Scalable Incremental Nonconvex Optimization Approach for Phase Retrieval

Ji Li1 · Jian-Feng Cai2 · Hongkai Zhao3

Received: 25 May 2019 / Revised: 23 January 2021 / Accepted: 8 February 2021 / Published online: 19 March 2021 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

We aim to find a solution $x \in \mathbb{C}^n$ to a system of quadratic equations of the form $b_i = |a_i^* x|^2$, $i = 1, 2, \ldots, m$, e.g., the well-known NP-hard phase retrieval problem. As opposed to recently proposed state-of-the-art nonconvex methods, we revert to the semidefinite relaxation (SDR) PhaseLift convex formulation and propose a successive and incremental nonconvex optimization algorithm, termed as IncrePR, to indirectly minimize the resulting convex problem on the cone of positive semidefinite matrices. Our proposed method overcomes the excessive computational cost of typical SDP solvers as well as the need of a good initialization for typical nonconvex methods. For Gaussian measurements, which is usually needed for provable convergence of nonconvex methods, restart-IncrePR solving three consecutive PhaseLift problems outperforms state-of-the-art nonconvex gradient flow based solvers with a sharper phase transition of perfect recovery and typical convex solvers in terms of computational cost and storage. For more challenging structured (non-Gaussian) measurements often occurred in real applications, such as transmission matrix and oversampling Fourier transform, IncrePR with several consecutive repeats can be used to find a good initial guess. With further refinement by local nonconvex solvers, one can achieve a better solution than that obtained by applying nonconvex gradient flow based solvers directly when the number of measurements is relatively small. Extensive numerical tests are performed to demonstrate the effectiveness of the proposed method.

Keywords Phase retrieval · Convex relaxation · Nonconvex optimization · Semidefinite programming

1 Beijing Computational Science Research Center, Beijing, China
2 Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
3 Department of Mathematics, University of California, Irvine, CA, USA
1 Introduction

1.1 The Problem

Consider the following \( m \) quadratic equations for \( x \in \mathbb{C}^n \)
\[ b_i = |a_i^* x|^2, \quad 1 \leq i \leq m, \] (1)
where the measurements \( b = [b_1, \ldots, b_m]^T \) and the design/sampling vectors \( \{a_i\}_{i=1}^m \in \mathbb{C}^n \) are known. Having information about \( |a_i^* x| \) means that the signs or phases of \( a_i^* x \) are missing. In general, problem (1) constitutes an instance of NP-hard nonconvex quadratic programming [14]. On the other hand, if the missing phases of the measurements \( \{b_i\}_{i=1}^m \) are recovered, the true solution, denoted by \( x^\diamond \) in this paper, can be reconstructed by solving a system of linear equations. Problem (1) is referred to the well-known (Fourier) phase retrieval problem when \( a_i \)'s are Fourier vectors. The phase retrieval problem is of paramount importance in many fields of physical sciences and engineering, e.g., X-ray crystallography [38], electron microscopy [39], X-ray diffraction imaging [46], optics [28] and astronomy [20] etc. In these applications, often one can only record intensity of the Fourier transform of a complex signal due to physical limitations of detectors, such as charge-coupled device (CCD) cameras.

In spite of its simple form and practical importance across various fields, solving problem (1) is challenging. Even checking the feasibility is an NP-hard problem. Much effort has recently been devoted to determining the number of such equations \( m \) necessary and/or sufficient for the uniqueness of the solution up to a global phase constant. It has been shown that, for the real case, \( m = 2n - 1 \) measurements with generic\(^1\) vectors \( a_i \)'s are sufficient and necessary for the uniqueness. For the complex case, the necessary condition for uniqueness requires \( m \geq 4n - 4 \) measurements [2,3]. Assuming that the system (1) admits an unique solution \( x^\diamond \) up to a phase constant, one needs an efficient and robust numerical algorithm to compute it. However, current state-of-the-art methods are either not well scaled with respect to the dimension of the signal, \( n \), in terms of computational complexity or can only guarantee local convergence when the number of measurements \( m \) is small. Our objective is to develop a conceptually simple while numerically efficient and stable algorithm to find the solution.

1.2 Prior Work

To develop provable algorithms and facilitate the analysis of phase retrieval, most recent works assume Gaussian random measurements [10,11,14,41,49], which are not very realistic in most applications. The measurement setup is said to be Gaussian model if the sampling vector \( a_i \sim \mathcal{N}(0, I) \) and \( a_i \sim \mathcal{CN}(0, I) = \mathcal{N}(0, I/2) + i\mathcal{N}(0, I/2) \) for the real and complex cases respectively. The randomness assumption is the key ingredient for the success of those proposed algorithms when the number of measurements is large enough. The only ambiguity of a solution to (1) for Gaussian model is a global phase shift, i.e., \( e^{i\theta} x^\diamond \) for some \( \theta \). For the classical oversampling Fourier phase retrieval, the ambiguity of solution becomes worse, which includes phase shift, translation shift and mirror image [22]. To mitigate the more severe ambiguity issue, redundant measurements are introduced, such as measurements with masks [7,19], short-time Fourier transform [24,44], and ptychography [44]. In practice, it is

\(^1\) For the definition of generic vectors, see the reference [2,3].
typically assumed that problem (1) admits an unique solution up to a global phase constant as long as the number of measurements is large enough. For most practical measurements, such as transmission matrix in the phase package [12] and Fourier transform, both the theory and those algorithms based on Gaussian model do not work as well as for Gaussian model.

Much effort has recently been devoted to devising provable phase retrieval solvers for Gaussian model [10,11,48–50]. These algorithms fall into two realms: convex and nonconvex ones. Convex approaches rely on the so-called lifting technique to turn the quadratic nonlinear constraints of $x$ into linear constraints of the matrix $X = xx^*$. Trace regularized semidefinite relaxation (SDR) formulation, termed as PhaseLift, can be designed [11], where low rank promoting trace regularization replaces the rank one constraint. Unlike nonconvex ones, the solution to SDR formulation is independent of initialization. However, such algorithms typically involve storing an $n \times n$ matrix variable and solving an SDP (semidefinite programming) problem, which suffer from prohibitive computation cost and storage for large scale problems. Another variant, PhaseCut, which is based on a different SDR reformulation, has the same scaling problem [21]. We say that the SDR for phase retrieval (1) is tight, if the optimal solution $X^\natural$ of SDR is unique and of rank one, i.e., $X^\natural = x^\natural x^\natural^*$. Then $x^\natural$ is the solution to (1) up to a phase shift. The tightness of PhaseLift with trace minimization for Gaussian model requires the number of measurements $m$ on the order of $n$ [8]. Another recent convex relaxation model [16,17] reformulated (1) as a linear programming problem in the natural parameter vector domain $C^n$. However, its performance depends on a good anchor vector, which has to form a small enough angle with the true solution and serves as a similar role as a good initial guess for nonconvex approaches to (1).

On the other hand, nonconvex approaches directly deal with (1) in the natural parameter space $C^n$. Based upon different formulations, recently proposed nonconvex approaches include: alternating projection methods on the basis of casting (1) as a two-set feasibility problem, such as alternating projection (AP, a.k.a. error reduction in [20]), HIO (hybrid input output) [20], RAAR (relaxed averaged alternating reflections) [31,34], graph projection splitting [29]; gradient flow based methods, such as Wirtinger flow (WF) and its truncated version (TWF) [10,14], truncated amplitude flow (TAF) [49] and reweighted amplitude flow (RAF) [50]; Wirtinger flow (WF) with an activation function [33]; message passing based methods, such as approximate message passing for amplitude-based optimization (AMP.A) [36]; and the prox-linear procedure via composite optimization [18]. These approaches lead to significant computational advantage over the lifting based SDR counterparts. However, due to the existence of possibly many stationary points, those gradient flow based nonconvex approaches in general are computationally intractable to find the true solution $x^\natural$ from an arbitrary initial guess. How to obtain an initialization that is good enough for these gradient flow based nonconvex methods is the key ingredient for a provable algorithm with convergence guarantee. In a nutshell, for gradient flow based nonconvex approaches for Gaussian model, there are generally two stages involved to obtain a global minimizer:

1. The initialization stage: for Gaussian model, a good initial guess can be computed efficiently by some routines, such as the spectral method [10,41], the truncated spectral method [14] and their variants, as well as the orthogonality-promoting method [49] (a.k.a. null method [13]) and the (re-weighted) maximal correlation method [50]. The obtained initial guess is proved to be good enough, provided the number of measurements $m$ is on the order of $n$.

2. The refinement stage: a refinement scheme, generally a gradient descent algorithm, is then used to push the initial guess toward to the global minimizer. The convergence to global
minimizer is shown through analysis of the population behavior of different objectives with the help of Gaussian model assumption [10,14,49,50].

For those state-of-the-art provable algorithms for phase retrieval, exact recovery for Gaussian model is guaranteed when the number of measurements \( m = cn \log n \) [10,16] or \( m = cn \) [8,14,49,50], where \( c \) is often a fixed but rather large constant independent of \( n \). The reported smallest numerical constant \( c \) for real case among the existing algorithms is 2 for RAF [50]. It is shown recently that for Gaussian model the intensity-based least-squares objective (used by WF) and its truncated modification with an activation function admits a benign geometric structure without spurious local minima from \( m = \mathcal{O}(n \log^3 n) \) [32,47] and \( m = \mathcal{O}(n) \) [33] measurements respectively. However, the number of measurements \( m = cn \) needs a large constant \( c \) far greater than 2 for real case. The required number of measurements for good recovery is even larger for non-Gaussian measurements, e.g., for ptychography [44].

When the number of measurements is large enough, a good initialization may be obtained and in this regime, those provable nonconvex algorithms often work well. For applications where the number of measurements is limited, a good initialization is difficult to obtain and nonconvex approaches usually fail to locate a satisfying solution. For example, due to the significant bias of the statistical properties of finite samples of Gaussian vectors, the initial guess becomes worse when \( m \) becomes smaller. Moreover, most initialization strategies only work for Gaussian model and fails for non-Gaussian measurements in most practices. The quality of initialization also downgrades with increased noise in the measurements. Without the help of a good initial guess, nonconvex approaches stagnate at local stationary points, which are far away from the true solution.

For practical Fourier phase retrieval, projection based solvers are the de facto approaches for their powerful performance. Projection based nonconvex solver unrelaxed HIO \((\beta = 1)\) and its variants can be recognized as nonconvex counterpart of Douglas-Rachford [4], which exhibits potential to avoid local minimum and shows powerful performance for phase retrieval. However, they are based on amplitude least-squares data fidelity, which does not match the noisy intensity measurements. This leads to worse reconstruction than our proposed method, as shown in numerical tests. Thus we focus on the optimization of intensity-based data fidelity formulation in this paper.

1.3 Contribution

Motivated by the global convergence of SDR PhaseLift approach, we revert to PhaseLift formulation of (1). Its tightness is proved for Gaussian model for which the rank one solution to PhaseLift provides the true solution to (1). To avoid the computation cost and storage of a typical SDP solver for PhaseLift, we propose a simple, well-scaled method to find the solution to PhaseLift by the decomposition of a Hermitian matrix \( X \) into \( YY^* \), where \( Y \in \mathbb{C}^{n \times p} \). Our method, termed as IncrePR, indirectly solves PhaseLift by a sequence of nonconvex optimizations in space \( \mathbb{C}^{n \times p} \) with incremental column number \( p \). Our incremental nonconvex optimization is similar to the approach for SDP proposed in literature [25]. However, it is the first time applied to solving the phase retrieval problem with the extension to Hermitian matrix. For Gaussian model, to pursue sharper phase transition for noiseless measurements, where phase transition is defined as the numerical limit of ratio \( m/n \) to successfully recover \( x^* \) from Eq. (1), we propose to solve three consecutive PhaseLift problems involving different parameters, and we call this procedure restart-IncrePR. The restart-IncrePR shows the sharpest phase transition among the state-of-the-art gradient flow based solvers for both real and complex cases. It achieves perfect recovery if the number of measurements \( m \geq 1.9n \).
and $m \geq 2.7n$ for real and complex cases respectively. It beats the theoretical guarantee of $m = 2n - 1$ [3] and $m \geq 4n - 4$ [2] for generic real and complex cases respectively. The extra decreasing of $m/n$ is due to the fact that Gaussian vector belongs to generic vector, so Gaussian model exhibits smaller phase transition than the above theoretical guarantee for generic vector measurements. For noisy measurements, restart-IncrePR locates good approximate solution with the help of global convergence and stability of PhaseLift [11].

Although the tightness of PhaseLift is no longer guaranteed for other type of more practical measurements, the rank one approximation of the solution to PhaseLift is generally close to the true solution to (1). We can then feed it to nonconvex approaches to recover a much better solution than that of direct nonconvex approaches from a random initialization. In other words, restart-IncrePR can be used as a good initializer. Numerical tests show that restart-IncrePR outperforms the state-of-the-art solver, reweighted amplitude flow (RAF) [50], for transmission matrix measurements (available in a public phase package [12]) when the number of measurements is small. For more difficult structured oversampling Fourier phase retrieval, restart-IncrePR is comparable to or better than the hybrid input-output (HIO) projection-based method [20], while all other gradient flow based methods built for Gaussian model fail to compute a meaningful reconstruction [7] when the number of measurements is not large enough.

1.4 Other Relevant Literature

The decomposition-based nonconvex optimization idea has been applied to solving linear cost function [6] and nonlinear cost function [25] with affine constraints on the cone of positive semidefinite matrices. It has also been extended to general (asymmetrical) low rank matrix recovery problems [15,40]. The first attempt to apply decomposition $X = YY^*$ within a fixed matrix dimension $\mathbb{C}^{n \times p}$ to solve PhaseLift [7] was proposed in [23], where the authors proposed a dynamic ranking decreasing strategy. In this work, we employ the incremental strategy and propose a post-verifiable condition for convergence and termination by establishing a relation between the stationary solutions to the nonconvex and convex problems respectively. Compared to the fixed dimension strategy, IncrePR performs better and terminates quickly with very few incremental steps in our experiments.

To develop a scalable nonconvex algorithm, authors in [51] considered applying the conditional gradient method to another SDR problem of (1) with constraint on the maximum trace of $X$. To avoid updating of the full matrix $X$, they derived a novel method, termed as SketchyCGM, driven by the measurement vectors and sketchy of the underlying low rank matrix during iterations. Compared to the standard PhaseLift solver, the algorithm outperforms when the number of measurements $m$ is large enough, and underperforms when $m$ is close to the phase transition point [12].

1.5 Notation and Organization

We use $\mathbb{H}^n$ and $\mathbb{H}^n_+$ to denote the set of complex and positive semidefinite $n \times n$ Hermitian matrices respectively. $\langle U, V \rangle = \text{Re}(\text{Tr}(U^*V))$ denotes the standard inner product in $\mathbb{H}^n$, where $\{\cdot\}^*$ is the conjugated transpose. We use bold font to denote vectors and bold capital font for matrices. $a \odot b$ and $a/b$ denote the entrywise multiplication and division between vectors $a$ and $b$, respectively. The organization of the paper is as follows. We first review the PhaseLift formulation and cite known results of exact and stable recovery for noiseless and noisy Gaussian models for completeness. Then we describe our incremental nonconvex approach
(IncrePR) for solving the PhaseLift problem. Combined with a restart strategy, IncrePR can achieve optimal phase transition for Gaussian model or obtain a good initialization for non-Gaussian measurements. Various numerical tests and comparisons are presented to demonstrate the performance of our proposed method. Conclusion is presented at the end.

**2 PhaseLift: A Review**

Note that our proposed method, IncrePR, works for both real and complex cases. For generality, we formulate the problem and the algorithm for the complex case.

The problem (1) is to find a vector \(x \in \mathbb{C}^n\) subject to the system of \(m\) quadratic equations \(|a_i^*x|^2 = b_i, i = 1, 2, \ldots, m\). It can be viewed as an instance of quadratic constraint quadratic programming (QCQP). Following the classical Schor’s relaxation [35], the vector variable \(x\) can be lifted into a matrix variable \(X = xx^*\) [7]. The nonlinear quadratic constraints for \(x\) can be cast as linear constraints of \(X\), i.e.,

\[|a_i^*x|^2 = \text{Tr}(A_iX) = b_i, \quad A_i = a_i a_i^*.\]

Thus problem (1) can be formulated as the following matrix problem

find \(X \in \mathbb{C}^{n \times n}\)

s.t. \(\text{Tr}(A_iX) = b_i, i = 1, 2, \ldots, m,\)

\(X \succeq 0, \quad \text{rank}(X) = 1.\)

The problem is still NP-hard due to nonconvexity of the rank one constraint. It can be equivalently transformed into a rank minimization problem:

\[
\min_X \quad \text{rank}(X)
\]

\[
\text{s.t.} \quad \text{Tr}(A_iX) = b_i, i = 1, 2, \ldots, m, \quad X \succeq 0.
\]

When the rank minimization is further relaxed to trace minimization, the new problem is of the following form and called semidefinite relaxation (SDR) or more specifically, PhaseLift,

\[
\min_X \quad \text{Tr}(X)
\]

\[
\text{s.t.} \quad \text{Tr}(A_iX) = b_i, i = 1, 2, \ldots, m, \quad X \succeq 0.
\]

(2)

Candès et al showed that the solution to problem (2) is of rank one and its rank one decomposition factor is exactly the solution to problem (1) with high probability for Gaussian model, provided the number of measurements \(m = \mathcal{O}(n)\) [8]. Furthermore, the solution is also stable for noisy measurements, no matter what distribution the noise is drawn from. The trace minimization in (2) is unnecessary in some cases. The most obvious case is that the intensity measurements determine the trace of \(X\), such as the Fourier phase retrieval and the case that the identity matrix is in the span of \(\{A_i\}_{i=1}^m\). When \(m = \mathcal{O}(n \log n)\) for Gaussian model, the trace minimization is unnecessary [16]. Note that there exists a gap for \(m\) between \(\mathcal{O}(n)\) and \(\mathcal{O}(n \log n)\), which is the critical transition of the tightness between the trace minimization problem (2) and the set feasibility problem, i.e., problem (2) without trace minimization. The available theoretical tightness and phase transition results are for
Gaussian model. For other type of measurements, the optimal solution to (2) may not be of rank one while its rank one approximation is often close to the rank one matrix $x^nx^n$.  

Due to possible noise in measurements, Candès et al [7] proposed a first-order scheme to solve the trace regularized SDP problem:

$$\begin{align*}
\min_X & \quad f_\lambda(X) = f_0(X) + \lambda \Tr(X), \\
\text{s.t.} & \quad X \succeq 0,
\end{align*}$$

where the parameter $\lambda$ trades off between the data fidelity term $f_0(X)$ and the low rank promoting trace term. Note that the theoretical guarantee is proved for the (2), i.e., PhaseLift with trace minimization, while problem (3), i.e., PhaseLift with trace regularization, is considered in numerical implementation. Hereafter, we call (3) PhaseLift for short without specification.

The data fidelity term $f_0(X)$ depends on the noise model. For least-squares loss, the objective function in matrix variable $X$ is

$$f_0(X) = \frac{1}{2m} \| A(X) - b \|^2.$$  

(4)

For Poisson likelihood estimation, the objective function is

$$f_0(X) = \langle 1, A(X) - b \circ \log(A(X)) \rangle = \sum_{i=1}^m \Tr(A_iX) - b_i \log(\Tr(A_iX)).$$  

(5)

The linear operator in (4) and (5) is defined as

$$A : \mathbb{H}^n \mapsto \mathbb{R}^m, A(X) = [\Tr(A_1X), \Tr(A_2X), \ldots, \Tr(A_mX)]^T.$$

And its adjoint operator is denoted as $A^* : \mathbb{R}^m \mapsto \mathbb{H}^n, A^*(b) = \sum_{i=1}^m b_iA_i$.

Problem (3) was solved by first-order method with matrix variable $X$ in [7], whereas the storage of $X \in \mathbb{H}^n$ and the projection onto $\mathbb{H}^n_+$ are intractable for large-scale problems. Motivated by decomposition of variable $X$ into $YY^*$, where $Y \in \mathbb{C}^{n \times p}$ and $p \ll n$ due to the low rank property of $X$, we propose an incremental nonconvex optimization approach to indirectly solve (3) with significantly reduced computational cost and storage.

### 3 Nonconvex Solver for PhaseLift

With the choices of objectives (4) or (5), problem (3) is a smooth convex optimization problem in the cone of $\mathbb{H}^n_+$. We substitute expression $X = YY^*$ with $Y \in \mathbb{C}^{n \times p}$ into (3), then the dimension of optimization variable decreases from $n^2$ to $np$. And the positive semidefiniteness of $X$ is automatically satisfied by the decomposition.

The resulting nonconvex problem in $Y$ is

$$\min_{Y \in \mathbb{C}^{n \times p}} \bar{f}_\lambda^p(Y) = f_0(YY^*) + \lambda \| Y \|^2_F.$$  

(6)

If $p = 1$, (6) becomes the $\ell_2$ regularized Wirtinger flow (a.k.a. intensity-based least-squares) in the natural parameter space $\mathbb{C}^n$.

The computation storage and cost of solving (6) is significantly less than solving (3) when $p \ll n$. Once $p$ is fixed, minimizing $\bar{f}_\lambda^p(Y)$ will locate a stationary point. The question is how to change $p$ appropriately to help us to find the true solution to (3) or a good approximation of it. The answer can be obtained from the analysis of the optimality conditions of the two associated problems.
3.1 Connection Between the Convex and Nonconvex Formulations

The possibility of indirectly solving problem (3) by solving (6) is implied by the relation between their first and second order optimality conditions. For notation simplicity, we omit the script $\lambda$ and $p$ in $f_\lambda(X)$ and $\tilde{f}_p(Y)$ when $\lambda$ and $p$ are fixed. Given a function $f : \mathbb{H}^n \to \mathbb{R} : X \mapsto f(X)$, we define the function

$$\tilde{f} : \mathbb{C}^{n \times p} \to \mathbb{R} : Y \mapsto \tilde{f}(Y) = f(YY^*).$$

For a differentiable function $f$, the notation $\nabla_X f(X_0)$ refers to the gradient of $f$ at $X_0$ with respect to the variable $X$ in the Wirtinger sense [10,27], since here we deal with the real-valued function of complex variables, i.e.,

$$[\nabla_X f(X_0)]_{i,j} = \frac{\partial f}{\partial X_{i,j}}(X_0).$$

In the real case, the gradient definition becomes the standard one.

The directional derivative of $f$ at $X_0$ in a direction $Z \in \mathbb{H}^n$ is defined as

$$D_X f(X_0)[Z] = \lim_{t \to 0} \frac{f(X_0 + tZ) - f(X_0)}{t}.$$ 

If $f$ is differentiable, we have

$$D_X f(X_0)[Z] = \langle \nabla_X f(X_0), Z \rangle,$$

and

$$\nabla_Y \tilde{f}(Y) = 2\nabla_X f(YY^*)Y, \quad D_Y \tilde{f}(Y)[Z] = \langle 2\nabla_X f(YY^*)Y, Z \rangle, \quad \forall Y, Z \in \mathbb{C}^{n \times p}.$$ 

We state the first-order KKT conditions for problems (3) and (6) in the following lemmas. The statements are standard for smoothing optimization with an inequality, the proofs are straightforward in the complex gradient sense. Thus we omit the proofs.

**Lemma 1** A global minimizer of problem (3) is a Hermitian matrix $X \in \mathbb{H}^n$ such that the first-order optimality conditions hold:

$$(\nabla_X f_0(X) + \lambda I)X = 0$$

$$\nabla_X f_0(X) + \lambda I \succeq 0$$

$$X \succeq 0.$$ (7)

The optimality conditions are sufficient and necessary for our convex optimization (3).

**Lemma 2** If $Y$ is a stationary point of problem (6), then it holds that

$$(\nabla_X f_0(YY^*) + \lambda I)Y = 0.$$ 

Comparing Lemma 1 and 2, we have the following relation between the stationary solutions of problems (3) and (6).

**Theorem 1** A stationary point $Y$ of the nonconvex problem (6) provides a global minimizer $YY^*$ of problem (3), if the matrix

$$S_Y = \nabla_X f_0(YY^*) + \lambda I$$

is positive semidefinite.
The stationary point $Y$ of problem (6) may be a saddle point. Here we state the second-order optimality condition for the unconstrained problem.

**Lemma 3** For a local minimizer $Y \in \mathbb{C}^{n \times p}$ of (6), it holds that
\[
\text{Tr}(Z^* D_Y \left( \nabla_Y (f_0(Y Y^*) + \lambda \|Y\|_F^2) \right) [Z]) \geq 0
\]
for any matrix $Z \in \mathbb{C}^{n \times p}$.

**Lemma 4** For any matrix $Z \in \mathbb{C}^{n \times p}$ such that $Y Z^* = 0$, the following equality holds:
\[
\frac{1}{2} \text{Tr}(Z^* D_Y \left( \nabla_Y (f_0(Y Y^*) + \lambda \|Y\|_F^2) \right) [Z]) = \text{Tr}(Z^* S_Y Z).
\]

**Proof** According to the definition of directional derivative, we have
\[
\frac{1}{2} \text{Tr}(Z^* D_Y \left( \nabla_Y (f_0(Y Y^*) + \lambda \|Y\|_F^2) \right) [Z]) = \text{Tr}(Z^* S_Y Z),
\]
where we use the equality $\text{Tr}(AB) = \text{Tr}(BA)$. \qed

The natural question is how to choose the column number $p$ of $Y$. We first give a sufficient condition by the following Theorem [25, which we extended to Hermitian matrices in a straightforward way.

**Theorem 2** ([25, Theorem 7]) A local minimizer $Y$ of problem (6) provides a global minimizer $X = YY^*$ of problem (3) if it is rank deficient.

**Proof** Assume the rank of $Y \in \mathbb{C}^{n \times p}$ is $r$, we have the full rank decomposition $Y = \tilde{Y} M^*$, where $\tilde{Y} \in \mathbb{C}^{n \times r}$ and $M \in \mathbb{C}^{r \times r}$. We take $M_\perp \in \mathbb{C}^{p \times (p-r)}$ as an orthogonal basis for the orthogonal complement of the column space of $M$. For any matrix $\tilde{Z} \in \mathbb{C}^{n \times (p-r)}$, the matrix $Z = \tilde{Z} M_\perp^*$ satisfies $Y Z^* = 0$. It follows that $\text{Tr}(Z^* S_Y Z) \geq 0$ for all matrices $Z$. Thus the matrix $S_Y$ is semi-definite and $X = YY^*$ is a global minimizer of problem (3). \qed

The worst case scenario is $p = n$, for which any local minimizer of problem (6) provides a global minimizer $X = YY^*$ of problem (3). Because, if $Y$ is rank deficient, then the local minimizer provides a global minimizer of problem (3). Otherwise, $Y$ is full rank, then the matrix $S_Y$ becomes zero by the equality $S_Y Y = 0$ from Lemma 2. Intuitively, for PhaseLift, $p$ should not be large. In particular, the unique solution to the PhaseLift with trace regularization (3) should be low rank if PhaseLift with trace minimization (2) is tight, e.g., for Gaussian model. In [23], $p$ was set to below 10 initially and then decreased further dynamically during iterations.

Here we propose to solve (6) starting from $p = p_0$ and a random initial guess $Y_0 \in \mathbb{C}^{n \times p_0}$. However, one may stagnate at a stationary point due to nonconvexity. The key idea is to escape from the current stationary point and decrease the objective function of (3) further by increasing the column number $p$ of the decomposition factor $Y$. Hence we construct the initialization $Y_0 \in \mathbb{C}^{n \times (p_0+1)}$ from the stationary point to problem (6) with $Y \in \mathbb{C}^{n \times p_0}$ that decreases the objective function. We successively increase $p$, until the minimum of the objective function is achieved at the global minimizer. In other words, we indirectly solve the original convex problem (3) by solving a sequence of nonconvex optimization problems (6).

The remaining question is what is the criterion for one to stop increasing $p$ and the iteration.
Theorem 1 provides the post-verifiable condition for our purpose. Although the feasible set may be expanded due to the increase of column number \( p \), finally we shall obtain the rank one solution, which is the global minimizer of the convex PhaseLift problem in the regime of \( m = \mathcal{O}(n \log n) \) for Gaussian measurements, regardless of intermediate iteration with higher rank factorization.

### 4 The Incremental Approach

When we solve the nonconvex problem (6) and reach a stationary point \( Y \), it may be a local minimizer or a saddle point. If condition in Theorem 1 is not satisfied, i.e., the matrix \( SY = \nabla_X f_0(YY^*) + \lambda I \) is not positive semidefinite, it means that we have not found a global minimum of the convex problem (3) yet. The incremental algorithm is designed to monotonically decrease the objective function (3) by increasing \( p \).

Starting from \( p = p_0 \) and assume we arrive at a stationary point \( Y \in \mathbb{C}^{n \times p} \) of problem (6). If \( SY = \nabla_X f_0(YY^*) + \lambda I \) is positive semidefinite, then \( YY^* \) is a solution to (3) and one can terminate. Otherwise we increase \( p \) to \( p + 1 \) and minimize problem (6) by constructing an initial point \( Y_p = [Y[0^p \times 1]] \in \mathbb{C}^{n \times (p+1)} \). It should be noted that if \((\nabla_X f_0(YY^*) + \lambda I)Y = 0,\) then \((\nabla_X f_0(Y_pY_p^*) + \lambda I)Y_p = 0,\) i.e., \( Y_p \) is also a stationary point of problem (6) in \( \mathbb{C}^{n \times (p+1)} \).

Since \( SY \) must have at least one negative eigenvalue, from Theorem 1 and Lemma 4, the matrix \( Z = [0^p \times p][v] \in \mathbb{C}^{n \times (p+1)} \), where \( v \) is the eigenvector responding to the smallest eigenvalue of \( SY \), satisfies

\[
\frac{1}{2} \text{Tr}(Z^*DYf(\nabla_X f_0(Y_pY_p^*) + \lambda I)(Z)) = v^TSY_pv \leq 0,
\]

since \( Y_pZ^* = 0 \). Thus \( Z \) is a descent direction of (6) at \( Y_p \) in \( \mathbb{C}^{n \times (p+1)} \). This allows us to escape the saddle point \( Y_p \) along the descent direction and minimize the objective function further through one backtracking step. After one decreasing step, we reach a new point \( Y_0 \in \mathbb{C}^{n \times (p+1)} \) and it provides a good initial guess for the next optimization problem (6) for \( Y \in \mathbb{C}^{n \times (p+1)} \). This procedure is done repeatedly till the termination of the algorithm. Note that simply padding a zero column to current stationary point provides a good starting point for the next stage. We summarize all the above elements in Algorithm 1. We call the incremental nonconvex optimization algorithm to solve PhaseLift (3) for phase retrieval IncrePR. The parameter \( \epsilon \) is a threshold on the eigenvalues of \( SY \) to decide the nonnegativity of this matrix. \( \epsilon \) is chosen depending on the problem and desired accuracy.

Based on Theorem 1, positive semidefiniteness of the matrix \( SY \) at a stationary point of the nonconvex problem (6) indicates global convergence of IncrePR. Since all eigenvalues of Hermitian matrix are real-valued, we just have to compute the smallest eigenvalue of \( SY \). In particular, we use the LOBPCG routine to locate the smallest eigenvalue without the need to form the matrix \( SY \) [26].

In general, one can employ two explicit criteria to terminate our incremental algorithm in terms of accuracy and the computational cost respectively. The first criteria is to check if the matrix \( SY \) is positive semidefinite at a stationary point of the nonconvex problem (6) within the threshold \( \epsilon \). The smaller the absolute value of the smallest negative eigenvalue of matrix \( SY \), the closer the computed solution is to a global minimizer of (3). In other words, absolute value of the smallest eigenvalue of matrix \( SY \) can be used as an indicator of the progress of the incremental algorithm and can be used to monitor the convergence. The upper limit of \( p \) for \( Y \) can be used as another criterion to forcibly limit the computation and storage.
Algorithm 1 Incremental algorithm (IncrePR) for solving PhaseLift (3)

Input: Initial column number \( p_0 \), initial guess \( Y_0 \in \mathbb{C}^{n \times p_0} \) and tolerance parameter \( \epsilon \).

Output: Solution \( X = Y_\rho Y_\rho^* \) to problem (3).

1: Initialize \( p = p_0 \)
2: repeat
3: Apply an optimization method to find a stationary point \( Y_\rho \) of problem (6) from initial guess \( Y_0 \).
4: Find the smallest eigenvalue \( \nu_{\min} \) of \( SY_\rho \) defined in (8).
5: if \( \nu_{\min} \geq -\epsilon \) then
6: break
7: else
8: Find the corresponding eigenvector \( v \) of the matrix \( SY_\rho \).
9: \( p \leftarrow p + 1 \), \( Y_\rho \leftarrow [Y_\rho|0] \)
10: Apply one descent step along descent direction \( Z_\rho = [0|v] \) (such as backtracking line-search), arrive at point \( Y_0 \).
11: end if
12: until convergence

Cost to solve the nonconvex problem (6). It is worth to note that our incremental approach always decreases the objective function (3) no matter what the error tolerance \( \epsilon \) is for each nonconvex problem. In particular, one does not need to set it too small at the beginning and then decrease it during the incremental process.

Compared to the dynamical rank decreasing strategy used in [23], our incremental strategy is more effective and natural. In our algorithm, one only needs to solve a sequence of unconstrained nonconvex problems (6). For each nonconvex problem, one simply use a first-order method to find a stationary point and use Theorem 1 to decide whether to continue or terminate. However, if only a stationary point is located for each nonconvex problem, Algorithm 1 may not find the solution and terminate even when \( p \) reaches \( n \). In this worst case scenario, one should compute a local minimizer of (6) by second-order methods and apply Theorem 2. Compared to first-order method, to find a local minimizer of (6), second-order optimality condition has to be checked when a stationary point is reached and, if the stationary point is a saddle point, a negative curvature direction needs to be found. For the application of phase retrieval, the worst case scenario happen very rarely. Thus we typically use a first order method to locate a stationary point in our algorithm. Although, applying first-order method may lead to a larger termination \( p \), numerical experiments show that both first-order and second-order methods for (6) work well and terminate with small \( p \) for phase retrieval applications. In the next section, we give the explicit gradient and Hessian for two data fidelity functions (4) and (5) and briefly review optimization methods for completeness.

5 Unconstrained Optimization

In this section, we provide components of the optimization scheme for solving the following nonconvex unconstrained problem:

\[
\min_{Y \in \mathbb{C}^{n \times p}} ~ \tilde{f}(Y) = f_0(YY^*).
\]

For the trace regularized minimization (6) \((\lambda \neq 0)\), the gradient and Hessian have an additional simple term. For simplicity we skip the trace term in the following derivation.
For the nonconvex objective $\tilde{f}_0$ of (4), its gradient and Hessian along the direction $\xi_Y$ are given by

$$\nabla \tilde{f}_0(Y) = \frac{1}{m} A^*(A(YY^*) - b)Y,$$

$$\nabla^2 \tilde{f}_0(Y)[\xi_Y] = \frac{1}{m} A^*(A(YY^*) - b)\xi_Y + \frac{1}{m} A^*(A(Y\xi_Y + \xi_Y Y^*))Y.$$ 

For the nonconvex objective $\tilde{f}_0$ of (5), its gradient and Hessian along the direction $\xi_Y$ are given by

$$\nabla \tilde{f}_0(Y) = 2A^* \left( 1 - \frac{b}{A(YY^*)} \right) Y,$$

$$\nabla^2 \tilde{f}_0(Y)[\xi_Y] = 2A^* \left( 1 - \frac{b}{A(YY^*)} \right) \xi_Y + 2A^* \left( \frac{b}{(A(YY^*))^2} \circ \left( A(Y\xi_Y + \xi_Y Y^*) \right) \right) Y.$$ 

The involved computation $A(YY^*)$ and $A^*(\cdot)$ is done by matrix-vector multiplication. When dealing with structured measurements, such as Fourier phase retrieval and the associated coded diffraction pattern [9], the computation can be done more efficiently.

Note that the function $\tilde{f}(Y)$ is invariant under an orthonormal transformation. Thus the stationary points are not isolated, which can cause complications for second-order optimization methods [1]. To locate a local minimizer for a nonconvex optimization, a second-order method, such as trust-region method [5], is needed. Since our nonconvex optimization is invariant up to an orthonormal transformation, one should consider the Riemannian optimization within the quotient space, e.g., the Manopt package [5], which only requires inputs of the gradient and Hessian in Euclidean space.

On the other hand, the invariance does not affect first-order methods. To avoid computing projections onto the quotient space at each iteration step, where Sylvester equation and Newton equation for Riemannian optimization need to be solved [5], one can use first-order line search method to locate a stationary point for each nonconvex optimization (6). First-order line search method constructs a sequence,

$$Y_{k+1} = Y_k + \alpha_k d_k, \quad (9)$$

where $d_k \in \mathbb{C}^{n \times p}$ is a decent direction and $\alpha_k \in \mathbb{R}$ is the step size along the descent direction at $k$th iteration. The simplest choice, a.k.a. gradient descent, of $d_k$ is $-\nabla \tilde{f}(Y_k)$. Although first-order method for (6) with fixed $p$ only converges to a stationary point, it is sufficient since the incremental approach decreases the function (3) by increasing $p$ to $p + 1$. Other more efficient methods with different choices of descent directions, such as CG and LBFGS [30,32] methods, can also be used.

There is no explicit line search solution for step size $\alpha_k$ for the likelihood estimation function (5). The inexact line search step size should satisfy the well known Wolfe conditions (whereas $g$ denotes $\nabla \tilde{f}$):

$$\tilde{f}(Y_k + \alpha_k d_k) \leq \tilde{f}(Y_k) + c_1 \alpha_k \text{Re}(d_k^* g_k) \quad (10a)$$

and

$$\text{Re}(d_k^* g_{k+1}) \geq c_2 \text{Re}(d_k^* g_k), \quad (10b)$$

where condition $0 < c_1 < c_2 < 1$ is satisfied. Equations (10a) and (10b) are known as the sufficient decrease and curvature condition respectively [42]. For intensity-based least-squares function (4), one can find the step size $\alpha_k$ from exact line search by solving a cubic
equation. A similar routine can be found in [32] to solve the nonconvex optimization (6) with $p = 1$.

6 Solving Phase Retrieval with Restart-IncrePR

For noiseless Gaussian model, when the number of measurements $m = \mathcal{O}(n \log n)$, the parameter $\lambda$ in (3) can be set to zero, one can solve (3) by IncrePR with $p_0 = 1$. If $m$ is between $\mathcal{O}(n)$ and $\mathcal{O}(n \log n)$, PhaseLift with trace minimization (2) is tight. We can solve a series of problems (3) with different $\lambda$’s to approach the solution to (2). Standard continuity approach of solving (3) with a series of decreasing $\lambda$’s can be exploited, where solution to current $\lambda$ is the initial guess for next $\lambda$. The continuity approach is used to select the optimal regularization parameter for low rank matrix recovery problem [40].

For Gaussian model, if the number of measurements $m = \mathcal{O}(n)$, the effect of trace regularization in (2) becomes important. We propose the following three-stage algorithm with a restart strategy. In stage I, we start with solving (3) with $\lambda = 0$ from a random initial guess. However, the solution $X_1$ is only aiming to fit the data and not of rank one in general, since the solution to the convex problem (3) with $\lambda = 0$ may not be unique from $\mathcal{O}(n)$ measurements. Then, In stage II, (3) with $\lambda = \lambda_0$ is solved starting from $X_1$ where trace regularization is used to promote low rank. The solution $X_2$ should be close to the solution to problem (2). Note that in this regime of $m = cn \log n$, we can directly solve PhaseLift with $\lambda = 0$ by IncrePR. However, the constant $c$ is unknown. Thus our strategy is to avoid the possible extra computational cost. Stage III can be regarded as a refinement stage. From a rank one approximation of $X_2$, the refinement can be done by a nonconvex solver, either alternating projection based or gradient flow based solvers can be applied, or another convex solver, such as PhaseLift with $\lambda = 0$. The restart-IncrePR can be used repeatedly. The repeated approach is applied to Fourier phase retrieval. To pursue sharper phase transition, solving PhaseLift with $\lambda \sim 0$ is necessary. Thus we call solving successive PhaseLift problems restart-IncrePR. The detailed description of the three-stage algorithm for synthetic Gaussian model is presented in Algorithm 2, where the refinement is done by solving PhaseLift. In other words, we adopt IncrePR to refine the solution instead of the typical SDP solver. SDP solver operates on an $n \times n$ matrix, while IncrePR operates on an $n \times p$ matrix, and $p$ is generally less than ten, as demonstrated in numerical tests. At each iteration, the computation complexity of the eigenvalue decomposition of SDP solver is $\mathcal{O}(n^3)$ for projection onto $\mathbb{H}_n^+$. For IncrePR, the main cost is checking the positive semi-definiteness of $SY_p$ defined in (8), i.e. computing its smallest eigenvalue and the corresponding eigenvector. This operation has to be performed each time when $p$ is increased. The restart-IncrePR can also achieve sharper phase transition with $\mathcal{O}(n)$ measurements. Since $\lambda$ is set to $10^{-3}$ in [7], the trace regularization is not strong enough to find a good approximate solution from relatively small number of measurements. Hence its best phase transition requires $\mathcal{O}(n \log n)$ measurements [16].
Algorithm 2 restart-IncrePR with convex refinement for Gaussian model

**Input:** Initial $p_0 = 1$, penalty parameter $\lambda_0$ and initialization $Y_0 \in \mathbb{C}^n$.

**Output:** An approximate solution $x$ to problem (1).

1: Initialize $p = p_0$ and $\lambda = 0$, use IncrePR to solve (3) and obtain solution $X_1 = Y_1 Y_1^*$.
2: if $\text{rank}(Y_1) = 1$ then
3: Set $x = Y_1$
4: break
5: else
6: Set $p_0$ to the column number of $Y_1$ and $\lambda = \lambda_0$. Initialize $Y_0 = Y_1$.
7: Use IncrePR to solve (3) and obtain solution $X_2 = Y_2 Y_2^*$.
8: Extract the best rank one approximation $y_2$ to $Y_2$.
9: Set $p_0 = 1$, $\lambda = 0$, and initialize $Y_0 = y_2$.
10: Refinement: Use IncrePR to solve (3) and obtain solution $X_3 = Y_3 Y_3^*$.
11: Extract the rank one approximation $x$ of $Y_3$.
12: end if

For other types of more practical measurements, such as the transmission matrix and Fourier measurements, the tightness of (2) is not guaranteed. However, one can still use IncrePR to obtain the global minimizer of SDR and use its rank one approximation as a good approximation to the global minimizer of the nonconvex problem (6) with $p = 1$. In particular, one can optimize (3) using restart-IncrePR (Algorithm 2) a few times, where the current solution is used as the initial guess for the next round of Algorithm 2. If needed, the rank one approximation obtained from IncrePR can be used as an initial guess for a nonconvex solver for the nonconvex problem (6) with $\lambda = 0$ for further improvement.

7 Numerical Tests

7.1 Gaussian Model

For Gaussian model, a benchmark for phase retrieval (1) is to compare the phase transition of each algorithm. The phase transition is the critical point of the ratio of the number of measurements $m$ to the length of the signal $n$. When $m/n$ is above the phase transition, the unique solution to (1) can always be located by the algorithm.

We would like to illustrate the empirical phase transition of restart-IncrePR (Alg. 2) for Gaussian model. This is done by Monte-Carlo simulations. Here we fix the length of signal $n = 400$ and vary the number of measurements $m = \{1n, 1.1n, \ldots, 5n\}$. For each pair $(n, m)$, we generate 50 instances of problem (1). For each instance, we generate a 1-dimension real/complex signal of length $n = 400$ at random, and the sampling vectors $a_r$’s drawn from real/complex Gaussian distribution and obtain the measurement intensity data. We then compute the recovery rate, i.e., the percentage of the successful recoveries for the 50 instances. We use the relative error to measure the success of recovery. We denote the solution computed by restart-IncrePR as $x$. The relative error is defined as

$$\text{relerr} = \min_{|c|=1} \frac{\|c x - x^\natural\|_2}{\|x^\natural\|_2},$$

where $c$ is to get rid of the effect of the constant phase shift for phase retrieval problem.

We consider a signal to be perfectly recovered if the relative error (11) is below $10^{-5}$. The least-squares data fidelity (4) is used in problem (3). For real and complex cases, $\lambda_0$ in Alg 2 is set to $10^2$. The maximum iteration for each nonconvex problem is set to 1000 and 3000.
for real and complex models respectively. For real case, the termination $\epsilon$ for incremental nonconvex approach is set to 100, 500 and 1 for the three convex problems respectively. For complex case, the termination $\epsilon$ is set to 800, 1000 and 1 respectively. We do not need to set $\epsilon$ too small for the first two stages, since they mainly serve to provide a good initial guess for the last stage through the restart strategy. Before exploring the phase transition, we first study the effect of first- and second-order methods on the termination column number $p$ for IncrPR.

### 7.1.1 Termination Column Number

We run 10 Monte Carlo trials and compare the performance of the conjugate gradient (CG) method with that of the second-order trust-region (TR) method. In Table 1, we list the termination column number $p$ of restart-IncrPR (the three-stage Algorithm 2) at each stage, where the average, minimum and maximum termination column number $p$ (in parenthesis separated by commas) are shown. For real case, we test three groups with the number of measurements $m = 2n, 2.5n, 3n$ respectively. For complex case, the number of measurements $m = 3n, 3.5n, 4n$ are considered. The termination $p$ for CG is typically larger than that of TR, since the first-order method only locates a stationary point, while TR finds a local minimizer, for the nonconvex problem (6). For the challenging case of minimal measurement limit, $m = 2n$ for real case and $m = 3n$ for complex case, there are tests for which the termination $p > 1$ at the third stage. It shows the necessity of utilizing IncrPR at the refinement stage to solve the convex problem. If nonconvex formulation (6) with fixed $p = 1$ is used to refine the solution, it may still stagnate at a stationary point.

### 7.1.2 Phase Transition

In the following experiments, we study the phase transition of restart-IncrPR (Alg. 2) with comparison to other algorithms. We still use the least-squares data-fidelity (4) and apply the conjugate gradient method to solve each nonconvex optimization. We compare restart-IncrPR (Algorithm 2) using random initial guess with state-of-the-art gradient flow based nonconvex methods, including reweighted amplitude flow (RAF) [50], truncated amplitude flow (TAF) [49], truncated Wirtinger flow (TWF) [14] as well as Wirtinger flow (WF) [10] with the same random initial guess as IncrPR or a good initial guess provided by the state-of-the-art reweighted maximal correlation method [50]. For the performance of projection-based nonconvex solvers, we refer to the numerical experiments in [29]. Here we only compare with nonconvex algorithms, since IncrPR and PhaseLift should share the same phase transition if the restart strategy is also utilized in PhaseLift. For fair comparison, we set the parameters involved in each method to their default values recommended in the literature. The maximum iteration number of the nonconvex solvers is set to 3000 and 6000 for real and complex cases respectively.

For each pair $(n, m)$, we compute the recovery rate of each method by averaging 50 Monte Carlo trials. Phase transition of each method is depicted in Fig. 1. The restart-IncrPR (Alg. 2) shows the sharpest phase transition for both real and complex cases. It achieves perfect recovery from $m \geq 1.9n$ and $m \geq 2.7n$ measurement data for the real and complex cases respectively. It beats the theoretical guarantee of $m = 2n - 1$ [3] and $m \geq 4n - 4$ [2] for generic real and complex cases respectively. The numerical results should not be alarming since the theoretical result is for generic measurement vectors, which include Gaussian model as a specific case. As stated in [50], the reweighted amplitude flow (RAF) can achieve perfect
Table 1 Comparison of termination column number $p$ of restart-Increment (Algorithm 2) by first/second-order method (CG/TR).

|       | $m = 2n$ | $m = 3n$ | $m = 4n$ |
|-------|----------|----------|----------|
| Real  |          |          |          |
|       | stage    | stage    | stage    |
| I     | 4.6 (4, 5) | 2.6 (2, 3) | 3.9 (3, 5) |
| II    | 6.3 (5, 8) | 2.9 (2, 3) | 3.9 (3, 5) |
| III   | 1.1 (1, 2) | 1.0 (1, 1) | 1.0 (1, 1) |
|       | CG       | TR       | CG       |
|       |          |          |          |
| Complex |        |          |          |
| I     | 2.2 (2, 3) | 3.3 (2, 4) | 3.3 (2, 4) |
| II    | 2.7 (2, 4) | 2.0 (2, 2) | 1.0 (1, 1) |
| III   | 1.4 (1, 2) | 1.2 (1, 2) | 1.0 (1, 1) |

The average, minimum and maximum termination $p$'s over 10 runs are shown for real and complex Gaussian models.
Fig. 1 Phase transition for noiseless Gaussian model. The legend with letter 'r' stands for starting from the same random initial guess as restart-IncrePR.

Fig. 2 Phase transition for noiseless Gaussian model with $n = 1000$.

Recovery for signal with dimension $n \geq 2000$ from $m = 2n$ measurement data, while it may encounter difficulty for small dimension signal ($n = 400$ here), as shown in Fig. 1. It results from the poor initialization by the reweighted maximal correlation method, which requires large signal dimension (e.g., $n \geq 2000$). It is also obvious that the performance of nonconvex methods degrades significantly when starting from a random initialization. The restart-IncrePR (Alg. 2) outperforms all the compared nonconvex solvers markedly from random initialization. This is due to the advantage of convex PhaseLift model. As suggested by the reviewer, we also test the case of signal’s length $n = 1000$. In Algorithm 2, the maximum iteration for each nonconvex subproblem is set to 1200, and we set $\lambda_0 = 5e + 3$. When solving each PhaseLift problem in the three stages, we set different tolerance levels. We set $\epsilon_1 = 1e + 3, \epsilon_2 = 3e + 3$ and $\epsilon_3 = 1e + 2$. In addition, to further save the computational cost, we set the upper allowable column number $p$ for each PhaseLift problem to 3, 5 and 2 respectively. We illustrate the recovery rate in Fig. 2. Note that the phase transition behavior for each method is similar and the phase transition point of $n = 1000$ signal matches with the case of $n = 400$ signal.
7.1.3 Effect of Trace Penalty

As mentioned before, the trace minimization in (2) is not necessary provided the number of measurements \(m\) is on the order of \(n \log n\). In this case, restart-IncrePR terminates at stage I (step 4) in Algorithm 2, i.e., solution \(X_1\) is of rank one. To demonstrate the improvement of restart-IncrePR on phase transition when \(m\) is inbetween, we compare IncrePR for solving PhaseLift with \(\lambda = 0\) (actually only Stage I) and restart-IncrePR (Algorithm 2) applied to Gaussian phase retrieval. For fair comparison, we consider noiseless case, thus IncrePR and restart-IncrePR are followed by refinement (Stage III in Algorithm 2). From the solutions, to compute the recovery quality, we extract the rank one approximations \(x_1\) and \(x_3\) to the solutions \(Y_1\) and \(Y_3\) respectively. If \(Y_1\) is of rank one, we take \(Y_3 = Y_1\). We track \(x_1\) after Stage I and refinement, and call it the solution from ‘bare’ IncrePR. We calculate the relative error between \(x_1\) (\(x_3\)) and \(x^*\) to judge whether perfect recovery is obtained or not. Figure 3 depicts the effect of trace regularization. The phase transition curves and the average relative errors of IncrePR and Alg. 2 are illustrated in Fig. 3a and 3b respectively. The phase transition points of Alg. 2 and IncrePR are \(m = 1.9n\) (\(m = 2.7n\)) and \(m = 2.4n\) (\(m = 3.6n\)) for real (complex) Gaussian models respectively.

7.1.4 Noisy Measurements

We demonstrate the stability of restart-IncrePR (Alg. 2) for noisy Gaussian model. Since we choose the least-squares data fidelity, we consider additive Gaussian noise for the real case. We vary the noise level from 10dB to 50dB with an increment of 5dB. We only compare IncrePR with the RAF method for its superior performance among the compared nonconvex solvers. Figure 4 illustrates the stability of Alg. 2. For all tests, restart-IncrePR shows better performance, compared to RAF. The improved recoveries for \(m = 2n\) and \(m = 3n\) are in particular remarkable. Due to the stability of convex PhaseLift formulation (2), the capability of IncrePR in dealing with noisy data from minimal measurements is very promising. The poor performance for RAF is due to its poor initialization from noisy and limited measurement data.
7.2 Real Transmission Measurements

Now we test on a real example for which the measurement transmission matrix is provided by Phasepack [12]. The package is used to benchmark the performance of various phase retrieval algorithms. The transmission matrix dataset provides measurement matrices at three different image resolutions (16 × 16, 40 × 40 and 64 × 64). The rows of the transmission matrices are calculated using a measurement process (also a phase retrieval problem), and some are more accurate than the others. Each measurement matrix comes with a per-row residual to measure the accuracy of that row. For our test, we use the measurement matrix A_prVAMP.mat of image with resolution 16 × 16. The measurement matrix describes the multiple-scattering medium to model the forward image process. It comes from the solution to a phase retrieval problem, which is solved by the vector approximate message passing (VAMP) algorithm. Thus the measurement matrix contains an unknown level of noise and there is a quality measuring the precision of the dataset. Here we use the quantity to select the measurement data with top precision. See the details in [37]. We can cut off the measurement matrix to a smaller size by only loading the most accurate rows. We obtain the measurement matrix by setting the residual constant to 0.

Since we are considering a non-Gaussian model, the PhaseLift model is not tight for this example. We would not expect that the rank one approximation obtained from the PhaseLift model is the true solution except that it may be close to the global minimizer of the non-convex problem (6) with \( p = 1 \). We directly adopt Algorithm 2 and extract the rank one approximation of the solution, while replacing the refinement stage III by a nonconvex solver. Specifically, we refine the rank one approximation by nonconvex optimization of (6) with \( \lambda = 0 \) by RAF. Our numerical experiments show that the local minimizer found by IncrePR plus refinement from a random initialization (Alg.2 + r) is better than that found by directly solving the nonconvex problem (6) with \( p = 1 \) by RAF from random initialization (RAF+r) or from the reweighted maximal correlation initialization (RAF).

We compare Alg.2+r with RAF+r and RAF for the transmission dataset for the number of measurements \( m = 2n, 3n, 4n, 5n \). For IncrePR, the least-squares data fidelity (4) is used and the termination \( \epsilon \) is set to 100. We run 10 experiments for Alg.2 + r and RAF+r.
The relative errors of reconstructions are shown for real transmission measurements. It shows the average and standard deviation over 10 runs.

| m = 2n | m = 3n | m = 4n | m = 5n |
|--------|--------|--------|--------|
| Alg.2+r | 0.5407 ± 0.0153 | 0.5156 ± 0.0067 | 0.5113 ± 0.0033 | 0.5008 ± 0.0022 |
| RAF | 0.7532 | 0.6770 | 0.5904 | 0.5524 |
| RAF+r | 0.8250 ± 0.0017 | 0.8251 ± 0.0023 | 0.8154 ± 0.0154 | 0.8111 ± 0.0118 |

Fig. 5: The best reconstructions by Algorithm 2 + refinement (Alg.2+r), RAF+r and RAF over 10 runs from the number of measurements m = 2n, 3n, 4n, 5n. Top row are reconstructions by Alg.2+r, middle row are reconstructions by RAF, bottom row are reconstructions by RAF+r with the same random initialization. For RAF, the initialization is fixed by the reweighted maximal correlation method. The maximum iteration number of every inner nonconvex optimization for IncrePR is set to 3000, while the maximum number of iterations for RAF+r and RAF is three times as many. The average error and its standard deviation are listed in Table 2. Relative reconstruction error is calculated by the method provided in the source code accompanying [12]. From Table 2, superiority of Algorithm 2 and the fact that the more measurement data, the better the recovery are clear. The best reconstructions for each number of measurements over ten runs are plotted in Figure 5. The subcaption below each image is the relative reconstruction error. For m = 2n and m = 3n, the reconstruction of Alg.2+r is much better. This is because RAF depends critically on the initialization which is sensitive to both measurement model and the number of measurements.
7.3 Oversampling Fourier Phase Retrieval

In this section, we consider the more difficult oversampling Fourier phase retrieval problem of reconstructing an image from its diffraction data. Unlike Gaussian model, where ambiguity of the solution only comes from a constant phase shift, there are three different factors that can cause ambiguity [22]. All the state-of-the-art nonconvex solvers fail to find an even mediocre solution. Trace minimization convex problem (2) is no longer tight, and its direct solver also fails, as reported in [7]. Here we test the performance of restart-IncrePR for recovering the synthetic images of Cameraman and Barbara, all of size 128 × 128, from oversampling data. If an image is of size \( n_1 \times n_2 \), we pad the image with zeros to form an image of size \( 2n_1 \times 2n_2 \) and record the FFT intensities of the padded image. For the 2-d image, here the oversampling ratio is 2.\(^2\) The data fidelity function is the Poisson maximum likelihood estimation (5), due to the inapplicability of intensity least-squares for oversampling Fourier phase retrieval [30]. We investigate the performance of IncrePR and HIO [20] for noiseless and noisy cases. For the noiseless Fourier phase retrieval, the measurement data determine the trace of the unknown lifting matrix, i.e., the Frobenius norm of the image. Hence, we just solve this problem by Algorithm 2 (without the Stage II of solving (3) with \( \lambda = \lambda_0 \)) a few times. For noisy cases, Poisson noise, which is more appropriate for phase retrieval problem, is generated for the measurements. To obtain different noise levels, we first scale the noise-free intensity data \( b \) by \( \alpha \) and generate noisy data by Matlab function poissrnd, then we scale the output back by \( \alpha \), that is \( b_{\text{obs}} = \alpha \text{poissrnd}(b/\alpha) \), to output the desired noisy measurement with the given SNR. The \( \alpha \) can be approximated by 
\[
\alpha = \|b\|_2^2 / (10^{\text{SNR}/10} \sum b_i).
\]
The amplitude data feeding to HIO is \( \sqrt{b_{\text{obs}}} \).

After obtaining the rank one approximation factor of current solution, starting from it, we can run IncrePR repeatedly. The number of repeat \( K \) is set to 4. The termination \( \epsilon \) for each repeat is set to 0.1. We set the maximum iteration for each subproblem inIncrePR to 80. The initial elements of \( Y_0 \) are integers randomly chosen from 0 to 100. To avoid large-scale SVD computations in our IncrePR algorithm to extract the optimal rank one approximation at the end of each repeat, we just choose the column with the maximum norm, which turns out to work as well as SVD approach for rank one approximation. Since HIO with \( \beta = 0.9 \) is more stable than HIO with \( \beta = 1 \) by tracking the consecutive relative error in our tests, we set \( \beta = 0.9 \) in HIO for all runs. The maximal number of iterations is set to 3000. As suggested by the reviewer, we show the average image of best 10 recoveries form 100 runs for different noise levels. The noisy levels span from 30dB to 70dB with increment 10dB. Each run of Alg. 2 and HIO starts from the same 100 random initial guesses for each image. The relative error and SSIM (structural similarity) index are collected to quantitatively compare the results. As suggested by the reviewer, we show the average and standard deviation in Table 3 to test the stability and consistency of different runs. We emphasize that the two metrics seem not consistent with the eyeball norm (see Figure 6). For noiseless case, the recovery of HIO is better than that of IncrePR with larger relative error. For HIO, we also compare the effect of nonnegativity constraint on recovery. For Alg. 2, however, with nonnegativity constraints makes the relations between the convex problem (3) and non-convex problem (6) established in Theorem 1 and Theorem 2 invalid. Thus we only apply real-valuedness and support constraints. Our method recovers better images for noisy cases. When the oversampling ratio equals to 3, our method still recover better image for noisy case (not shown here). To the best of our knowledge, it is the first nonconvex optimization based method that works well for the Fourier phase retrieval problem. Although IncrePR is more expensive than HIO in

\(^2\) Here, the oversampling ratio \( d \) means to sample \( d \) times more points in each direction.
Table 3: Relative error and SSIM for IncrePR and HIO with/without nonnegativity constraints. It shows the average and standard deviation over best 10 recoveries from 100 runs.

| SNR | 30 | 40 | 50 | 60 | 70 | Noiseless |
|-----|----|----|----|----|----|-----------|
| rel.err |     |     |     |     |     |           |
| IncrePR | 0.3158 ± 0.0059 | 0.2552 ± 0.0122 | 0.1971 ± 0.0182 | 0.1586 ± 0.0154 | 0.1426 ± 0.0164 | 0.1414 ± 0.0218 |
| HIO | 0.6614 ± 0.0843 | 0.6250 ± 0.0906 | 0.5689 ± 0.0185 | 0.5169 ± 0.0009 | 0.5057 ± 0.0005 | 0.5012 ± 0.0003 |
| HIO+nonneg | 0.8403 ± 0.1243 | 0.7471 ± 0.1233 | 0.5822 ± 0.0305 | 0.5142 ± 0.0008 | 0.5045 ± 0.0006 | 0.5007 ± 0.0002 |
| SSIM |     |     |     |     |     |           |
| IncrePR | 0.0623 ± 0.0303 | 0.1054 ± 0.0600 | 0.1497 ± 0.1128 | 0.2105 ± 0.1537 | 0.2155 ± 0.1842 | 0.1623 ± 0.1896 |
| HIO | 0.0205 ± 0.0119 | 0.0370 ± 0.0190 | 0.0611 ± 0.0164 | 0.2012 ± 0.0062 | 0.3180 ± 0.0070 | 0.4843 ± 0.0209 |
| HIO+nonneg | 0.0051 ± 0.0095 | 0.0122 ± 0.0153 | 0.0508 ± 0.0176 | 0.2214 ± 0.0058 | 0.3341 ± 0.0101 | 0.5133 ± 0.0246 |

It shows the average and standard deviation over best 10 recoveries from 100 runs.
terms of computation cost, better reconstructions are obtained for noisy cases. The artifacts of the recoveries can be further removed by incorporating TV minimization into the recovery procedure [29,43,45].

8 Conclusion

In this work, we propose an incremental nonconvex approach, \textit{IncrePR}, to solve the phase retrieval problem based on the convex semidefinite relaxation (SDR) PhaseLift formulation. For Gaussian model, restart-\textit{IncrePR} obtains the sharpest phase transition compared with the state-of-the-art gradient flow based nonconvex solvers. Since \textit{IncrePR} solves the PhaseLift SDR problem, it avoids sensitive dependence of initialization and achieves global convergence, even when the number of measurements is close to the theoretical limit. It can escape stationary points and decrease the object function of PhaseLift by increasing the column number of the matrix decomposition factor successively. Compared to standard SDP solvers for PhaseLift, it reduces storage demand by matrix decomposition and improves computation efficiency by avoiding eigenvalue decomposition at each iteration. For phase
retrieval for non-Gaussian models, such as transmission matrix and oversampling Fourier measurements, although PhaseLift formulation is not tight and the optimal solution is not necessarily of rank one, one can still use restart-IncrePR to find a good approximation and then improve it further by a nonconvex solver, which often produces a better solution than that obtained by a direct nonconvex solver.

Acknowledgements Ji Li was supported by China Postdoctoral Science Foundation grant No. 2017M620589 and the Young Scientists Fund of the National Natural Science Foundation of China Grant No. 11801025. JFC was supported in part by Hong Kong Research Grant Council (HKRGC) Grant 16306317. The work of Hongkai Zhao is partially supported by the summer visiting program at CSRC.

References
1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press (2009)
2. Balan, R.: Reconstruction of signals from magnitudes of redundant representations: The complex case. Foundations of Computational Mathematics pp. 1–45. https://doi.org/10.1007/s10208-015-9261-0 (2015)
3. Balan, R., Casazza, P., Edidin, D.: On signal reconstruction without phase. Appl. Comput. Harmon. Anal. 20(3), 345–356 (2006). https://doi.org/10.1016/j.acha.2005.07.001
4. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization. J. Opt. Soc. Am. A 19(7), 1334–1345 (2002). https://doi.org/10.1364/josaa.19.001334
5. Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a matlab toolbox for optimization on manifolds. J. Mach. Learn. Res. 15(1), 1455–1459 (2013)
6. Burer, S., Monteiro, R.D.C.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Math. Program. 95(2), 329–357 (2003)
7. Candès, E.J., Eldar, Y.C., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM Rev. 57(2), 225–251 (2015). https://doi.org/10.1137/151005099
8. Candès, E.J., Li, X.: Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. Found. Comput.1 Math. 14(5), 1017–1026 (2013). https://doi.org/10.1007/s10208-013-9162-2
9. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval from coded diffraction patterns. Appl. Comput. Harmon. Anal. 39(2), 277–299 (2015). https://doi.org/10.1016/j.acha.2014.09.004
10. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via wirtinger flow: Theory and algorithms. IEEE Trans. Inf. Theory 61(4), 1985–2007 (2015). https://doi.org/10.1109/tit.2015.2399924
11. Candès, E.J., Strohmer, T., Voroninski, V.: PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. 66(8), 1241–1274 (2012). https://doi.org/10.1002/cpa.21432
12. Chandra, R., Zhong, Z., Hontz, J., McCulloch, V., Studer, C., Goldstein, T.: Phasepack: A phase retrieval library. In: 2017 51st Asilomar Conference on Signals, Systems, and Computers, pp. 1617–1621. IEEE (2017)
13. Chen, P., Fannjiang, A., Liu, G.R.: Phase retrieval with one or two diffraction patterns by alternating projections with the null initialization. J. Fourier Anal. Appl. 16(2), 1–40 (2015)
14. Chen, Y., Candès, E.J.: Solving random quadratic systems of equations is nearly as easy as solving linear systems. Commun. Pure Appl. Math. 70(5), 739–747 (2017)
15. Ciliberto, C., Stamos, D., Pontil, M.: Reexamining low rank matrix factorization for trace norm regularization (2017)
16. Demanet, L., Hand, P.: Stable optimizationless recovery from phaseless linear measurements. J. Fourier Anal. Appl. 20(1), 199–221 (2013). https://doi.org/10.1007/s00041-013-9305-2
17. Dhillon, O., Thrampoulidis, C., Lu, Y.M.: Phase retrieval via linear programming: Fundamental limits and algorithmic improvements. In: 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1071–1077. IEEE (2017)
18. Duchi, J.C., Ruan, F.: Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. Inf. Inf. A J. IMA 8(3), 471–529 (2018)
19. Fannjiang, A.: Absolute uniqueness of phase retrieval with random illumination. Inv. Prob. 28(7), 075008 (2012). https://doi.org/10.1088/0266-5611/28/7/075008
20. Fienup, J.R.: Phase retrieval algorithms: a comparison. Appl. Opt. 21(15), 2758–2769 (1982). https://doi.org/10.1364/ao.21.002758

21. Fogel, F., Waldspurger, I., Dâ ˘A ´ZAspremont, A.: Phase retrieval for imaging problems. Math. Programm. Comput. 8(3), 311–335 (2016)

22. Hayes, M.H.: The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform. IEEE Trans. Acoust. Speech Sig. Process. 30(2), 140–154 (1982). https://doi.org/10.1109/taspp.1982.1163863

23. Huang, W., Gallivan, K.A., Zhang, X.: Solving phaselift by low-rank riemannian optimization methods for complex semidefinite constraints. SIAM J. Sci. Comput. 39(5), B840–B859 (2017). https://doi.org/10.1137/16M1072838

24. Jaganathan, K., Eldar, Y.C., Hassibi, B.: Phase retrieval: An overview of recent developments. Mathematics (2015)

25. JournÃl'e, M., Bach, F., Absil, P.A., Sepulchre, R.: Low-rank optimization on the cone of positive semidefinite matrices. Siam J. Optim. 20(5), 2327–2351 (2010)

26. Knyazev, A. V.: Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. Siam J. Sci. Comput. 23(2), 517–541 (2006)

27. Kreutz-Delgado, K.: The complex gradient operator and the cr-calculus . http://arxiv.org/abs/0906.4835v1 (2009)

28. Kuznetsova, T.I.: On the phase retrieval problem in optics. Soviet Phys. Uspekhi 31(4), 364 (1988). https://doi.org/10.1070/pu1988v031n04abeh005755

29. Li, J., Zhao, H.: Solving phase retrieval via graph projection splitting. Inv. Prob. 36(5), 055003 (2020)

30. Li, J., Zhou, T.: Numerical optimization algorithm of wavefront phase retrieval from multiple measurements. Inv. Prob. Imag.11(4), (2016)

31. Li, J., Zhou, T.: On relaxed averaged alternating reflections (raar) algorithm for phase retrieval with structured illumination. Inv. Prob. 33(2), 025012 (2017)

32. Li, J., Zhou, T., Wang, C.: On global convergence of gradient descent algorithms for generalized phase retrieval problem. J. Comput. Appl. Math. 329, 202–222 (2017)

33. Luke, D.R.: Relaxed averaged alternating reflections for diffraction imaging. Inv. Prob. 21(1), 37–50 (2004). https://doi.org/10.1088/0266-5611/21/1/004

34. Luo, Z., Ma, W., So, A.M., Ye, Y., Zhang, S.: Semidefinite relaxation of quadratic optimization problems. IEEE Sig. Process. Mag. 27(3), 20–34 (2010)

35. Ma, J., Ji, X., Maleki, A.: Approximate message passing for amplitude based optimization. In: International Conference on Machine Learning (2018)

36. Metzler, C.A., Sharma, M.K., Nagesh, S., Baraniuk, R.G., Cossairt, O., Veeraraghavan, A.: Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In: 2017 IEEE International Conference on Computational Photography (ICCP), pp. 1–16. IEEE (2017)

37. Millane, R.P.: Phase retrieval in crystallography and optics. J. Opt. Soc. Am. A 7(3), 394–411 (1990). https://doi.org/10.1364/josaa.7.000394

38. Misell, D.L.: A method for the solution of the phase problem in electron microscopy. J. Phys. D Appl. Phys. 6(1), L6 (1973). https://doi.org/10.1088/0022-3727/6/1/102

39. Mishra, B., Meyer, G., Bach, F., Sepulchre, R.: Low-rank optimization with trace norm penalty. Siam J. Optim. 23(4), 2124–2149 (2013)

40. Netrapalli, P., Jain, P., Sanghavi, S.: Phase retrieval using alternating minimization. IEEE Trans. Sig. Process. 63(18), 4814–4826 (2015). https://doi.org/10.1109/tsp.2015.2448516

41. Nocedal, J., Wright, S.: Numerical Optimization. Springer (2006)

42. Pham, M., Yin, P., Rana, A., Osher, S., Miao, J.: Generalized proximal smoothing (gps) for phase retrieval. Opt. Exp. 27(3), 2792–2808 (2019)

43. Qian, J., Yang, C., Schirotzek, A., Maia, F., Marchesini, S.: Efficient algorithms for ptychographic phase retrieval. In: Contemporary Mathematics. American Mathematical Society (AMS). https://doi.org/10.1090/conm/615/12259 (2014)

44. Rodriguez, J.A., Xu, R., Chen, C.C., Zou, Y., Miao, J.: Oversampling smoothness: an effective algorithm for phase retrieval of noisy diffraction intensities. J. Appl. Crystallogr. 46(2), 312–318 (2013)

45. Shechtman, Y., Eldar, Y.C., Cohen, O., Chapman, H.N., Miao, J., Segev, M.: Phase retrieval with application to optical imaging: A contemporary overview. IEEE Signal Processing Magazine 32(3), 87–109. https://doi.org/10.1109/msp.2014.2352673 (2015)

46. Sun, J., Qu, Q., Wright, J.: A geometric analysis of phase retrieval. Found. Comput. Math. 18, 1131–1198 (2016)
48. Wang, G., Giannakis, G.B., Chen, J., Akğakaya, M.: Sparta: Sparse phase retrieval via truncated amplitude flow. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3974–3978 (2017)
49. Wang, G., Giannakis, G.B., Eldar, Y.C.: Solving systems of random quadratic equations via truncated amplitude flow. IEEE Trans. Inf. Theory 64(2), 773–794 (2018)
50. Wang, G., Giannakis, G.B., Saad, Y., Chen, J.: Solving almost all systems of random quadratic equations (2017)
51. Yurtsever, A., Udell, M., Tropp, J.A., Cevher, V.: Sketchy decisions: Convex low-rank matrix optimization with optimal storage. In: 20th International Conference on Artificial Intelligence and Statistics (AISTATS2017), CONF, pp. 1188–1196 (2017)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.