} = \sqrt{\|A^*A \circ \overline{A^*A}\|_2}.$$

It would be interesting to see how fast $\|\mathcal{A}\|$ grows with dimensions n and m when A is a complex-valued random Gaussian matrix. In this setting, $\mathcal{A}$ enjoys approximate ℓ_1-isometry properties as revealed by Lemma 3.1 of [8] (in the complex case). Here we are most interested in the part that concerns the upper bound:

LEMMA 3.3 ([8]). *Suppose $A \in \mathbb{C}^{n \times n}$ is random Gaussian. Fix any $\delta > 0$ and assume $m \geq 16\delta^{-2}n$. Then with probability at least $1 - e^{-m\epsilon^2/2}$, where $\delta/4 = \epsilon^2 + \epsilon$,*

$$\frac{1}{m}\|\mathcal{A}(X)\|_1 \leq (1+\delta)\|X\|_*$$

holds for all $X \in \mathbb{C}^{n \times n}$.

Under assumptions of Lemma 3.3, with high probability we have

$$\frac{1}{\sqrt{m}}\|\mathcal{A}(X)\|_2 \leq \frac{1}{m}\|\mathcal{A}(X)\|_1 \leq (1+\delta)\|X\|_* \leq (1+\delta)\sqrt{n}\|X\|_F,$$

which implies $\|\mathcal{A}\| = O(\sqrt{mn})$. For the phase retrieval problem to be well-posed, $m = O(n)$ is required; for instance, $m = 4n$ would be sufficient according to [2]. Then we expect that $\|\mathcal{A}\|$ is on the order of n. This can be validated by a simple numerical experiment whose results are shown in Table 3.1 below.

4. Algorithms

In this section, we consider the computational aspects of the minimization problem (1.4).

n	32	64	128	256	512	1024
$\|\mathcal{A}\|$	148	291	577	1149	2295	4584

TABLE 3.1. *Fixing $m=4n$, $\|\mathcal{A}\|$ is nearly linear in n, where $\|\mathcal{A}\|=\sqrt{\|A^*A\circ\overline{A^*A}\|_2}$ with A being a complex-valued Gaussian matrix. For each n, the value of $\|\mathcal{A}\|$ is averaged over 10 independent samples of A using MATLAB.*

4.1. Difference of convex functions algorithm. The DCA is a descent method without line search developed by Tao and An [1, 25]. It addresses the problem of minimizing a function of the form $f(x)=g(x)-h(x)$ on the space $\mathbb{R}^n$, with g, h being lower semicontinuous proper convex functions:

$$\min_{x\in\mathbb{R}^n} f(x)$$

$g-h$ is called a DC decomposition of f, while the convex functions g and h are DC components of f. The DCA involves the construction of two sequences $\{x^k\}$ and $\{y^k\}$, the candidates for optimal solutions of primal and dual programs respectively. At the $(k+1)$-th step, we choose a subgradient of $h(x)$ at x^k, namely $y^k\in\partial h(x^k)$. We then linearize h at x^k, which permits a convex upper envelope of f. More precisely,

$$f(x)=g(x)-h(x)\leq g(x)-(h(x^k)+\langle y^k,x-x^k\rangle),\forall x\in\mathbb{R}^n$$

with equality at $x=x^k$.

By iteratively computing

$$\begin{cases} y^k\in\partial h(x^k), \\ x^{k+1}=\arg\min_{x\in\mathbb{R}^n} g(x)-(h(x^k)+\langle y^k,x-x^k\rangle) \end{cases}$$

we have

$$f(x^k)\geq g(x^{k+1})-(h(x^k)+\langle y^k,x^{k+1}-x^k\rangle)\geq g(x^{k+1})-h(x^{k+1})=f(x^{k+1}).$$

This generates a monotonically decreasing sequence $\{f(x^k)\}$, leading to its convergence if $f(x)$ is bounded from below.

We can readily apply the DCA to (1.4), where the objective naturally has the DC decomposition

$$\varphi(X)=(\frac{1}{2}\|\mathcal{A}(X)-b\|_2^2+\lambda\mathrm{Tr}(X))-\lambda\|X\|_F. \qquad (4.1)$$

Since $\varphi(X)\geq 0$ for all $X\succeq 0$, the scheme

$$\begin{cases} \Delta^k\in\partial\|X^k\|_F, \\ X^{k+1}=\arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2+\lambda\mathrm{Tr}(X)-\lambda(\|X^k\|_F+\langle\Delta^k,X-X^k\rangle)\,\mathrm{s.t.}\,X\succeq 0 \end{cases}$$

yields a decreasing and convergent sequence $\{\varphi(X^k)\}$. Note that $\|X\|_F$ is differentiable with gradient $\frac{X}{\|X\|_F}$ at all $X\neq 0$ and that $0\in\partial\|X\|_F$ at $X=0$. By ignoring constants we iterate

$$X^{k+1}=\begin{cases} \arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2+\lambda\mathrm{Tr}(X) & \mathrm{s.t.}\quad X\succeq 0 & \mathrm{if}\quad X^k=0, \\ \arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2+\lambda\langle X,I_n-\frac{X^k}{\|X^k\|_F}\rangle & \mathrm{s.t.}\quad X\succeq 0 & \mathrm{otherwise.} \end{cases} \qquad (4.2)$$

Since $X^k - X^{k-1} \to 0$ as $k \to \infty$ (Proposition 4.1 (2)), we stop the DCA when

$$\frac{\|X^k - X^{k-1}\|_F}{\max\{\|X^k\|_F, 1\}} < \texttt{tol},$$

for some given tolerance $\texttt{tol} > 0$. In practice the above iteration takes only a few steps to convergence. While the problem (1.4) is nonconvex, empirical studies have shown that the DCA usually produces a global minimizer with a good initialization. In particular, our initialization here is $X^0 = 0$, as suggested by the observations in [27]. This amounts to employing the (PhaseLift) solution of the regularized trace-norm minimization problem (1.2) as a start.

4.2. Convergence analysis. We proceed to show that the sequence $\{X^k\}$ is bounded and $X^{k+1} - X^k \to 0$, and limit points of $\{X^k\}$ are stationary points of (1.4) satisfying KKT optimality conditions. Standard convergence results for the general DCA (e.g. Theorem 3.7 of [25]) take advantage of strong convexity of the DC components. However, the DC components in (4.1) only possess weak convexity as $\ker(\mathcal{A}^*\mathcal{A})$ is generally nontrivial. In this sense, our analysis below is novel.

LEMMA 4.1. *Suppose $X \succeq 0$. Then $\varphi(X) \to \infty$ as $X \to \infty$.*

Proof. It suffices to show that for any fixed nonzero $X \succeq 0$, $\varphi(cX) \to \infty$ as $c \to \infty$.

$$\varphi(cX) = \frac{1}{2}\|c\mathcal{A}(X) - b\|_2^2 + c\lambda(\mathrm{Tr}(X) - \|X\|_F) \geq \frac{1}{2}(c\|\mathcal{A}(X)\|_2 - \|b\|_2)^2.$$

Since $X \succeq 0$ and X is nonzero, by Lemma 2.1 (3), $\|\mathcal{A}(X)\|_2 > 0$. Hence, $c\|\mathcal{A}(X)\|_2 - \|b\|_2 \to \infty$ as $c \to \infty$, which completes the proof. □

LEMMA 4.2. *Let $\{X^k\}$ be the sequence generated by the DCA. For all $k \in \mathbb{N}$, we have*

$$\varphi(X^k) - \varphi(X^{k+1}) \geq \frac{1}{2}\|\mathcal{A}(X^k - X^{k+1})\|_2^2 + \lambda(\|X^{k+1}\|_F - \|X^k\|_F - \langle \Delta^k, X^{k+1} - X^k \rangle) \geq 0, \quad (4.3)$$

where $\Delta^k \in \partial\|X^k\|_F$.

Proof. We first calculate

$$\varphi(X^k) - \varphi(X^{k+1}) = \frac{1}{2}\|\mathcal{A}(X^k - X^{k+1})\|_2^2 + \langle \mathcal{A}(X^k - X^{k+1}), \mathcal{A}(X^{k+1}) - b \rangle$$
$$+ \lambda\mathrm{Tr}(X^k - X^{k+1}) + \lambda(\|X^{k+1}\|_F - \|X^k\|_F). \quad (4.4)$$

Recall that the $(k+1)$-th DCA iteration is to solve

$$X^{k+1} = \arg\min_{X \in \mathbb{C}^{n \times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \lambda\langle X, I_n - \Delta^k \rangle \quad \text{s.t.} \quad X \succeq 0,$$

where $\Delta^k \in \partial\|X^k\|_F$. Then by the KKT conditions at X^{k+1}, there exists Λ^{k+1} such that

$$\mathcal{A}^*(\mathcal{A}(X^{k+1}) - b) + \lambda(I_n - \Delta^k) = \Lambda^{k+1}, \quad (4.5)$$
$$X^{k+1} \succeq 0, \ \Lambda^{k+1} \succeq 0, \ \langle \Lambda^{k+1}, X^{k+1} \rangle = 0.$$

Multiplying (4.5) by $X^k - X^{k+1}$ (inner product) gives

$$\langle \mathcal{A}(X^k - X^{k+1}), \mathcal{A}(X^{k+1}) - b \rangle + \lambda\mathrm{Tr}(X^k - X^{k+1}) = \langle \Lambda^{k+1}, X^k \rangle - \lambda\langle \Delta^k, X^{k+1} - X^k \rangle. \quad (4.6)$$

In (4.6), $\langle \Delta^k, X^{k+1} - X^k \rangle \leq \|X^{k+1}\|_F - \|X^k\|_F$ since $\Delta^k \in \partial\|X^k\|_F$, and $\langle \Lambda^{k+1}, X^k \rangle \geq 0$ by Lemma 2.1 (1). Combining (4.4) and (4.6) gives

$$\varphi(X^k) - \varphi(X^{k+1})$$
$$= \frac{1}{2}\|\mathcal{A}(X^k - X^{k+1})\|_2^2 + \lambda(\|X^{k+1}\|_F - \|X^k\|_F) + \langle \Lambda^{k+1}, X^k \rangle - \lambda\langle \Delta^k, X^{k+1} - X^k \rangle$$
$$\geq \frac{1}{2}\|\mathcal{A}(X^k - X^{k+1})\|_2^2 + \lambda(\|X^{k+1}\|_F - \|X^k\|_F - \langle \Delta^k, X^{k+1} - X^k \rangle)$$
$$\geq 0.$$

□

We are now in the position to prove convergence results of the DCA for solving the PhaseLiftOff problem (1.4).

PROPOSITION 4.1. *Let $\{X^k\}$ be the sequence produced by the DCA starting with $X^0 = 0$.*

1. *$\{X^k\}$ is bounded.*

2. *$X^{k+1} - X^k \to 0$ as $k \to \infty$.*

3. *Any nonzero limit point $\tilde{X}$ of the sequence $\{X^k\}$ is a first-order stationary point, which means there exists $\tilde{\Lambda}$, such that the following KKT conditions are satisfied:*
- *Stationarity: $\mathcal{A}^*(\mathcal{A}(\tilde{X}) - b) + \lambda(I_n - \frac{\tilde{X}}{\|\tilde{X}\|_F}) = \tilde{\Lambda}$.*
- *Primal feasibility: $\tilde{X} \succeq 0$.*
- *Dual feasibility: $\tilde{\Lambda} \succeq 0$.*
- *Complementary slackness: $\langle \tilde{X}, \tilde{\Lambda} \rangle = 0$.*

Proof.

(1) By Lemma 4.1, the level set $\Omega := \{X \in \mathbb{C}^{n \times n} : X \succeq 0, \ \varphi(X) \le \varphi(0)\}$ is bounded. Since $\{\varphi(X^k)\}$ is decreasing, $\{X^k\} \subseteq \Omega$ is also bounded.

(2) Letting $k = 0$ and substituting $\Delta^0 = 0$ in (4.3), we obtain

$$\varphi(0) - \varphi(X^1) \ge \frac{1}{2}\|\mathcal{A}(X^1)\|_2^2 + \lambda\|X^1\|_F.$$

If $X^1 \ne 0$, then $\varphi(0) > \varphi(X^1) \ge \cdots \ge \varphi(X^k)$, so $X^k \ne 0, \ \forall k \ge 1$. Otherwise $X^k \equiv 0$.

Assuming $X^k \ne 0$, we show that $X^{k+1} - X^k \to 0$ as $k \to \infty$ in what follows. Note that $\{\varphi(X^k)\}$ is decreasing and convergent, and that $\Delta^k = \frac{X^k}{\|X^k\|_F}$ when $k \ge 1$. Combining this with (4.3), we have the following key information about $\{X^k\}$:

$$\|\mathcal{A}(X^k - X^{k+1})\|_2 \to 0 \tag{4.7}$$

$$\|X^{k+1}\|_F - \langle \frac{X^k}{\|X^k\|_F}, X^{k+1} \rangle \to 0. \tag{4.8}$$

Define $c^k := \frac{\langle X^k, X^{k+1} \rangle}{\|X^k\|_F^2} \ge 0$ and $E^k := X^{k+1} - c^k X^k$, then it suffices to prove $E^k \to 0$ and $c^k \to 1$. A simple computation shows

$$\|E^k\|_F^2 = \|X^{k+1}\|_F^2 - \frac{\langle X^k, X^{k+1} \rangle^2}{\|X^k\|_F^2} \to 0,$$

where (4.8) was used. Thus, from (4.7) it follows that

$$0 = \lim_{k \to \infty} \|\mathcal{A}(X^k - X^{k+1})\|_2 = \lim_{k \to \infty} \|\mathcal{A}((c^k - 1)X^k - E^k)\|_2 = \lim_{k \to \infty} |c^k - 1|\|\mathcal{A}(X^k)\|_2.$$

Suppose $\lim_{k \to \infty} c^k \ne 1$, then there exists a subsequence $\{X^{k_j}\}$ such that $\|\mathcal{A}(X^{k_j})\|_2 \to 0$. Since, by Lemma 2.1 (3), $\mathcal{A}(X) = 0 \Leftrightarrow X = 0$ for $X \succeq 0$, we must have $X^{k_j} \to 0$ and $\varphi(X^{k_j}) \to \varphi(0)$, which leads to a contradiction because

$$\varphi(X^{k_j}) \le \varphi(X^1) < \varphi(0).$$

Therefore $c^k \to 1$ and $X^{k+1} - X^k \to 0$, as $k \to \infty$.

(3) Let $\{X^{k_j}\}$ be a subsequence of $\{X^k\}$ converging to some limit point $\tilde{X} \ne 0$, then the optimality conditions at the k_j-th step read:

$$\mathcal{A}^*(\mathcal{A}(X^{k_j}) - b) + \lambda(I_n - \frac{X^{k_j - 1}}{\|X^{k_j - 1}\|_F}) = \Lambda^{k_j},$$

$$X^{k_j} \succeq 0, \ \Lambda^{k_j} \succeq 0, \ \langle \Lambda^{k_j}, X^{k_j} \rangle = 0.$$

Define

$$\tilde{\Lambda} := \lim_{k_j \to \infty} \Lambda^{k_j}$$

$$= \lim_{k_j \to \infty} \mathcal{A}^*(\mathcal{A}(X^{k_j}) - b) + \lambda(I_n - \frac{X^{k_j - 1}}{\|X^{k_j - 1}\|_F})$$

$$= \lim_{k_j \to \infty} \mathcal{A}^*(\mathcal{A}(X^{k_j}) - b) + \lambda(I_n - \frac{X^{k_j}}{\|X^{k_j}\|_F}) + \lambda(\frac{X^{k_j}}{\|X^{k_j}\|_F} - \frac{X^{k_j - 1}}{\|X^{k_j - 1}\|_F})$$

$$= \mathcal{A}^*(\mathcal{A}(\tilde{X}) - b) + \lambda(I_n - \frac{\tilde{X}}{\|\tilde{X}\|_F}).$$

In the last equality, we used $\lim_{k_j \to \infty} X^{k_j} = \tilde{X} \neq 0$ and $X^{k_j} - X^{k_j - 1} \to 0$. Since $X^{k_j} \succeq 0$, $\Lambda^{k_j} \succeq 0$, their limits are $\tilde{X} \succeq 0$ and $\tilde{\Lambda} \succeq 0$. It remains to check that $\langle \tilde{\Lambda}, \tilde{X} \rangle = 0$. Using $\langle \Lambda^{k_j}, X^{k_j} \rangle = 0$, we have

$$\langle \tilde{\Lambda}, \tilde{X} \rangle = \langle \tilde{\Lambda} - \Lambda^{k_j}, \tilde{X} - X^{k_j} \rangle + \langle \Lambda^{k_j}, \tilde{X} \rangle + \langle \tilde{\Lambda}, X^{k_j} \rangle.$$

Letting $k_j \to \infty$ on the right hand side above, we have $\langle \tilde{\Lambda}, \tilde{X} \rangle = 0$.   □

REMARK 4.1.   In light of the proof of Proposition 4.1 (2), one can see that for all $k \geq 1$, either $X^k \equiv 0$ or $\|X^k\|_F > \eta$ for some $\eta > 0$. A sufficient condition to ensure that the DCA does not yield $\tilde{X} = 0$ is as follows:

$$\frac{1}{2}\|b\|_2^2 > \frac{1}{2}\|e\|_2^2 + \lambda \text{Tr}(\hat{X}) \Leftrightarrow \lambda < \frac{\|b\|_2^2 - \|e\|_2^2}{2\text{Tr}(\hat{X})},$$

where $\hat{X}$ is the ground truth obeying $b = \mathcal{A}(\hat{X}) + e$. The above condition would guarantee that $X^1 \neq 0$. Though the equivalence between (1.3) and (1.4) follows from Theorem 3.1 as long as λ is sufficiently large, in practice λ cannot get too large because the DCA iterations may stall at $X^0 = 0$.

4.3. Solving the subproblem.   At the $(k+1)$-th DC iteration, one needs to solve a convex subproblem of the form:

$$X^{k+1} = \arg\min_{X \in \mathbb{C}^{n \times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \langle X, W \rangle \quad \text{s.t.} \quad X \succeq 0. \tag{4.9}$$

In our case, $W = \lambda I_n$ or $\lambda(I_n - \frac{X^k}{\|X^k\|_F})$ is a known Hermitian matrix.

This problem can be treated as a weighted trace-norm regularization problem, which has been studied in [6]. The authors of [6] suggest using FISTA [4, 6] which is a variant of Nesterov's accelerated gradient descent method [24]. An alternative choice is the alternating direction method of multipliers (ADMM). We only discuss the naive ADMM here, though this algorithm could be further accelerated by incorporating Nesterov's idea [21]. To implement ADMM, we introduce a dual variable Y and form the augmented Lagrangian

$$\mathcal{L}_\delta(X, Y, Z) = \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \langle X, W \rangle + \langle Y, X - Z \rangle + \frac{\delta}{2}\|X - Z\|_F^2 + g_\succeq(Z), \tag{4.10}$$

where

$$g_\succeq(Z) = \begin{cases} 0 & \text{if} \quad Z \succeq 0, \\ \infty & \text{otherwise}. \end{cases}$$

ADMM consists of updates on both the primal and dual variables [5]:

$$\begin{cases} X^{l+1} = \arg\min_X \mathcal{L}_\delta(X, Y^l, Z^l) \\ Z^{l+1} = \arg\min_Z \mathcal{L}_\delta(X^{l+1}, Y^l, Z) \\ Y^{l+1} = Y^l + \delta(X^{l+1} - Z^{l+1}) \end{cases}$$

Algorithm 1 ADMM for solving (4.9)

---

while not converged **do**

$X^{l+1} = (\mathcal{A}^*\mathcal{A} + \delta\mathcal{I}_n)^{-1}(\mathcal{A}^*(b) - W + \delta Z^l - Y^l)$

$Z^{l+1} = \mathcal{P}_{\succeq}(X^{l+1} + Y^l/\delta)$

$Y^{l+1} = Y^l + \delta(X^{l+1} - Z^{l+1})$

end while

The first two steps have closed-form solutions, which are detailed in Algorithm 1.

In the X-update step, one needs to know the expression of $(\mathcal{A}^*\mathcal{A} + \delta\mathcal{I}_n)^{-1}$. The celebrated Woodbury formula implies

$$(\mathcal{A}^*\mathcal{A} + \delta\mathcal{I}_n)^{-1} = \frac{1}{\delta}(\mathcal{I}_n - \mathcal{A}^*(\mathcal{A}\mathcal{A}^* + \delta\mathcal{I}_m)^{-1}\mathcal{A}).$$

By Lemma 3.2, $\mathcal{A}\mathcal{A}^* = A^*A \circ \overline{A^*A}$, so we have

$$
\begin{aligned}
(\mathcal{A}^*\mathcal{A} + \delta\mathcal{I}_n)^{-1}(X) &= \frac{1}{\delta}(X - \mathcal{A}^*((\mathcal{A}\mathcal{A}^* + \delta\mathcal{I}_m)^{-1}\mathcal{A}(X))) \\
&= \frac{1}{\delta}(X - A\mathrm{Diag}((A^*A \circ \overline{A^*A} + \delta I_m)^{-1}\mathrm{diag}(A^*XA))A^*).
\end{aligned}
$$

In the Z-update step, $\mathcal{P}_{\succeq} : \mathbb{H}^{n\times n} \to \mathbb{H}^{n\times n}$ represents the projection onto the positive semidefinite cone. More precisely, if X has the eigenvalue decomposition $X = U\Sigma U^*$, then

$$\mathcal{P}_{\succeq}(X) = U\max\{\Sigma, 0\}U^*.$$

According to [5], the stopping criterion here is given by:

$$\|R^l\|_F \le n\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}}\max\{\|X^l\|_F, \|Z^l\|_F\}, \quad \|S^l\|_F \le n\epsilon^{\mathrm{abs}} + \epsilon^{\mathrm{rel}}\|Y^l\|_F,$$

where $R^l = X^l - Z^l$, $S^l = \delta(Z^l - Z^{l-1})$ are primal and dual residuals respectively at the l-th iteration. $\epsilon^{\mathrm{abs}} > 0$ is an absolute tolerance and $\epsilon^{\mathrm{rel}} > 0$ is a relative tolerance, and they are both algorithm parameters. δ is typically fixed, but one can also adaptively update it during iterations following the rule in [5]; for instance,

$$\delta^{l+1} = \begin{cases} 2\delta^l & \text{if} \quad \|R^l\|_F > 10\|S^l\|_F, \\ \delta^l/2 & \text{if} \quad 10\|R^l\|_F < \|S^l\|_F, \\ \delta^l & \text{otherwise.} \end{cases}$$

4.4. Real-valued, nonnegative signals. If the signal is known to be real or nonnegative, we should add one more constraint to the complex PhaseLiftOff (1.4):

$$\min_{X\in\mathbb{C}^{n\times n}} \varphi(X) \quad \text{s.t.} \quad X \succeq 0, X \in \Omega. \tag{4.11}$$

Here Ω is $\mathbb{R}^{n\times n}$ (or resp., $\mathbb{R}_+^{n\times n}$), which means each entry of X is real (or resp., nonnegative). Thus we need to modify the DCA (4.2) accordingly:

$$X^{k+1} = \begin{cases} \arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \lambda\mathrm{Tr}(X) \,\text{s.t.}\, X \succeq 0, X \in \Omega & \text{if } X^k = 0, \\ \arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \lambda\langle X, I_n - \frac{X^k}{\|X^k\|_F}\rangle \,\text{s.t.}\, X \succeq 0, X \in \Omega & \text{otherwise.} \end{cases} \tag{4.12}$$

The above subproblem at each DCA iteration can also be solved by ADMM. Specifically, we want to solve the optimization problem of the following form:

$$\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \langle X, W\rangle \quad \text{s.t.} \quad X \succeq 0, X \in \Omega. \tag{4.13}$$

In ADMM form, (4.13) is reformulated as

$$\min_{X,Z\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2 + \langle X,W\rangle + g_\Omega(X) + g_\succeq(Z) \quad \text{s.t.} \quad X-Z=0,$$

where $g_\succeq(Z)$ is the same as in (4.10), and

$$g_\Omega(X) = \begin{cases} 0 & \text{if} \quad X\in\Omega, \\ \infty & \text{otherwise.} \end{cases}$$

Having defined the augmented Lagrangian

$$\mathcal{L}_\delta(X,Y,Z) = \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2 + \langle X,W\rangle + \langle Y,X-Z\rangle + \frac{\delta}{2}\|X-Z\|_F^2 + g_\Omega(X) + g_\succeq(Z),$$

we arrive at Algorithm 2 by alternately minimizing $\mathcal{L}_\delta$ with respect to X, minimizing with respect to Z, and updating the dual variable Y. The operator $\mathcal{P}_\Omega : \mathbb{H}^{n\times n}\to\Omega$ in Algorithm 2 represents the projection onto the set Ω. In particular, $\mathcal{P}_\Omega(X) = \text{Re}(X)$ is the real part of X for $\Omega = \mathbb{R}^{n\times n}$, whereas $\mathcal{P}_\Omega(X) = \max\{\text{Re}(X),0\}$ for $\Omega = \mathbb{R}_+^{n\times n}$. Algorithm 2 is almost identical to Algorithm 1 except that an extra projection $\mathcal{P}_\Omega$ is performed in the X-update step.

Algorithm 2 ADMM for solving (4.13)

---

    while not converged **do**
        $X^{l+1} = \mathcal{P}_\Omega((\mathcal{A}^*\mathcal{A}+\delta\mathcal{I}_n)^{-1}(\mathcal{A}^*(b)-W+\delta Z^l - Y^l))$
        $Z^{l+1} = \mathcal{P}_\succeq(X^{l+1}+Y^l/\delta)$
        $Y^{l+1} = Y^l + \delta(X^{l+1}-Z^{l+1})$
    end while

---

5. Numerical experiments

In this section, we report numerical results. Besides the proposed (1.4) and the regularized PhaseLift (1.2), we also discuss the following reweighting scheme from [6], which is an extension of reweighted ℓ_1 algorithm in the regime of compressed sensing introduced in [9]:

$$X^{k+1} = \arg\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2 + \lambda\langle W^k,X\rangle \quad \text{s.t.} \quad X\succeq 0, \tag{5.1}$$

where $W^0 = I_n$ and $W^k = (X^k+\varepsilon I_n)^{-1}$ for $k\geq 1$ and for some $\varepsilon > 0$. The aim of this scheme is to provide more accurate solutions with lower rank than that of PhaseLift. Note that W^k is exactly the gradient of $\log(\det(X+\varepsilon I_n))$ at X^k, the reweighting scheme is in essence an implementation of the DCA attempting to solve the nonconvex problem

$$\min_{X\in\mathbb{C}^{n\times n}} \frac{1}{2}\|\mathcal{A}(X)-b\|_2^2 + \lambda\log(\det(X+\varepsilon I_n)) \quad \text{s.t.} \quad X\succeq 0. \tag{5.2}$$

Here the DC components are $\frac{1}{2}\|\mathcal{A}(X)-b\|_2^2$ and $-\lambda\log(\det(X+\varepsilon I_n))$. We hereby remark that with positive semidefinite constraint, the DCA is basically equivalent to the reweighting scheme. In [6], (1.2) and the subproblem (5.1) of (5.2) are solved by FISTA. Here we solve them using ADMM (Algorithm 1) instead as we find it more efficient.

5.1. Exact recovery from noise-free measurements. We set up a phase retrieval problem by 1) generating a random complex-valued signal $\hat{x}$ of length $n=32$ whose real and imaginary parts are Gaussian, 2) sampling a Gaussian matrix $A\in\mathbb{C}^{n\times m}$ with $m=60,62,\ldots,150$, and 3) computing the measurements $b=\mathcal{A}(\hat{x}\hat{x}^*)$. We then solve (1.4), (1.2) and (5.2) to get approximations to $\hat{x}\hat{x}^*$. The ultimate goal of phase retrieval is to reconstruct the signal $\hat{x}$ rather than the rank-1 matrix $\hat{x}\hat{x}^*$. So given a solution $\tilde{X}$, we need to compute the

relative mean squared error (rel. MSE) between $\tilde{x} = \sqrt{\sigma_1(\tilde{X})}u_1$ and $\hat{x}$ modulo a global phase term to measure the recovery quality, where $\sigma_1(\tilde{X})$ is the largest singular value (or eigenvalue) of $\tilde{X}$ and u_1 the corresponding unit-normed eigenvector. More precisely, the rel. MSE is given by

$$\min_{c \in \mathbb{C}:|c|=1} \frac{\|c\tilde{x} - \hat{x}\|_2^2}{\|\hat{x}\|_2^2}.$$

It is easy to show that its minimum occurs at

$$\tilde{c} = \frac{\langle \tilde{x}, \hat{x} \rangle}{|\langle \tilde{x}, \hat{x} \rangle|}.$$

A recovery is considered as a success if the rel. MSE is less than 10^{-6} (or equivalently, relative error $< 10^{-3}$). For each $m = 60, 63, \ldots, 150$, we repeat the above procedures 100 times and record the success rate for each model.

For (1.2), we set $\lambda = 10^{-4}$, $\epsilon^{\text{rel}} = 10^{-5}$, and $\epsilon^{\text{abs}} = 10^{-7}$ in its ADMM algorithm; for (1.4), $\lambda = 10^{-4}$, $\epsilon^{\text{rel}} = 10^{-5}$, $\epsilon^{\text{abs}} = 10^{-7}$, and $\texttt{tol} = 10^{-2}$; parameters for (5.2) are the same as those for (1.4) except that there is an additional parameter $\varepsilon = 2$. In addition, the maximum iteration set for all ADMM algorithms is 5000 and that for the DCA and the reweighting algorithm are both 10. All three methods start with the same initial point $X^0 = 0$.
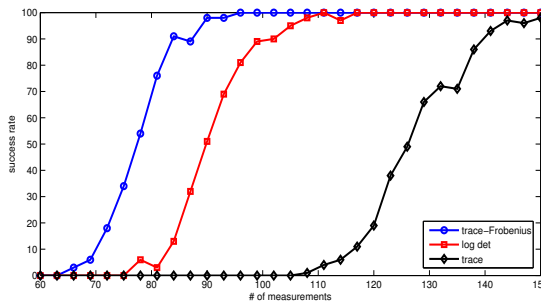

FIG. 5.1. *success rate v.s. number of measurements with parameters* $n = 32$, $m = 60, 63, \ldots, 150$, *and 100 runs at each m.*

The success rate vs. number of measurements plot is shown in Figure 5.1. The result validates that nonconvex proxy for the rank functional gives significantly better recovery quality than the convex trace norm. A similar finding has been reported in the regime of compressed sensing [27]. We also observe that PhaseLiftOff outperforms log-det regularization. This is not surprising as the former always captures rank-1 solutions. In Figure 5.1, one can see that when the number of measurements is $m \approx 3n = 96$, solving our model by the DCA guarantees exact recovery with high probability. Recall that in theory [2] at least $3n - 2$ measurements are needed to recover the signal exactly. This is an indication that the proposed method is likely to provide the optimal practical results one can hope for.

5.2. Robust recovery from noisy measurements. We investigate how the proposed method performs in the presence of noise. The test signal $\hat{x}$ is a Gaussian complex-valued signal of length $n = 32$. We sample $m = 4n$ Gaussian measurement vectors in $\mathbb{C}^n$ and compute the measurements $b \in \mathbb{R}^m$, followed by adding additive white Gaussian noise by means of the MATLAB function $\texttt{awgn(b,snr)}$. There are 6 noise levels varying from 5dB to 55dB. We then apply the DCA to achieve a reconstruction $\tilde{X}$ and compute the signal-to-noise ratio (SNR) of reconstruction in dB defined as $-10\log_{10}(\text{rel. MSE})$. The SNR of reconstruction for each noise level is finally averaged over 10 independent runs.

A crucial point to address here is how we set the value of λ. Theorem 3.1 predicts that provided the noise amount $\|e\|_2$ is known, when $\lambda > \frac{\|\mathcal{A}\|\|e\|_2}{\sqrt{2}-1} \approx 2.414\|\mathcal{A}\|\|e\|_2$, the PhaseLiftOff (1.4) is equivalent to the phase retrieval problem (1.3), and its solution is no longer related to λ. From computational perspective, however, λ cannot be too large as the algorithm may often get stuck at a local solution. An extreme example is that if λ is exceedingly large, the DCA will be trapped at the initial guess $X^0 = 0$. On the other hand, if λ is too small, the reconstruction will be of course far from the ground truth as $\mathcal{A}(X) = b$ tends to be enforced. But can we choose λ that is less than $2.414\|\mathcal{A}\|\|e\|_2$? The answer is yes, since this bound only provides a sufficient condition for equivalence.
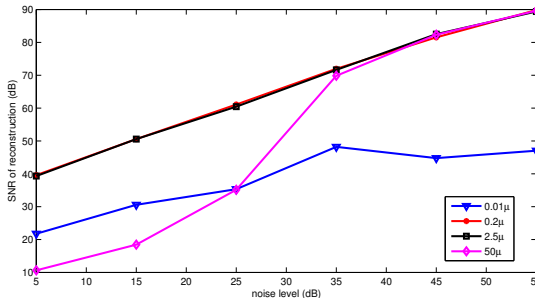

FIG. 5.2. *SNR of signal recovery v.s. noise level in measurement (in SNR dB). Parameters are:* $n = 32$*, noise level = 5dB, 15dB, ..., 55dB;* $\lambda = 0.01\mu, 0.2\mu, 2.5\mu, 50\mu$*; with 10 runs at each noise level.*

Suppose the noise amount $\|e\|_2$ (or its estimate) is known. Defining

$$\mu := \|\mathcal{A}\|\|e\|_2 = \sqrt{\|A^*A \circ \overline{A^*A}\|_2}\|e\|_2,$$

we try 4 different values of λ in each single run. They are multiples of μ, namely 0.01μ, 0.2μ, 2.5μ, and 50μ. The maximum outer and inner iterations are 10 and 5000 respectively. The other parameters are $\epsilon^{\mathrm{rel}} = 10^{-5}$, $\epsilon^{\mathrm{abs}} = 10^{-7}$, $\mathtt{tol} = 10^{-2}$. The reconstruction results are depicted in Figure 5.2. The two curves for 0.2μ and 2.5μ nearly coincide, and they are almost linear, which strongly suggests stable recoveries. In contrast, the algorithm with 0.01μ and 50μ performed poorly. Although $\lambda = 50\mu$ yields comparable reconstruction when there is little noise, the DCA clearly encounters local minima in the low SNR regime. On the other hand, 0.01μ is too small. Summarizing these observations, we conclude that for the DCA method, a reasonable value for λ lies in (but is not limited to) the interval $[0.2\mu, 2.5\mu]$.

6. Conclusions

We introduced and analyzed a novel penalty (trace minus Frobenius norm) for phase retrieval in the PhaseLiftOff least squares regularization problem. We proved its equivalence with rank-1 least squares and stable recovery for noisy measurement at high probability. The DC algorithm for energy minimization is proved to converge to a stationary point satisfying KKT conditions without imposing strict convexity of convex components of the energy (a step beyond the standard DCA theory [1]). Numerical experiments showed that the PhaseLiftOff method outperforms PhaseLift and its nonconvex variant (log-det regularization). The minimal number of measurements for exact recovery by PhaseLiftOff approaches the theoretical limit. In future work, we shall further explore the potential of PhaseLiftOff in phase retrieval applications and rank-1 optimization problems. We are also interested in developing faster optimization algorithms that are more robust to the value of λ.

REFERENCES

[1] Le-Thi Haoi An and Pham-Dinh Tao, *Solving a class of linearly constrained indefinite quadratic problems by D.C. algorithms*, J. Global Opt., 11, 253–285, 1997.

[2] R. Balan, P. Casazza, and D. Edidin, *On signal reconstruction without noisy phase*, Appl. Comp. Harm. Anal., 20, 345–356, 2006.

[3] R. Balan, B. Bodemann, P. Casazza, and D. Edidin, *Painless reconstruction from magnitudes of frame coefficients*, J. Fourier Anal. Appl., 15, 488–501, 2009.

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2(1), 183–202, 2009.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3(1), 1–122, 2011.

[6] E.J. Candès, Y.C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM J. Imaging Sci., 6, 199–225, 2013.

[7] E.J. Candès and X. Li, *Solving quadratic equations via phaseLift when there are about as many equations as unknowns*, Found. Comput. Math., 14, 1017–1026, 2014.

[8] E.J. Candès, T. Strohmer, and V. Voroninski, *PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming*, Commun. Pure Appl. Math., 66, 1241–1274, 2011.

[9] E.J. Candès, M.B. Wakin, and S. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14(5), 877–905, 2008.

[10] A. Chai, M. Moscoso, and G. Papanicolaou, *Array imaging using intensity-only measurements*, Inverse Problems, 27, 015005, 2011.

[11] L. Demanet and P. Hand, *Stable optimizationless recovery from phaseless linear measurements*, J. Fourier Anal. Appl., 20, 199–221, 2014.

[12] E. Esser, Y. Lou, and J. Xin, *A Method for finding structured sparse solutions to non-negative least squares problems with applications*, SIAM J. Imaging Sci., 6(4), 2010–2046, 2013.

[13] A. Fannjiang and W. Liao, *Phase retrieval with random phase illumination*, J. Optic. Soc. America A, 29, 1847–1859, 2012.

[14] A. Fannjiang and W. Liao, *Phasing with phase-uncertain mask*, Inverse Problems, 29, 125001, 2013.

[15] M. Fazel, *Matrix Rank Minimization with Applications*, PhD thesis, Stanford University, 2002.

[16] M. Fazel, H. Hindi, and S. Boyd, *Log-det heuristics for matrix rank minimization with applications to Hankel and Euclidean distance metrics*, Proc. Am. Control Conf, 2156–2162, 2003.

[17] J. Finkelstein, *Pure-state informationally complete and "really" complete measurements*, Phys. Rev. A, 70:052107, 2004.

[18] J. Fienup, *Reconstruction of an object from the modulus of its Fourier transform*, Optics Letters, (3), 27–29, 1978.

[19] J. Fienup, *Phase retrieval algorithms: A comparison*, Appl. Optics, 21, 2758–2769, 1982.

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

[21] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, *Fast alternating direction optimization methods*, UCLA CAM-report, 12–35, 2012.

[22] R. Harrison, *Phase problem in crystallography*, J. Optic. Soc. America A, 10(5), 1046–1055, 1993.

[23] Z. Mou-Yan and R. Unbehauen, *Methods for reconstruction of 2-D sequences from Fourier transform magnitudes*, IEEE Transaction on Image Processing, 6(2), 222–233, 1997.

[24] Y. Nesterov, *Introductory Lectures on Convex Optimization*, New York, NY: Kluwer Academic Press, 2004.

[25] Pham-Dinh Tao and Le-Thi Haoi An, *A D.C. optimization algorithm for solving the trust-region subproblem*, SIAM J. Optim., 8(2), 476–505, 1998.

[26] I. Waldspurger, A. D'Aspremont, and S. Mallat, *Phase Recovery, MaxCut and Complex Semidefinite Programming*, Math. Program., Ser. A, DOI 10.1007/s10107-013-0738-9, 2013.

[27] P. Yin, Y. Lou, Q. He, and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, UCLA CAM-report, 14–01, 2014, to appear in SIAM J. Sci. Computing, 2015.