Solving phase retrieval via graph projection splitting

Ji Li¹,³ and Hongkai Zhao²

¹ Beijing Computational Science Research Center, Beijing, People’s Republic of China
² Department of Mathematics, University of California, Irvine, CA, United States of America

E-mail: matliji@nus.edu.sg, keelee@csrc.ac.cn and zhao@math.uci.edu

Received 19 October 2019, revised 2 February 2020
Accepted for publication 25 February 2020
Published 8 April 2020

Abstract
Phase retrieval with prior information can be cast as a nonsmooth and nonconvex optimization problem. To decouple the signal and measurement variables, we introduce an auxiliary variable and reformulate it as an optimization with an equality constraint. We then solve the reformulated problem by graph projection splitting (GPS), where the two proximity subproblems and the graph projection step can be solved efficiently. With slight modification, we also propose a robust graph projection splitting (RGPS) method to stabilize the iteration for noisy measurements. Contrary to intuition, RGP$^S$ outperforms GPS with fewer iterations to locate a satisfying solution even for noiseless case. Based on the connection between GPS and Douglas–Rachford iteration, under mild conditions on the sampling vectors, we analyze the fixed point sets and provide the local convergence of GPS and RGP$^S$ applied to noiseless phase retrieval without prior information. For noisy case, we provide the error bound of the reconstruction. Compared to other existing methods, thanks for the splitting approach, GPS and RGP$^S$ can efficiently solve phase retrieval with prior information regularization for general sampling vectors which are not necessarily isometric. For Gaussian phase retrieval, compared to existing gradient flow approaches, numerical results show that GPS and RGP$^S$ are much less sensitive to the initialization. Thus they markedly improve the phase transition in noiseless case and reconstruction in the presence of noise respectively. GPS shows sharpest phase transition among existing methods including RGP$^S$, while it needs more iterations than RGP$^S$ when the number of measurement is large enough. RGP$^S$ outperforms GPS in terms of stability for noisy measurements. When applying RGP$^S$ to more general non-Gaussian measurements with prior information, such as support, sparsity and TV minimization, RGP$^S$ either outperforms
state-of-the-art solvers or can be combined with state-of-the-art solvers to improve their reconstruction quality.

Keywords: phase retrieval, graph projection splitting, local convergence, Douglas–Rachford iteration

(Some figures may appear in colour only in the online journal)

1. Introduction

We consider the phase retrieval problem with prior information expressed as:

\[
\min_{x} g(x) \\
\text{s.t.} \quad |a_i^* x| = b_i + \epsilon_i, \quad i = 1, \ldots, m,
\]

where the objective function \(g(x)\) corresponds to the prior information, such as \(\ell_1\) norm for sparsity, total variation for piecewise constant, or the indicator function of solution set \(X \subseteq \mathbb{C}^n\). \(\{a_i\}_{i=1}^m \subseteq \mathbb{C}^n\) are the sampling vectors, the nonnegative \(b_i\)'s are amplitude measurements and \(\epsilon_i\)'s are the corruption noise. Without loss of generality, we assume that \(b_{\min} = \min\{b_1, \ldots, b_m\} > 0\), as we can put the equalities with \(b_i = 0\) as constraints included in \(g(x)\). Finding \(x\) amounts to solving a system of quadratic equations, which is generally an NP-hard problem. The main difficulty of solving (1) stems from the lack of the phase information and the nonconvexity of the amplitude measurement constraints. When the sampling vectors are drawn from Fourier transform basis, problem (1) is the so-called Fourier phase retrieval, which has a wide range of imaging applications in science and engineering, such as x-ray crystallography [1], electron microscopy [2], x-ray diffraction imaging [3], optics [4] and astronomy [5], just to name a few. When \(a_i\)'s are drawn from (complex) Gaussian distribution, problem (1) is called Gaussian phase retrieval, which is the model problem in recent research of phase retrieval, due to its nice statistical properties that lead to provable theoretical results.

Much efforts have been devoted to developing provable algorithms for Gaussian phase retrieval without prior information, i.e., \(g(x) = 0\). For Gaussian phase retrieval, the solution is unique up to a global phase offset in noiseless case \((\epsilon_i = 0)\) and it is also stable in noisy case when the number of measurements \(m = \mathcal{O}(n)\) [6, 7]. There are many provable algorithms to locate the solution to (1) from a good initial guess for Gaussian measurements. These algorithms include Wirtinger flow [8], truncated Wirtinger flow [7], amplitude truncated flow (ATF) [9], reweighted amplitude flow (RAF) [10] and alternating minimization [11]. Most of them are based on gradient flows for different loss functions starting from a specific initialization. Thus the choice of step size is important to ensure convergence and achieve fast convergence rate. When ratio \(m/n\) is large enough, a good initial guess can be generated by various initialization schemes, such as spectral method [8] and reweighted maximal correlation method [10]. However, finding a good initial guess stably is not a simple task in general, for examples, when the number of Gaussian measurements is not large enough or the measurements are not Gaussian. There is a line of works to explore the geometric landscape of nonconvex optimization. These works meant to point out that nonconvexity is not necessarily a challenge for intensity-based least-squares for phase retrieval and gradient flow based methods can actually avoid local minimum. However, this benign behavior depends on the number of measurements [12, 13] for Gaussian phase retrieval, or on the measurement setting, such as diversity between the two scanning surface data or the two probes [14].

The crucial dependence of a good initial guess can be avoided by the lifting technique for quadratic programming. Convex semidefinite relaxation (SDR) approaches, such as PhaseLift [15] and PhaseCut [16], have been proposed to solve (1). However, the extended dimension of
SDR is prohibitive for high dimensional phase retrieval applications. Another convexification algorithm PhaseLin is introduced with the help of anchor vector [17], which plays the similar role as initialization in gradient flow algorithms. To reduce the high computation cost of PhaseLift, matrix-factorization based approach, IncrePR [18], with improved phase transition is developed. The main issue of these solvers is their inapplicability to tackle practical phase retrieval problems with non-Gaussian measurements, where a good initial guess can not be easily obtained for gradient flow solvers and the tightness of SDR does not hold for convex solvers. Although these methods can be adapted to the inclusion of a regularization term $g(x)$ in the objective, they usually fail to find a satisfactory solution due to the stagnation of nonconvex optimization without a good initial guess.

For different types of sampling vectors, the uniqueness and stability of the solution to (1) may differ. For Fourier phase retrieval, the uniqueness is almost guaranteed up to three trivial ambiguities—translation, mirror shift, global phase offset—and their combinations. To mitigate these difficulties and improve the algorithmic efficiency, some additional constraints on the solution are imposed, such as support set, real-valuedness, nonnegativity, and sparsity. With the prior information, the popular solver HIO (hybrid input–output) and RAAR are widely used to solve Fourier phase retrieval [19, 20]. For the ‘far-field’ Fourier phase retrieval, the popular alternating projection (AP) method performs worse [5], though some success had been reported for other related problems, such as antenna pattern synthesis [21, 22]. This instability results from the stagnation of AP. However, they do not apply in the case of non-isometric measurements with more general prior constraints, e.g., total variation regularization. Recently, RAAR has been adapted to nonisometric measurements [23], but it does not support prior information. Thus these existing algorithms for non-Gaussian measurements have different kinds of restrictions. In this paper, we propose an unified algorithmic framework to solve phase retrieval that supports general measurements and prior information simultaneously.

For noiseless phase retrieval (1), we stack the sampling vectors into a matrix $A = [a_1, \ldots, a_m] \in \mathbb{C}^{n \times m}$, then the measurements can be written as $|A^* x | = b$. Throughout this paper, we assume $A^*$ is full column rank. With the indicator function $f(y) = \mathbb{1}_{\{|y| = b\}}(y)$, problem (1) can be reformulated as an unconstrained optimization to minimize $f(A^* x) + g(x)$.

To tackle the difficulty of the involvement of $A^*$ in $f$, we consider the following splitting form:

$$\begin{align*}
\min_{x} & \quad f(y) + g(x) \\
\text{s.t.} & \quad A^* x = y.
\end{align*}$$

Equation (2) can be solved by the following standard ADMM:

$$\begin{align*}
\lambda^{k+1} &= \lambda^k + A^* x^{k+1} - y^{k+1}
\end{align*}$$

where $k$ is the iteration number. The proximity operator is defined as

$$\text{prox}_f(z) = \mathbb{1}_{\{|y| = b\}}(y) \circ z$$

where we suppress the parameter $\rho$ in our notation. Operator $\text{prox}_f(z)$ is just the projection of $z$ onto the set $\{y \in \mathbb{C}^m | |y| = b\}$ with the expression

$$\text{prox}_f(z) = b \circ \frac{z}{|z|}$$
Note that the division is elementwise. When $|z_i| = 0$, we set $z_i/|z_i| = 0$.

Although solving the first subproblem is not straightforward in general, it can be solved easily in some special situations. For example, if $m = n$ and matrix $A^*$ satisfies $AA^* = A'A = I$, then the first subproblem is equivalent to

$$x^{k+\frac{1}{2}} = \arg\min_x g(x) + \frac{\mu}{2} \|x - A(y^k - \lambda^k)\|^2.$$ 

Furthermore, if $g(x)$ is the indicator function of a set, such as nonnegativity, real-valuedness, the ADMM method is the same as the case considered in [24].

If $g(x) = 0$ and $m \geq n$, the first subproblem is just a least-squares problem, we have $x^{k+\frac{1}{2}} = (AA^*)^{-1}A(y^k - \lambda^k)$. Let $y^k_{DR} = \lambda^k + y^k$ and substitute it to (3), we have

$$x^{k+\frac{1}{2}} = (AA^*)^{-1}A(2y^k - y^k_{DR})$$
$$y^{k+1} = \text{prox}_f\left(A'x^{k+\frac{1}{2}} + y^k_{DR} - y^k\right)$$
$$y^k_{DR} = A'x^{k+\frac{1}{2}} + y^k_{DR} - y^k.$$

So after one iteration, we have $y^{k+1} = \text{prox}_f(y_{DR}^{k+1})$. With the relation, we have the equivalent Douglas–Rachford (DR) iteration

$$y^k_{DR} = y^k_{DR} + A'(AA^*)^{-1}A \left(2b \circ \frac{y^k_{DR} - y^k_{DR}}{|y^k_{DR} - y^k_{DR}|} - b \circ \frac{y^k_{DR} - y^k_{DR}}{|y^k_{DR} - y^k_{DR}|}\right)$$  \hspace{1cm} (DR)

Then we can generate the solution sequence $x^k$ by the expression $(AA^*)^{-1}A \text{prox}_f(y^k_{DR})$. DR can also be reduced from solving a set feasible problem [23]. Although the above DR is simple, it can not deal with the general case when $g(x) \neq 0$. Later we will include this DR algorithm in comparison when $g(x) = 0$ in our numerical experiments.

To develop an efficient algorithm for the general case, we revert to graph projection splitting (GPS) to solve (2) instead, where each subproblem can be solved efficiently. To the best of our knowledge, it is the first time GPS is used to solve phase retrieval problem (1). For phase retrieval problem with noise, we propose a robust GPS (RGPS) method. We would like to point out several advantages of using graph projection splitting in $(x, y) \in \mathbb{C}^{n+m}$ graph space

- Prior information for both signal and measurement can be easily incorporated.
- No difficult parameters, e.g., time step for gradient based method, to tune.
- Graph projection finds a pair $(x, y)$ that satisfies the exact relation in GPS.
- Graph projection updates $x, y$ simultaneously without bias in an optimal way.
- $(I + AA^*)^{-1}$ is better conditioned (than $(AA^*)^{-1}$).
- Using the distance to the graph in $(x, y)$ space in RGPS is more intrinsic and unbiased to measure the violation of the constraint.

Details of both algorithms are presented in section 2. Actually RGPS can also be used for noiseless case and performs better than GPS near the solution. Using the equivalence relation between GPS and Douglas–Rachford, we show local convergence of GPS and RGPS by formulating them in single-variable-updating forms in section 3 and defer some technical details in section 4. Numerical experiments are conducted to demonstrate the performance of GPS and RGPS in section 5. Section 6 provides the conclusion.

**Notation.** We use bold fonts for vectors and matrices. We denote the inner product in $\mathbb{C}^n$ by $\langle u, v \rangle = u^*v$. $u$ is said orthogonal to $v$ if Re$(u, v) = 0$. The indicator function for a set $C$ is
denoted as $\mathbb{I}_C$, which is defined by $\mathbb{I}_C(x) = 0$ for $x \in C$ and $\mathbb{I}_C(x) = +\infty$ otherwise. Notation $u \odot v$ and $u / v$ is the elementwise multiplication and division between vectors.

Reproducible research. The accompany code for this paper can be found online at: https://github.com/Chilie/GPS.

2. Algorithms

2.1. Graph projection splitting

Instead of recognizing $y$ as the auxiliary variable, we view splitting form (2) as an optimization with a stacked variable $(x, y) \in \mathbb{C}^{n+m}$. Then we apply graph projection splitting (GPS) to solve (2), where the equality constraint indicates the stacked variable is in the graph set $C = \{(x, y) \in \mathbb{C}^{n+m} | Ax = y\}$. Given initial primal variables $x^0, y^0$ and dual variables $\lambda^0, \nu^0$, the GPS iteration is given by

\[
\begin{align*}
(x^{k+\frac{1}{2}}, y^{k+\frac{1}{2}}) &= \Pi_A^* (x^k - \lambda^k + y^k - \nu^k) \quad (5a) \\
(x^{k+1}, y^{k+1}) &= \left( \text{prox}_g \left( x^{k+\frac{1}{2}} + \lambda^k \right), \text{prox}_f \left( y^{k+\frac{1}{2}} + \nu^k \right) \right) \quad (5b) \\
(\lambda^{k+1}, \nu^{k+1}) &= \left( \lambda^k + x^{k+\frac{1}{2}} - x^{k+1}, \nu^k + y^{k+\frac{1}{2}} - y^{k+1} \right), \quad (5c)
\end{align*}
\]

where operator $\Pi_A^*$ denotes projection onto the graph set $C$ and implies the dependence on $A^*$. The graph projection step is the most dominant computation, as the two proximal steps (5b) are easily obtained for $f$ and $g$ in our problems. Typically, we initialize $\lambda^0, \nu^0$ with zero vectors, $x^0$ with a random vector and $y^0 = A^* x^0$. The most appealing advantage of GPS is the parameter-free iteration.

The graph projection splitting names after the graph projection step in the algorithm. Actually, GPS can be interpreted as a specific ADMM in the stack variable $z = (x, y) \in \mathbb{C}^{n+m}$. By setting $\phi(z) = f(y) + g(x)$, we are to minimize $\phi(z)$ with the constraint $z \in C = \{z = (x, y) | Ax = y\}$. The ADMM iteration for it is as follows

\[
\begin{align*}
z^{k+\frac{1}{2}} &= \Pi_A^* (z^k - \zeta^k) \\
\hat{z}^{k+1} &= \text{prox}_\phi \left( z^{k+\frac{1}{2}} + \zeta^k \right) \\
\zeta^{k+1} &= \zeta^k + z^{k+\frac{1}{2}} - z^{k+1}, \quad \text{(ADMM)}
\end{align*}
\]

where $\zeta = (\lambda, \nu) \in \mathbb{C}^{n+m}$ is the dual variable. With the splitting of our problem and the separation of function $\phi(z)$, substituting the stacked variable $z$ into ADMM iteration, we get GPS iteration (5). Note that for convex problem, the ADMM iteration ensures global convergence, and so does GPS. The ADMM/GPS can be directly applied to phase retrieval without modification, although the global convergence is not guaranteed. We will show its local convergence near the solution. Note that although GPS and the following RGPS are still based on non-convex optimization, their applications to Gaussian phase retrieval exhibit global convergence numerically.

2.2. Robust GPS

Note that the graph projection step (5a) is associated with the proximal operator for the indicator function $\mathbb{I}_C(\cdot)$. When we consider noisy measurements, we should change the
function $f(y)$. One way is to replace the indicator function by least-squares function $f(y) = \frac{1}{2} \|y - b\|^2$, which shows instability in our experiments. Instead, we introduce a relaxed projection by replacing $\mathbb{I}_C(\cdot)$ with $\text{dist}^2_C(s)$, which is square of the distance from $s$ to its projection $\Pi_C(s)$.

Hence, the graph projection step (5a) given by

$$(x^k + \frac{1}{2}, y^k + \frac{1}{2}) = \arg \min_{x, y} \mathbb{I}_C(x, y) + \frac{\rho}{2} \|x - c^k\|^2 + \frac{\rho}{2} \|y - d^k\|^2,$$

where $c^k = x^k - \lambda^k, d^k = y^k - \nu^k$, is replaced by

$$(x^k + \frac{1}{2}, y^k + \frac{1}{2}) = \arg \min_{x, y} \frac{1}{2} \text{dist}^2_C(x, y) + \frac{\rho}{2} \|x - c^k\|^2 + \frac{\rho}{2} \|y - d^k\|^2.$$

Lemma 2.1. Given $C \subset \mathbb{C}^n$ is a nonempty, closed convex set, then the distance function $\frac{1}{2} \text{dist}^2_C(s)$ is differentiable, and its gradient is

$$\nabla \left( \frac{1}{2} \text{dist}^2_C(s) \right) = (I - \Pi_C)s.$$

Proof. The proof is straightforward. Since

$$\frac{1}{2} \text{dist}^2_C(s) = \frac{1}{2} \|(I - \Pi_C)s\|^2,$$

the gradient is given by $(I - \Pi_C)s$. □

Moreover, since $C = \{(x, y) \in \mathbb{C}^{n+m}| A^*x = y\}$ is a linear space, we have the following properties.

Lemma 2.2. The inverse of the operator $(I - t\Pi_C)$ for $t \in \mathbb{R}, t \neq 1$ is $I + \frac{t}{1 - t}\Pi_C$.

Proof. It is easy to check that

$$(I - t\Pi_C) \left( I + \frac{t}{1 - t}\Pi_C \right) = I.$$

□

Theorem 2.3. The proximity operator for function $\frac{1}{2} \text{dist}^2_C(s)$ is given by

$$\text{prox}_{\frac{1}{2} \text{dist}^2_C}(s) = \left( \frac{\rho}{1 + \rho} I + \frac{1}{1 + \rho} \Pi_C \right)(s).$$

Proof. The proximity point $x$ satisfies

$$x = \text{prox}_{\frac{1}{2} \text{dist}^2_C}(s) = \arg \min_y \frac{1}{2} \text{dist}^2_C(y) + \frac{\rho}{2} \|y - s\|^2.$$

According to the first-order optimal condition, $x$ satisfies

$$(I - \Pi_C)x + \rho(x - s) = 0.$$
The optimal solution \( x \) is

\[
x = \frac{\rho}{1 + \rho} \left( I - \frac{1}{1 + \rho} \Pi_C \right)^{-1} s = \frac{\rho}{1 + \rho} \left( I + \frac{1}{\rho} \Pi_C \right) s = \left( \frac{\rho}{1 + \rho} I + \frac{1}{1 + \rho} \Pi_C \right) (s),
\]

where lemma 2.2 is used.

We obtain the robust graph projection,

\[
\left( x^{k+\frac{1}{2}}, y^{k+\frac{1}{2}} \right) = \left( tI + (1 - t)\Pi_{A^*} \right) (x^k - \lambda^k, y^k - \nu^k).
\]

(6)

When it replaces (5a) in GPS, we call the resulting algorithm robust GPS (RGPS). To ensure the local convergence, the allowable range of values of \( t \) falls into \((0, t_{\text{max}})\), where the upper \( t_{\text{max}} \) depends on the singular values of \( A^* \), see section 3.

Remark 1. We propose the robust GPS motivated by the infeasibility of graph set in noisy case. Actually, as we will see, besides noisy case, RGPS can also be used for noiseless case and outperforms GPS around the solution. This phenomenon can be somewhat explained by the local convergence behaviors of the two methods in section 3.

Multiplying with \( A^* \) on both sides of the first step of (4) and viewing \( A^* x^{k+\frac{1}{2}} \) as a whole variable, we recognize the resulting step as the projection \( 2x^k - y^{k}_{\text{DR}} \) onto the range of \( A^* \), by the same argument, robust Douglas–Rachford (RDR) can be also proposed. The iteration reads as follows

\[
y^{k+1}_{\text{DR}} = y^k_{\text{DR}} + (tI + (1 - t)A^*(AA^*)^{-1}) \left( 2b \circ \frac{y^k_{\text{DR}}}{|y^k_{\text{DR}}|} - y^k_{\text{DR}} \right) - b \circ \frac{y^k_{\text{DR}}}{|y^k_{\text{DR}}|}.
\]

(RDR)

2.3. Graph projection step

For \( c \in \mathbb{C}^n, d \in \mathbb{C}^m \), the projection of \((c, d)\) onto the set graph set \( C = \{(x, y)|A^*x = y, x \in \mathbb{C}^n, y \in \mathbb{C}^m\} \) can be computed explicitly as follows.

Theorem 2.4. The projection \( \Pi_{A^*}(c, d) \) is given by the solution to the linear system

\[
\begin{pmatrix} I & A \\ A^* & -I \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c + Ad \\ 0 \end{pmatrix}.
\]

(7)

Proof. The projection is to solve the following optimization problem with linear constraint

\[
\begin{align*}
\min_{x,y} & \quad \frac{1}{2} ||x - c||^2 + \frac{1}{2} ||y - d||^2 \\
\text{s.t.} & \quad A^*x = y.
\end{align*}
\]

According to the KKT condition, there exists \( \lambda \) such that

\[
\begin{align*}
x - c + A\lambda &= 0 \\
y - d - \lambda &= 0 \\
A^*x &= y.
\end{align*}
\]

After some substitutions, we reach the conclusion. \( \square \)
In most applications $m \geq n$, it is more efficient to compute the projection by the following

\[
x = (I + AA^*)^{-1}(c + Ad)
\]

\[
y = A^*x.
\]  \hspace{1cm} (8)

Using the Cholesky decomposition $LL^*$ of $I + AA^*$, where $L \in \mathbb{C}^{n \times n}$ is lower triangular matrix, we compute $x$ by the following forward and backward substitutions,

\[
L\tilde{x} = c + Ad \quad \text{(forward substitution)}
\]

\[
L^*x = \tilde{x} \quad \text{(backward substitution)}
\]

**Remark 2.** If we are to solve a phase retrieval problem with sparsity prior, $m$ may be less than $n$. In this case, we compute the projection as follows

\[
y = (I + A^*A)^{-1}A^*(c + Ad)
\]

\[
x = c + A(d - y).
\]  \hspace{1cm} (9)

**Remark.** If the sampling matrix $A^*$ satisfies $AA^* = ll^I$, such as coded diffraction pattern (CDP) [25], in which $l$ is the number of Fourier measurements with phase mask. In this case, we have

\[
(I + AA^*)^{-1} = \frac{1}{l + 1} I.
\]

Once the Cholesky decomposition is done once and the factor $L$ is stored, the only computations involved are matrix–vector multiplication in the computation. The computation of $I + AA^*$ needs $O(mn^2)$ flops and the computation of Cholesky decomposition needs $O(n^3)$ flops. At each iteration, the forward and backward substitution take $O(n^2)$ flops. Each graph projection costs $O(mn)$. Compared to other state-of-the-art nonconvex solvers, GPS and RGPS have the same order of computation cost in each step excluding the additional precomputation of Cholesky decomposition which is done once for all. In addition to the parameter-free advantage, another main advantage of GPS and RGPS is its flexibility to handle general measurements with prior information for $x$ due to the introduction of $y$ as an independent variable.

3. Local convergence analysis

In the section, we provide the local convergence of GPS and RGPS without regularization (for prior information), i.e. $g(x) = 0$. Although one may argue, in this case, DR involving only $y_{DR}^k$ should be used instead of GPS since GPS needs storage of both $\hat{x}^k$ and $\hat{\lambda}^k$. However, GPS seems to have the more tendency of escaping stagnation point than DR, which will be illustrated by numerical tests. Thus GPS (RGPS) achieves sharper phase transition than DR (RDR). Since this specific case is the basis of our splitting formulation, its convergence study will shed insight for our method in more general situations and we defer local convergence study with general regularization for future research. Hereafter, in the following convergence analysis, $g(x) = 0$ is always assumed.

For our analysis, we need assumptions of the given sampling vectors to ensure the uniqueness of solution to phase retrieval and related magnitude retrieval up to some trivial ambiguities.
Assumption 1. Throughout the paper, the following requirements of sampling vectors hold:

(a) Measurement matrix $A^*$ has full column rank.
(b) The solution to phase retrieval problem (1) is unique up to the trivial ambiguities, including translation shift, mirror flipping and global phase shift. The solution up to these ambiguities is denoted by $x^\dagger$ in our paper.
(c) The solution to magnitude retrieval problem

\[ \text{find } x \in \mathbb{C}^n \quad \text{s.t.} \quad \frac{a_i^*x}{|a_i^*x|} = \pm \frac{a_i^*x^\dagger}{|a_i^*x^\dagger|}, \quad i = 1, 2, \ldots, m \]

is unique up to a constant magnitude difference, where the $\pm$ sign is element-by-element, i.e., its solution is $cx^\dagger, c \in \mathbb{R} \setminus \{0\}$.

The assumptions almost hold for oversampling Fourier phase retrieval, see [26]. And we believe they are satisfied for Gaussian phase retrieval, while the uniqueness of magnitude retrieval needs proof.

3.1. Equivalence of GPS and Douglas–Rachford

The connection of (ADMM) and Douglas–Rachford (DR) is well-known. As DR involves only one sequence, it is easier to study than (ADMM). To see this connection, we take $z_k^{DR} = z_k^\dagger + \zeta_k$ and substitute it into (ADMM), we have

\[ z_k^{k+\frac{1}{2}} := \Pi_{A^*}(2z_k^\dagger - z_k^{DR}) \]
\[ z_k^{k+1} := \text{prox}_\phi \left( z_k^{k+\frac{1}{2}} + z_k^{DR} - z_k^k \right) \]
\[ z_k^{DR} := z_k^{k+\frac{1}{2}} + z_k^{DR} - z_k^k. \]

After one iteration, we have the relation $z_k^{k+1} = \text{prox}_\phi(z_k^{DR})$. So the iteration becomes:

\[ z_k^{k+\frac{1}{2}} := \Pi_{A^*}(2 \text{prox}_\phi(z_k^{DR}) - z_k^{DR}) \]
\[ z_k^{k+1} := z_k^{DR} + z_k^{k+\frac{1}{2}} - \text{prox}_\phi(z_k^{DR}). \]

It is easy to see the iteration falls into DR iteration. Thus sequence $\{z_k^k\}$ in (ADMM) can be generated by the following scheme

\[ z_k^{DR} = z_k^{DR} + \Pi_{A^*}\left(2 \text{prox}_\phi(z_k^{DR}) - z_k^{DR}\right) - \text{prox}_\phi(z_k^{DR}) \] (10a)
\[ z_k^{k+1} = \text{prox}_\phi(z_k^{DR}). \] (10b)

The obvious benefit is that we deal with a single-variable in DR iteration which is helpful for the local convergence study of GPS and RGPS for phase retrieval problem (1).

\[ \text{This is the general case for Fourier phase retrieval, the global phase shift is the only ambiguity for Gaussian phase retrieval.} \]
We let $z^k_{DR} = \begin{pmatrix} x^k_{DR} \\ y^k_{DR} \end{pmatrix}$ be the iterative sequence of DR, where $x^k_{DR} \in \mathbb{C}^n$ and $y^k_{DR} \in \mathbb{C}^m$. By equality (10b), we have $x^k = x^k_{DR}$. The DR iteration of GPS (5) is

$$
\begin{pmatrix} x^{k+1}_{DR} \\ y^{k+1}_{DR} \end{pmatrix} = \begin{pmatrix} x^k_{DR} \\ y^k_{DR} \end{pmatrix} + M \left( 2 \text{prox}_\phi \begin{pmatrix} x^k_{DR} \\ y^k_{DR} \end{pmatrix} - \begin{pmatrix} x^k_{DR} \\ y^k_{DR} \end{pmatrix} \right) - \text{prox}_\phi \begin{pmatrix} x^k_{DR} \\ y^k_{DR} \end{pmatrix},
$$

(11)

where

$$M = \begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} (I + AA^*)^{-1} & (I + AA^*)^{-1} \\ (I + AA^*)^{-1} & (I + AA^*)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & A \end{pmatrix}$$

is the corresponding matrix for the graph projection operator for the case $m \geq n$. Hereafter, without specific note, we assume $m \geq n$.

From the Cholesky decomposition $I + AA^* = LL^*$, we have $(I + AA^*)^{-1} = (L^{-1})^*L^{-1}$. The matrix $M$ can be expressed as

$$M = \begin{pmatrix} (L^{-1})^* \\ A^*L^{-1} \end{pmatrix} \begin{pmatrix} L^{-1} & L^{-1}A \end{pmatrix} : = C^*C.$$

It can be easily verified that matrix $C^*$ is isometric, i.e., $CC^* = I$.

### 3.2. Fixed point set

Now we derive the fixed point set of (11), and investigate the local convergence around the fixed point. This analysis can be adapted to investigate the local convergence of DR which we omit here.

Denote the solution to (1) by $x^\ast$, we define the set

$$\mathcal{F} = \left\{ \begin{pmatrix} \alpha x^\ast \\ \alpha(b + \theta) \circ \frac{y^\ast}{|y^\ast|} \end{pmatrix} | |\alpha| = 1, \quad \theta \in \mathbb{R}^m, \quad A \left( \theta \circ \frac{y^\ast}{|y^\ast|} \right) = \theta + b \geq 0 \right\},$$

where $y^\ast = A^*x^\ast$ with $|y^\ast| = b > 0$.

**Lemma 3.1.** Upon assumption 1, the fixed point set of iteration scheme (11) is equal to $\mathcal{F}$.

**Proof.** Obviously, every element in $\mathcal{F}$ is a fixed point.

On the other hand, the fixed point $z^\infty_{DR} = \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix}$ for iteration (11) satisfies

$$M \left( 2 \text{prox}_\phi \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix} - \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix} \right) - \text{prox}_\phi \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix} = \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix}.$$

By $M^2 = M$, we have

$$M \left( \text{prox}_\phi \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix} \right) = M \left( x^\infty_{DR} \right) = \text{prox}_\phi \begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix}.$$

By the second equality and the uniqueness of problem (1), we have $\begin{pmatrix} x^\infty_{DR} \\ y^\infty_{DR} \end{pmatrix} \in \mathcal{F}$. □
Given a point $z^k_{\text{DR}}$, we define its projection onto the fixed point set $F$

$$P_f z^k_{\text{DR}} = \arg \min_{z \in F} \| z - z^k_{\text{DR}} \|_2.$$ 

Denote $z^k := P_f z^k_{\text{DR}} = \begin{pmatrix} \alpha_k x^k \\ \alpha_k y^k \end{pmatrix} = \begin{pmatrix} \alpha_k \bar{x}^k \\ \alpha_k (b + \theta_k) \circ \frac{\bar{y}^k}{|\bar{y}^k|} \end{pmatrix},$ where $\theta_k \in C = \{ \theta \in \mathbb{R}^m \mid A (\theta \circ \frac{\bar{x}^k}{|\bar{x}^k|}) = 0 \text{ and } b + \theta \geq 0 \}$. When we study the local convergence around solution $z^* := \begin{pmatrix} x^* \\ y^* \end{pmatrix},$ for each iteration number $k$, we will consider an open ball $V_k \in \mathbb{C}^n$ of radius $b_{\text{min}}/2$ centered at $\alpha_k z^*$, where $\alpha_k$ is determined by $P_f z^k_{\text{DR}}$.

**Lemma 3.2.** If $b_{\text{min}} > 0$ and $z^k_{\text{DR}} \in V_k$, then we have that

$$\left\langle \frac{\alpha_k x^k}{|x^k|} \circ (y^k_{\text{DR}} - \alpha_k y^k), \theta \right\rangle = 0, \ \forall \theta \in C',$$

where $C' = \{ \theta \in \mathbb{R}^m \mid A (\theta \circ \frac{\bar{x}^k}{|\bar{x}^k|}) = 0 \}$. Therefore $\text{Re} \left( \frac{\alpha_k x^k}{|x^k|} \circ (y^k_{\text{DR}} - \alpha_k y^k) \right) \perp \theta$ and $\text{Im} \left( \frac{\alpha_k x^k}{|x^k|} \circ (y^k_{\text{DR}} - \alpha_k y^k) \right) \perp \theta$.

**Proof.** We consider set $F'$ with the definition

$$F' = \left\{ \begin{pmatrix} \alpha_k x^k \\ \alpha_k (b + \theta) \circ \frac{\bar{y}^k}{|\bar{y}^k|} \end{pmatrix} \mid \theta \in \mathbb{R}^m, \ A (\theta \circ \frac{\bar{y}^k}{|\bar{y}^k|}) = 0 \right\}.$$ 

From $z^k_{\text{DR}} \in V_k$, i.e., $\| z^k_{\text{DR}} - \alpha_k z^* \| \leq \frac{b_{\text{min}}}{2}$, by the projection property, it implies that $$\| z^k_{\text{DR}} - P_{F'} z^k_{\text{DR}} \| \leq \| z^k_{\text{DR}} - \alpha_k z^* \| \leq \frac{b_{\text{min}}}{2}.$$ 

Hence $\| P_{F'} z^k_{\text{DR}} - \alpha_k z^* \| \leq b_{\text{min}}$. Denote

$$P_{F'} z^k_{\text{DR}} = \begin{pmatrix} \alpha_k x^k \\ \alpha_k (b + \theta_k) \circ \frac{\bar{y}^k}{|\bar{y}^k|} \end{pmatrix},$$

we have $b + \theta_k \geq 0$, which implies $P_{F'} z^k_{\text{DR}} = P_{F'} z^k_{\text{DR}}$. By the definition of projection $P_{F'}$,

$$\theta_k = \arg \min_{\theta \in C'} \left\| y^k_{\text{DR}} - \alpha_k (b + \theta) \frac{\bar{y}^k}{|\bar{y}^k|} \right\|_2^2,$$

one can derive the optimality condition

$$\left\langle \frac{\alpha_k x^k}{|x^k|} \circ (y^k_{\text{DR}} - \alpha_k y^k), \theta \right\rangle = 0.$$
3.3. Local convergence of GPS

At the $k$th step, given $z_{DR}^k$, using the fixed point $P_{F^k_{DR}} = \begin{pmatrix} \alpha_1 x^k \\ \alpha_2 y^k \end{pmatrix}$, we have the recursive relation

$$
\begin{pmatrix}
   x_{DR}^{k+1} - \alpha_1 x^k \\
   y_{DR}^{k+1} - \alpha_2 y^k
\end{pmatrix} = \begin{pmatrix}
   x_{DR}^k - \alpha_1 x^k \\
   y_{DR}^k - \alpha_2 y^k
\end{pmatrix} + C^\ast C \begin{pmatrix}
   \Omega_1 \left( 2 \text{prox}_\phi \left( x_{DR}^k - \alpha_1 x^k \right) \right) - 2 \text{prox}_\phi \left( x_{DR}^k - \alpha_1 x^k \right) & \Omega_2 \left( \text{prox}_\phi \left( y_{DR}^k - \alpha_2 y^k \right) \right) - \text{prox}_\phi \left( y_{DR}^k - \alpha_2 y^k \right) \\

\end{pmatrix}
$$

(12)

Note that $\text{prox}_\phi(x_{DR}^k) = x_{DR}^k$. At the $k$th step, let $B = C\Omega^k$, where $\Omega^k = \begin{pmatrix} \Omega_t \\ \Omega_t \end{pmatrix}$, and $\Omega_t = \text{diag} \left( \frac{\alpha_1 x^k}{|x^k|}, \frac{\alpha_2 y^k}{|y^k|} \right)$. Hence the matrix $B$ depends on the iteration number $k$, but we do not explicitly show the dependence to simplify the notation. The proximity term in (12) is nonlinear which has the following directional derivative\(^5\) for $z \in \mathbb{C}^n$ and $h \in \mathbb{C}^n$,

$$
D \left( \frac{z}{|z|} \right)(h) = \frac{h}{|z|} - \frac{z \text{Re}(\overline{z} h)}{|z|^2} = \frac{iz \text{Im}(\overline{z} h)}{|z|^2}.
$$

(13)

Hence, the linear approximation to (12) around point $\begin{pmatrix} \alpha_1 x^k \\ \alpha_2 y^k \end{pmatrix}$ reads

$$
\begin{pmatrix}
   x_{DR}^{k+1} - \alpha_1 x^k \\
   y_{DR}^{k+1} - \alpha_2 y^k
\end{pmatrix} = \Omega(I - B^*B) \begin{pmatrix}
   \Omega_1^t(x_{DR}^k - \alpha_1 x^k) \\
   \Omega_2^t(y_{DR}^k - \alpha_2 y^k)
\end{pmatrix}
$$

$$
+ \Omega(2B^*B - I) \begin{pmatrix}
   \Omega_1^t(x_{DR}^k - \alpha_1 x^k) \\
   \Omega_2^t(y_{DR}^k - \alpha_2 y^k)
\end{pmatrix} + \text{h.o.t.}
$$

Denote $w^{k+1} = \begin{pmatrix} \Omega_1^t(x_{DR}^{k+1} - \alpha_1 x^k) \\ \Omega_2^t(y_{DR}^{k+1} - \alpha_2 y^k) \end{pmatrix}$ and $w^k = \begin{pmatrix} \Omega_1^t(x_{DR}^k - \alpha_1 x^k) \\ \Omega_2^t(y_{DR}^k - \alpha_2 y^k) \end{pmatrix}$, then the iteration scheme becomes

$$
w^{k+1} = (I - B^*B) \begin{pmatrix}
   \Omega_1^t(x_{DR}^k - \alpha_1 x^k) \\
   \Omega_2^t(y_{DR}^k - \alpha_2 y^k)
\end{pmatrix}
$$

$$
+ (2B^*B - I) \begin{pmatrix}
   \Omega_1^t(x_{DR}^k - \alpha_1 x^k) \\
   \Omega_2^t(y_{DR}^k - \alpha_2 y^k)
\end{pmatrix} + \text{h.o.t.}
$$

$$
:= J_k(w^k) + \text{h.o.t.}
$$

\(^5\)Here we exploit the $\mathbb{C} - \mathbb{R}$ calculus of real-valued function in complex-valued variables, please refer to [27].
Let \( J_0(v^k) = (I - B^*B) \left( \Omega^k_R(x^k_{DR} - \alpha k x^k) \right) + (2B^*B - I) \left( \Omega^k_R(y^k_{DR} - \alpha k y^k) \right) \), we infer from the continuity of the operator \( J_k \) around \( J_0 \) that for any \( \theta \in C \) such that \( \| \theta \| \leq \varepsilon_2 \),

\[
\| J_k - J_0 \| \leq \varepsilon_1.
\]

Given point \( z^k_{DR} \) and denote \( z^k = \begin{pmatrix} x^k \\ y^k \end{pmatrix} \), we assume \( \| z^k_{DR} - z^k \| < \varepsilon_2 < \frac{\alpha_0}{b_{\max}} \), then \( z^k_{DR} = P x z^k_{DR} = (\alpha_0 x^k, \alpha_0 y^k) \) satisfies \( \| \alpha_0 y^k - y^k \| < 2 \| z^k_{DR} - z^k \| < \varepsilon_2 < \frac{\alpha_0}{b_{\max}} \), and hence \( \| \alpha_0 y^k - y^k \| < \varepsilon_2 \), a.k.a. \( \| \theta \| \leq \varepsilon_2 \). The quantity \( \varepsilon_1, \varepsilon_2 \) and \( \varepsilon_3 \) will be determined accordingly. Starting from this point, it can be shown that \( \| \alpha_0 y^k - y^k \| \leq \varepsilon_2 \) for all \( k \geq 1 \). Therefore

\[
\| w^{k+1} \| \leq \| J_k(v^k) \| + o(\| v^k \|) \leq \| J_0(v^k) \| + (\varepsilon_1 + o(1)) \| v^k \|.
\]

Denote \( v^k = \text{Re}(v^k) + i \text{Im}(v^k) = \begin{pmatrix} e \\ f \end{pmatrix} + i \begin{pmatrix} g \\ h \end{pmatrix} \), where \( e, g \in \mathbb{R}^n \) and \( f, h \in \mathbb{R}^m \), we have

\[
J_0(v^k) = (I - B^*B) \text{Re}(v^k) + iB^*B \text{Im}(v^k) + (2B^*B - I) \left( \text{Re} \left( \Omega^k_R(x^k_{DR} - \alpha k x^k) \right) \right)
\]

\[
= B^*B \begin{pmatrix} e \\ 0 \end{pmatrix} + (I - B^*B) \begin{pmatrix} 0 \\ f \end{pmatrix} + dB^*B \begin{pmatrix} g \\ h \end{pmatrix}
\]

\[
= B^*B \begin{pmatrix} e + ig \\ ih \end{pmatrix} + (I - B^*B) \begin{pmatrix} 0 \\ f \end{pmatrix}.
\]

Since \( B^* \) is isometric, \( B^*B \) is a projection. It is easy to see that

\[
\| J_0(v^k) \|^2 = \| B^*B \left( \begin{pmatrix} e + ig \\ ih \end{pmatrix} \right) \|^2 + \| (I - B^*B) \begin{pmatrix} 0 \\ f \end{pmatrix} \|^2
\]

\[
\leq \| e \|^2 + \| g \|^2 + \| h \|^2 + \| f \|^2 = \| v^k \|^2.
\]

One can further bound \( \| J_0(v^k) \| \) from the following facts.

**Lemma 3.3.** \( e = \text{Re} \left( \Omega^k_R(x^k_{DR} - \alpha k x^k) \right) \); \( f = \text{Re} \left( \Omega^k_R(y^k_{DR} - \alpha k y^k) \right) \) and provided \( z^k_{DR} \in V_1 \), there exist two constants \( 0 < \delta_1, \delta_2 < 1 \), such that

\[
\| B^*B \begin{pmatrix} e \\ 0 \end{pmatrix} \| \leq \frac{1}{\sqrt{1 + (\text{Re}(A))^2}} \| e \| := \delta_1 \| e \|,
\]

\[
\| B^*B \begin{pmatrix} 0 \\ f \end{pmatrix} \| \geq \frac{\text{smin}(G(A^k))}{\text{smax}(A)^2} \| f \| := \delta_2 \| f \|,
\]

where \( \text{smax}(A) \) and \( \text{smin}(A) \) are the largest and smallest non-zero singular values of matrix \( A \) respectively and \( G(A) = \begin{pmatrix} \text{Re}(A) \\ \text{Im}(A) \end{pmatrix} \).
\textbf{Proof.} By direct calculation and the isometry of $C$,  
\[
\left\| B^*B \begin{pmatrix} e \\ 0 \end{pmatrix} \right\| = \left\| \Omega^* C \Omega \begin{pmatrix} e \\ 0 \end{pmatrix} \right\| = \left\| \Omega \begin{pmatrix} e \\ 0 \end{pmatrix} \right\| = \left\| L^{-1} \Omega e \right\| .
\]
where $L$ satisfies $LL^* = I + AA^*$. By the relation of the eigenvalues of $I + AA^*$ and the singular values of $A$, which has full rank, one has the first bound. For the second inequality, we write down its explicit expression  
\[
B^*B \begin{pmatrix} 0 \\ f \end{pmatrix} = \begin{pmatrix} (\Omega_1^*(I + AA^*)^{-1}A\Omega_1)^2 \\ (\Omega_1^*A^*(I + AA^*)^{-1}A\Omega_1)^2 \end{pmatrix}.
\]
In principle, when the vector $f$ is in the null space of $A\Omega_1$, the lower bound of the norm can vanish. By the positive definiteness of matrix $(I + AA^*)^{-1}$, so  
\[
\left\| \Omega_1^*(I + AA^*)^{-1}A\Omega_1 f \right\| \geq \frac{1}{1 + (s_{\max}(A))^2} ||A\Omega_1 f||.
\]
By lemma 3.2, we have that $f \perp \theta, \forall \theta \in C'$, so  
\[
||A\Omega_1 f|| = ||G(A\Omega_1) f|| \geq s_{\min}(G(A\Omega_1)) ||f||,
\]
where $s_{\min}(G(A\Omega_1))$ denotes the smallest nonzero singular value of $G(A\Omega_1)$. Note that $s_{\min}(G(A\Omega_1))$ is the same for different iteration number $k$. Therefore,  
\[
\left\| B^*B \begin{pmatrix} 0 \\ f \end{pmatrix} \right\| \geq \frac{s_{\min}(G(A\Omega_1))}{1 + (s_{\max}(A))^2} ||f|| := \delta_2 ||f||.
\]
It is also obvious that $0 < \delta_1, \delta_2 < 1$ if $A$ is column full rank. \hfill \Box

As a consequence, we have  
\[
\left\| (I - B^*B) \begin{pmatrix} 0 \\ f \end{pmatrix} \right\| = ||f||^2 - \left\| B^*B \begin{pmatrix} 0 \\ f \end{pmatrix} \right\|^2 \leq (1 - \delta_2^2) ||f||^2.
\]
To bound the first norm, we will prove the following fact in next section.

\textbf{Lemma 3.4.} \textit{With the specific expression of $v^k = \begin{pmatrix} e \\ f \end{pmatrix} + i \begin{pmatrix} g \\ h \end{pmatrix}$, we have}  
\[
\left\| B^*B \begin{pmatrix} e + ig \\ ih \end{pmatrix} \right\|^2 \leq (\sigma_1^2 + (1 - \sigma_2^2)\delta_1^2)||e||^2 + \sigma_2^2||g||^2 + ||h||^2.
\]
\textit{for some $0 < \sigma_2 < 1$.}

Combining the above two lemmas, we have  
\[
\left\| J_0(v^k) \right\|^2 \leq (\sigma_2^2 + (1 - \sigma_2^2)\delta_1^2)||e||^2 + \sigma_2^2||g||^2 + ||h||^2 + (1 - \delta_2^2) ||f||^2.
\]
Therefore, there exists $0 < \beta < 1$, such that  
\[
\left\| w^{k+1} \right\| \leq (\beta + \epsilon_1) \left\| v^k \right\|.
\]
And we can choose an appropriate $\epsilon_1$ such that $\beta + \epsilon_1 < 1$. By the definition of projection onto $\mathcal{F}$, we have
\[
\|v^{k+1}\| \leq \|w^{k+1}\| \leq (\beta + \epsilon_1) \|v^k\|.
\]

The remaining thing needs to be shown is that when the iteration gets close enough to the true solution $z^*$ (up to a global phase), then it will stay close so that $\|x_k - J_0\| \leq \epsilon_1$ is true for later iterations. Suppose $\|v^1\| \leq \|z^{1}_{\text{DR}} - z^*\| < \epsilon_3 < \frac{\beta}{2}$, then $\|\alpha_1 z^*_k - z^*\| \leq 2\epsilon_3 < \epsilon_2$, and for all $\alpha_k z^*_k$ for $k = 2, \ldots$, we show $\|\alpha_k z^*_k - z^*\| \leq \epsilon_2$. According to the projection property, it implies
\[
\|\alpha_{k+1} z^*_k - z^*\| = \|(\alpha_{k+1} z^*_k - z^{1}_{\text{DR}} + (z^{1}_{\text{DR}} - \alpha_k z^*_k))\|
\leq \|z^{1}_{\text{DR}} - z^*\| + \|z^{1}_{\text{DR}} - \alpha_k z^*_k\|
\leq 2 \|z^{1}_{\text{DR}} - z^*\| \leq 2(\beta + \epsilon_1) \|v^k\|.
\]

Therefore,
\[
\|\alpha_{k+1} z^*_k - z^*\| = \|\alpha_{k+1} z^*_k - \alpha_1 z^*_1 + \alpha_1 z^*_1 - z^*\|
\leq \|\alpha_{k+1} z^*_k - \alpha_1 z^*_1\| + \|\alpha_1 z^*_1 - z^*\|.
\]

Iterating the above inequality backward, we have
\[
\|\alpha_{k+1} z^*_k - z^*\| \leq \|\alpha_1 z^*_1 - z^*\| + \sum_{j=1}^{k} \|\alpha_{j+1} z^*_j - \alpha_1 z^*_1\|
\leq \|\alpha_1 z^*_1 - z^*\| + \|z^*_{\text{DR}} - z^*\| + 2(\beta + \epsilon_1) \sum_{j=1}^{k} \|v^j\|.
\]
\[
< 2\epsilon_3 + 2 \|v^1\| \sum_{j=1}^{k} (\beta + \epsilon_1)^j
< \frac{2\epsilon_3}{1 - (\beta + \epsilon_1)}.
\]

Hence, one can choose $\epsilon_3 = \frac{1 - (\beta + \epsilon_1)}{2} \epsilon_2$ small enough such that $\alpha_k y^k$ is uniformly close to $y^k$ such that $\|x_k - J_0\| \leq \epsilon_1$.

Note that when $\|z^{1}_{\text{DR}} - z^*\|$ is small enough, both $\|v^k\|$ and $\|\alpha_k z^*_k - z^*\|$ can be controlled. Therefore, the high order term along with the linear approximation of $J_2(v^k)$ at each step can also be uniformly bounded. From the above contraction property of the DR iteration (11), we have the following linear convergence bound for $\text{dist}(x^k, x^*):= \min_{i=1}^{m} \|x^k - \alpha_i x^*\|$.
\[
\text{dist}(x^k, x^*) \leq \|x^k - \alpha_0 x^*\| = \|x^k_{\text{DR}} - \alpha_0 x^*\| \leq \|v^k\| \leq (\beta + \epsilon_1)^{k-1} \|v^1\|.
\]

Thus the convergence of $x^k$ to $x^*$ (up to a global phase) can be ensured.

Another thing needs to verify is that $x^k_{\text{DR}} \in \mathcal{V}_k$, i.e., $\|x^k_{\text{DR}} - \alpha_0 z^*\| \leq \frac{b_{\text{min}}}{\beta}$ for $k \geq 1$, where $\alpha_0$ is determined by $P_J z^{1}_{\text{DR}}$. Note that $\|x^k_{\text{DR}} - \alpha_0 z^*\| \leq \|x^k_{\text{DR}} - \alpha_k z^*_k\| + \|\theta_i\| \leq \frac{b_{\text{min}}}{\beta} + \frac{b_{\text{min}}}{\beta} \leq \frac{2b_{\text{min}}}{\beta}$. Here is the main result for the noiseless case of phase retrieval problem.

**Theorem 3.5.** For problem (1) in noiseless case ($\epsilon_i^{\text{noise}} = 0, i = 1, \ldots, m$) with the sampling vectors satisfying the assumption 1, if $\|z^{1}_{\text{DR}} - z^*\| \leq \epsilon_3 < \frac{b_{\text{min}}}{\beta}$, where $\epsilon_3$ depending on $\epsilon_1$.
Proof. We first compute
\[ \text{dist}(x^k, x^*) \leq \gamma^{k-1} \| z^k_{\text{DR}} - z^* \| \leq \gamma^{k-1} \| z^1_{\text{DR}} - z^* \|, \]
where
\[ z^1_{\text{DR}} = \left( \frac{\lambda^k + x^k}{\nu^k + y^k} \right) \text{ and } z^* = P_{\mathcal{F}} z^1_{\text{DR}}. \]

**Remark 4.** In the basin of local convergence, dist \((x^k, x^*)\) is generally not monotonically decreasing, it may oscillate. However, from theorem 3.5, it goes to zero in a controlled way that it is bounded by a linear convergence.

### 3.4. Error bound in noisy case

In noisy case, the measurement \( \tilde{b} = b + e^{\text{noise}} \), then we have the following result.

**Theorem 3.6 (Noisy case).** Assume that the sampling vectors \( \{a_i\}_{i=1}^m \) satisfy assumption 1, and the measurements \( b \) are corrupted by noise \( e^{\text{noise}} \). If \( \| z^k_{\text{DR}} - z^* \| \) is small enough but greater than the noise level, i.e., \( \| e^{\text{noise}} \| \leq \| e^{\text{noise}} \| \) is small enough such that \( \| \| \) is still in \( \mathcal{V}_k \) and \( \| e^{\text{noise}} \| \) is small enough such that \( \| e^{\text{noise}} \| \leq \epsilon_2 \), we have
\begin{equation}
\text{dist}(x^k, x^*) \leq (\beta + \epsilon_1) \| z^k_{\text{DR}} - z^* \| + \frac{1}{1 - (\beta + \epsilon_1)} \| e^{\text{noise}} \|. \tag{14}
\end{equation}

**Proof.** We first compute
\[ w^{k+1} := \left( \begin{array}{c}
\Omega^*_y (z^1_{\text{DR}} - \alpha_k x^k) \\
\Omega^*_y (z^1_{\text{DR}} - \alpha_k y^k)
\end{array} \right) \]
\[ = J_k(u^k) + (2B'B - I) \left( \begin{array}{c}
0 \\
\Omega^*_y \left( e^{\text{noise}} \circ \frac{y^1_{\text{DR}}}{||y^1_{\text{DR}}||} \right)
\end{array} \right) + \text{h.o.t.}
\]

Using the same argument as for the noiseless case, we have
\begin{equation}
\| w^{k+1} \| \leq \| w^{k+1} \| \leq (\beta + \epsilon_1) \| w^k \| + \| e^{\text{noise}} \|. \tag{15}
\end{equation}

Iterating the above inequality backward, we get
\begin{equation}
\| w^1 \| \leq (\beta + \epsilon_1) \| w^1 \| + \sum_{j=0}^{k-2} (\beta + \epsilon_1)^j \| w^j \| + \frac{1}{1 - (\beta + \epsilon_1)} \| e^{\text{noise}} \|. \tag{16}
\end{equation}

By \( \text{dist}(x^k, x^*) \leq \| w^1 \| \), we obtain the estimation of reconstruction in noisy case. \( \square \)

Applying GPS for noisy phase retrieval, its performance heavily depends on the noisy level. The requirement on noisy level is very strict. Thus for noisy case, RGPS is more appropriate and the performance is much better.
3.5. Local convergence of RGPS

Considering the robust GPS (RGPS) for noiseless setting, the operator matrix $M$ in (11) is replaced by the following matrix

$$
\tilde{M} = I + (1 - t)M, \quad 0 < t < t_{\text{max}} < 1.
$$

When considering the local convergence of RGPS via equivalent Douglas–Rachford iteration, we note that the fixed point set of the iteration scheme is no longer the set $\mathcal{F}$ of GPS. Thus local convergence of RGPS is different from the above analysis.

Define set $\mathcal{F}_0 = \left\{ \frac{\alpha x^2}{|\alpha|^2} \left| |\alpha| = 1 \right. \right\}$, we can check that elements of $\mathcal{F}_0$ belong to the fixed point set of RGPS$^6$. In the following, we investigate the local convergence of RGPS around the point $z^* = \left( \frac{x^2}{y^2} \right)$. And we assume there is no other fixed points around the solution up to a global phase shift. At $k$th step, following the same argument as before, given $z^{k}_{\text{DR}}$, we denote $P_{\mathcal{F}_0}(z^{k}_{\text{DR}}) = \left( \frac{\alpha x^2}{|\alpha|^2} \right)$ and $\Omega = \left( \begin{array}{c} \Omega_x \\ \Omega_y \end{array} \right) = \text{diag} \left( \frac{\alpha x^2}{|\alpha|^2}, \frac{\alpha y^2}{|\alpha|^2} \right)$, we have the iteration scheme ($B = C\Omega$)

$$
\begin{pmatrix}
    x^{k+1}_{\text{DR}} - \alpha x^2 \\
    y^{k+1}_{\text{DR}} - \alpha y^2
\end{pmatrix} = \Omega (1 - t)(I - B^*B) \begin{pmatrix}
    \Omega_x(x^{k}_{\text{DR}} - \alpha x^2) \\
    \Omega_y(y^{k}_{\text{DR}} - \alpha y^2)
\end{pmatrix} + \Omega((2(1 - t))B^*B - (1 - 2t)I) \begin{pmatrix}
    \Omega_x(x^{k}_{\text{DR}} - \alpha x^2) \\
    \Omega_y(y^{k}_{\text{DR}} - \alpha y^2)
\end{pmatrix} + \text{h.o.t.}
$$

where

$$
\tilde{v}^k = \begin{pmatrix}
    \Omega_x(x^{k}_{\text{DR}} - \alpha x^2) \\
    \Omega_y(y^{k}_{\text{DR}} - \alpha y^2)
\end{pmatrix} = \frac{\alpha x^2}{|\alpha|^2} \left( x^{k}_{\text{DR}} - \alpha x^2 \right) - \frac{\alpha y^2}{|\alpha|^2} \left( y^{k}_{\text{DR}} - \alpha y^2 \right)
$$

$$
= \text{Re}(\tilde{v}^k) + \text{Im}(\tilde{v}^k)
$$

$$
= \begin{pmatrix}
    \bar{e} \\
    \bar{f}
\end{pmatrix} + i \begin{pmatrix}
    \bar{g} \\
    \bar{h}
\end{pmatrix}.
$$

We still compute

$$
J^*_0(\tilde{v}^k) = (1 - t)(I - B^*B) \begin{pmatrix}
    0 \\
    \bar{f}
\end{pmatrix} + \left( I + (1 - t)B^*B \right) \begin{pmatrix}
    \bar{e} + i\bar{g} \\
    \bar{h}
\end{pmatrix} = (1 - t)J^*_0(\tilde{v}^k) + t \begin{pmatrix}
    \bar{e} + i\bar{g} \\
    \bar{h}
\end{pmatrix}.
$$

$^6$ We can show that the elements of $\mathcal{F}$ with $\theta \neq 0$ is not a fixed point of RGPS. Besides $\mathcal{F}_0$, there may exist other nontrivial fixed points.
Therefore, we have
\[ \|J_0(\bar{u}^k)\| \leq (1 - t) \|J_0(\bar{u}^k)\| + t \left\| \left( \frac{\bar{e} + i\bar{g}}{i\bar{h}} \right) \right\| . \]

Using the same argument for lemma 3.4, there exists 0 < \beta < 1, such that
\[ \|J_0(\bar{u}^k)\| \leq (\sigma_2^2 + (1 - \sigma_2^2)\delta_1^2)\|\bar{e}\|^2 + \sigma_2^2(\|\bar{g}\|^2 + \|\bar{h}\|^2) \leq \beta^2(\|\bar{e}\|^2 + \|\bar{g}\|^2 + \|\bar{h}\|^2) \leq \beta^2(\|\bar{e}\|^2 + \|\bar{g}\|^2 + \|\bar{h}\|^2) . \]

However, due to the possibility that \( \bar{f} \) may be in the null space of \( \tilde{\mathbf{B}} \), one can only have unity before \( \| \bar{f} \|^2 \).

Define \( r(t) = (1 - t)\sqrt{\beta^2(\|\bar{e}\|^2 + \|\bar{g}\|^2 + \|\bar{h}\|^2) + \|\bar{f}\|^2} + t\sqrt{\|\bar{e}\|^2 + \|\bar{g}\|^2 + \|\bar{h}\|^2} \), then
\[ \|J_0(\bar{u}^k)\| \leq \max(r(0),r(t_{\text{max}})) . \]

Now we show that if \( t \) satisfies some condition, 0 < \( t < t_{\text{max}} = \frac{2(1-\beta)}{2-\beta} \), where \( \beta = \max\{\sigma_2^2 + (1 - \sigma_2^2)\delta_1^2, \sigma_2^2\} \), there exists a constant \( \gamma < 1 \) such that \( \|J_0(\bar{u}^k)\| \leq \gamma \|\bar{u}^k\| \), which implies local convergence for RGPS.

**Lemma 3.7.** When 0 < \( t < t_{\text{max}} = \frac{2\alpha}{\alpha + 1} \) and \( \alpha_e = 1 - \beta^2 < 1 \), then we have that
\[ (1 - t)\sqrt{\|\bar{e}\|^2 - \alpha_e\|\bar{e}\|^2} + t\sqrt{\|\bar{e}\|^2} \leq \sqrt{(1 - t)^2 + \frac{\gamma^2}{\alpha_e}\|\bar{u}^k\|^2} := \gamma \|\bar{u}^k\| . \]

**Proof.** In this proof, for fixed \( t \), we denote \( f(u,s) = (1 - t)\sqrt{u - \alpha_e s} + ts \) where 0 < \( s \leq u \). By studying the partial derivatives \( \partial_s f \) and \( \partial_t f \), we know that \( f(u,s) \) is monotonically increasing with respect to \( u \) and \( f(u,s) \) attains maximum at \( s = \frac{r_u}{(1 - t)^2 + \alpha_e r_u} \). At this point \( f(u,s) \leq \sqrt{(1 - t)^2 + \frac{\gamma^2}{\alpha_e} u} \). If 0 < \( t < t_{\text{max}} = \frac{2\alpha}{\alpha + 1} \), then this scaling factor is strictly less than unity.

By the same argument as before for GPS, we can obtain the following local convergence result for RGPS.

**Theorem 3.8.** For RGPS, if the parameter \( t \) satisfies 0 < \( t < t_{\text{max}} = \frac{2(1-\beta)}{2-\beta} < 1 \), where \( \tilde{\beta} \) depends on the singular values of \( \mathbf{A} \), and \( \text{dist}(\mathbf{z}_{\text{DG}},\mathbf{z}^2) = \|\bar{u}^1\| \) is small enough such that \( \mathcal{F}_0 \) is the only fixed point set around \( \mathbf{z}^2 \), there is a \( 0 < \gamma < 1 \) such that
\[ \text{dist}(\mathbf{x}^1,\mathbf{x}^2) \leq (\gamma + o(1))^{k-1} \|\bar{u}^1\| . \]

The estimation of reconstruction of RGPS in noisy case can also be derived. Note that the requirement of small noise level is minimized. In this view, there is no doubt that it outperforms GPS for noisy measurements.

**Remark 5.** We note that the dynamic range of \( t \) is generally narrow. That is to say the iteration of RGPS is not far away from GPS, where \( t = 0 \). As shown by our numerical experiments, although a slight perturbation of GPS, RGPS stabilizes the iterations, in particular around the solution.
We conclude that all of GPS, RGPS, DR and RDR exhibit local convergence when they are applied to phase retrieval with \( g(x) = 0 \). The local convergence of DR and RDR can be derived by mimicking the above analysis. In general, GPS (DR) seems to converge to the contraction basin near the solution easier and faster due to exact graph projection. However, near the solution, GPS (DR) tends to oscillate more than its robust version, RGPS (RDR). This is because the fixed point set for GPS (DR) \( F \) may not be simple, i.e., more than \( z^\perp \) (up to a global phase), especially when the number of measurement is large compared to the dimension of the signal. In this case, moreover, \( F \) connect all the way continuously to \( z^\perp \). Compared to DR, GPS tends to escape stagnation stage easier due to the graph projection in \((x, y)\) space.

4. Technical details for proof

We provide the main technical detail used in previous proofs in this section. The key is to show the eigen structure of the isometric matrix \( B \). The local convergence analysis for specific Douglas–Rachford algorithms for Fourier phase retrieval problem can be found in [23, 28]. However, their convergence results require \( m \geq 2n \) while our analysis is valid for arbitrary \( m \) and \( n \). Here we also prove convergence of the newly proposed robust version.

First, we study the singular values associated with the matrix \( B \). Note that matrix \( B^* \) is isometric, then \( \| B^* x \| = \| x \| \), i.e, \( BB^* = I \). Note that \( B \in \mathbb{C}^{n \times (n + m)} \) (here \( m \geq n \), since we have no additional information on \( x \)),\(^7\) we write down the real form of complex matrix \( B \):

\[
B := \begin{pmatrix}
\text{Re}(B) \\
\text{Im}(B)
\end{pmatrix} \in \mathbb{R}^{2n \times (n + m)}.
\]

Accordingly, we define the operator \( G \), which maps a complex vector to its real and imaginary parts, i.e.,

\[
G(z) := \begin{pmatrix}
\text{Re}(z) \\
\text{Im}(z)
\end{pmatrix}.
\]

**Lemma 4.1.** By the definition of \( G \), we have the following relation

\[
G(B^* u) = \begin{pmatrix}
B^T G(u) \\
B^T G(-iu)
\end{pmatrix}, \quad \text{where } u \in \mathbb{C}^n.
\]

**Proof.** By straightforward calculation, we have

\[
G(B^* u) = G \left( (\text{Re}(B^T) - i \text{Im}(B^T))(\text{Re}(u) + i \text{Im}(u)) \right)
\]

\[
= G \left( (\text{Re}(B^T)\text{Re}(u) + \text{Im}(B^T)\text{Im}(u)) + i (\text{Re}(B^T)\text{Im}(u) - \text{Im}(B^T)\text{Re}(u)) \right)
\]

\[
= \begin{pmatrix}
\text{Re}(B^T)\text{Re}(u) + \text{Im}(B^T)\text{Im}(u) \\
\text{Re}(B^T)\text{Im}(u) - \text{Im}(B^T)\text{Re}(u)
\end{pmatrix}.
\]

\[\square\]

We assume that \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{2n} \geq \sigma_{2n+1} = \cdots = \sigma_{n+m} = 0 \) (\( m \geq n \)) are the singular values of \( B \), and \( \{ \eta_j \in \mathbb{R}^{n+m} \}_{j=1}^{n+m} \) and \( \{ \xi_j \in \mathbb{R}^{2n} \}_{j=1}^{2n} \) are its associated right singular vectors and left singular vectors respectively. By the definition, for \( j = 1, \ldots, 2n \),

\[\text{(16)}\]

\(^7\) We can also show the local convergence for \( m \leq n \) case using slightly modified analysis.
we have
\[ B\eta_j = \sigma_j \xi_j, \]
\[ B^\dagger \xi_j = \sigma_j n_j. \]

Accordingly, by denoting \( \xi_j = \begin{pmatrix} \xi^R_j \\ \xi^I_j \end{pmatrix} \), we have
\[
\begin{pmatrix}
\text{Re}(B) \eta_j \\
\text{Im}(B) \eta_j
\end{pmatrix} = \begin{pmatrix} \sigma_j \xi^R_j \\ \sigma_j \xi^I_j \end{pmatrix},
\]
\[
\begin{pmatrix}
\text{Re}(B^\dagger) \\
\text{Im}(B^\dagger)
\end{pmatrix} \begin{pmatrix} \xi^R_j \\ \xi^I_j \end{pmatrix} = \sigma_j n_j.
\]

Therefore, we have
\[ B\eta_j = \sigma_j G^{-1}(\xi_j), \]
\[ \text{Re}(B^\dagger) \xi^R_j + \text{Im}(B^\dagger) \xi^I_j = \sigma_j n_j. \]

The last equality can be written as
\[ \text{Re}(B^* G^{-1}(\xi_j)) = \sigma_j n_j. \]

By the isometry of \( B^* \), we can directly obtain the singular values \( \sigma_1 = 1 \) and \( \sigma_2 = 0 \). Moreover, the following properties hold.

**Theorem 4.2.** ([28]). For \( i = 1, \ldots, n \), the singular values \( \{\sigma_i\}_{i=1}^{2n} \) and singular vectors \( \{\xi_i\}_{i=1}^{2n} \) have the following relations
\[ 1 = \sigma^2_1 + \sigma^2_{2n+1-i} \]
\[ \xi_{2n+1-i} = G(-iG^{-1}(\xi_i)) \]
\[ \xi_i = G(iG^{-1}(\xi_{2n+1-i})). \]

By the eigen structure \( \{\sigma_1, \ldots, \sigma_{2n}\} \) of the matrix \( B \), we have the following relations.

**Lemma 4.3.** ([28]). For each \( k = 1, \ldots, n \), the following matrix-vector multiplication can be expressed as
\[ B^* B \eta_k = \sigma_k \left( \sigma_k \eta_k + i \sigma_{2n+1-k} \eta_{2n+1-k} \right) \]
\[ B^* B \eta_{2n+1-k} = \sigma_{2n+1-k} \left( \sigma_{2n+1-k} \eta_{2n+1-k} - i \sigma_k \eta_k \right). \]

In the next, we give the leading singular vector of \( B \) explicitly.

**Lemma 4.4.** For the real matrix \( B \in \mathbb{R}^{2n \times (n+m)} \) generated by \( B = C\Omega_k \) with \( \Omega_k \) being
\[
\text{diag} \begin{pmatrix}
\alpha_k x^1 \\
x^2 \\
\alpha_k A x^3
\end{pmatrix},
\]
the left singular vector corresponding to the leading singular value \( \sigma_1 = 1 \) is
\[ \xi_1 = G(L^* \alpha_k x^3) \] and the left singular vector corresponding to the least singular value \( \sigma_{2n} = 0 \) is
\[ \xi_{2n} = G(-iL^* \alpha_k x^3). \]
First we have

$$B^*L^*\alpha_3x^3 = \begin{pmatrix}
\text{diag} \left( \frac{\alpha_3x^3}{|x^3|} \right) \\
\text{diag} \left( \frac{\alpha_3A^*x^3}{|A^*x^3|} \right)
\end{pmatrix} \left( L^{-1} \right)^* A \left( L^{-1} \right)^* L^*\alpha_3x^3 = \begin{pmatrix}
|x^3| \\
|A^*x^3|
\end{pmatrix},$$

which says

$$\text{Re}(B^*L^*\alpha_3x^3) = \begin{pmatrix}
|x^3| \\
|A^*x^3|
\end{pmatrix}.$$ 

Similarly, it is easy to show that $B \begin{pmatrix}
|x^3| \\
|A^*x^3|
\end{pmatrix} = L^*\alpha_3x^3$. Hence the leading singular value $\sigma_1 = 1$ and its associated left and right singular vectors are $\xi_1 = G(L^*\alpha_3x^3) \in \mathbb{R}^n$ and $\eta_1 = \begin{pmatrix}
|x^3| \\
|A^*x^3|
\end{pmatrix} \in \mathbb{R}^{n+m}$ respectively. Furthermore, by the fact

$$G(B^*u) = \begin{pmatrix}
B^TG(u) \\
B^T G(-iu)
\end{pmatrix} \text{ and } B^T G(-iL^*\alpha_3x^3) = 0,$$

we have $\sigma_{2n} = 0$ and $\xi_{2n} = G(-iL^*\alpha_3x^3)$.

**Lemma 4.5.** Both $\tilde{v}^k$ and $\tilde{v}^k$, defined in sections 3.3 and 3.5 respectively, provided both $\|\tilde{v}^k\|$ and $\|\tilde{v}^k\|$ are small enough, then they satisfy $\text{Im}(\tilde{v}^0) \perp \eta_1$, $\text{Im}(\tilde{v}^0) \perp \eta_1$.

**Proof.** According to the definition of $\tilde{v}^k$, we have

$$\langle \tilde{v}^k, i\eta_1 \rangle = \begin{pmatrix}
\frac{\alpha_3x^3}{|x^3|} (x_{DR}^k - \alpha_3x^3), i(x^3) \\
\frac{\alpha_3y^k}{|y^k|} (y_{DR}^k - \alpha_3y^k), i(y^k)
\end{pmatrix} = \langle x_{DR}^k - \alpha_3x^3, i\alpha_3y^k \rangle + \langle y_{DR}^k - \alpha_3y^k, i\alpha_3y^k \rangle.$$

Since $\alpha_3$ is determined by the projection, we have $\text{Re}(\langle \tilde{v}^k, i\eta_1 \rangle) = 0$, which says $\text{Im}(\tilde{v}^k) \perp \eta_1$. Similarly,

$$\langle v^k, i\eta_1 \rangle = \begin{pmatrix}
\frac{\alpha_3x^3}{|x^3|} (x_{DR}^k - \alpha_3x^3), i(x^3) \\
\frac{\alpha_3y^k}{|y^k|} (y_{DR}^k - \alpha_3y^k), i(y^k)
\end{pmatrix} = \langle x_{DR}^k - \alpha_3x^3, i\alpha_3y^k \rangle + \langle y_{DR}^k - \alpha_3y^k, i\alpha_3y^k \rangle = \langle x_{DR}^k - \alpha_3x^3, i\alpha_3x^3 \rangle + \langle y_{DR}^k - \alpha_3y^k, i\alpha_3y^k \rangle - \frac{\alpha_3y^k}{|y^k|} (y_{DR}^k - \alpha_3y^k), i\theta_k \rangle.$$

By lemma 3.2 and the projection property, we also have $\text{Re}(\langle v^k, i\eta_1 \rangle) = 0$, i.e., $\text{Im}(\tilde{v}^0) \perp \eta_1$. 

**Proof of lemma 3.4.** Since we have that

$$\begin{pmatrix}
g \\
h
\end{pmatrix} \perp \eta_1,$$
we know that \((\frac{ig}{ih})\) ∈ \(\text{span}\{i\eta_2, \ldots, i\eta_{n+m}\}\). For the vector \(\begin{pmatrix} e \\ 0 \end{pmatrix}\), we can decompose it as \(v_1 + v_2\) such that \(v_1 \in \text{span}\{\eta_1\}\) and \(v_2 \in \text{span}\{\eta_2, \ldots, \eta_{n+m}\}\). According to the operation \(B^*B\) on the singular vectors, see lemma 4.3, we know that \(B^*B \begin{pmatrix} e \\ 0 \end{pmatrix} = v_1 + B^*Bv_2\) where \(B^*Bv_2 \in \text{span}\{\eta_j, i\eta_j| j = 2, \ldots, 2n - 1\}\). Then

\[
\left\|B^*B \begin{pmatrix} e + ig \\ ih \end{pmatrix}\right\|^2 = \left\|v_1 + B^*B \left(v_2 + \begin{pmatrix} ig \\ ih \end{pmatrix}\right)\right\|^2 \\
= \left\|v_1\right\|^2 + \left\|B^*B \left(v_2 + \begin{pmatrix} ig \\ ih \end{pmatrix}\right)\right\|^2 \\
\leq \left\|v_1\right\|^2 + \sigma_2^2(\left\|v_2\right\|^2 + \left\|g\right\|^2 + \left\|h\right\|^2),
\]

where we use the fact \(v_2 + \begin{pmatrix} ig \\ ih \end{pmatrix} \in \text{span}\{\eta_j, i\eta_j| j = 2, \ldots, 2n - 1\}\). Since

\[
\left\|v_1\right\|^2 + \left\|B^*Bv_2\right\|^2 = \left\|v_1 + B^*Bv_2\right\|^2 \\
= \left\|B^*B \begin{pmatrix} e \\ 0 \end{pmatrix}\right\|^2 \leq \delta_1^2 \left\|e\right\|^2,
\]

we have

\[
\left\|v_1\right\|^2 + \sigma_2^2(\left\|v_2\right\|^2 + \left\|g\right\|^2 + \left\|h\right\|^2) = (1 - \sigma_2^2)\left\|v_1\right\|^2 + \sigma_2^2(\left\|v_2\right\|^2 + \left\|g\right\|^2 + \left\|h\right\|^2) \\
\leq (1 - \sigma_2^2)\delta_1^2 \left\|e\right\|^2 + \sigma_2^2(\left\|e\right\|^2 + \left\|g\right\|^2 + \left\|h\right\|^2).
\]

This completes the proof. □

The following lemma verifies that the second singular value \(\sigma_2\) is strictly less than unity.

**Lemma 4.6.** Let \(B = C\Omega\), where \(\Omega = \begin{pmatrix} \alpha_2x^2 \\ |x^2| \alpha_2A^2x^2 \\ |A^2x^2| \end{pmatrix}\) where \(C = (L^{-1} L^{-1} A) \in \mathbb{C}^{n \times (m+n)}\). Then \(\left\|\text{Im}(B^*x)\right\| = 1\) holds for a unit vector \(x\) if and only if

\[
\begin{pmatrix} (L^{-1})^y x \\ |(L^{-1})^y x| \\ A^*(L^{-1})^y x \\ |A^*(L^{-1})^y x| \end{pmatrix} = i\delta \circ \begin{pmatrix} x^2 \\ |x^2| \\ A^2x^2 \\ |A^2x^2| \end{pmatrix}
\]

where the components of \(\delta\) are either 1 or \(-1\).
Proof. We have
\[
\|
\text{Im}(B^*x)\| = \| \text{Im} \left( \frac{\alpha x^2}{|x|^2} \odot (L^{-1})^*x \right) \| = \| \text{Im} \left( \frac{\alpha A^2x^2}{|A^*x|^2} \odot A^*(L^{-1})^*x \right) \| = \| \text{Im} \left( \frac{\alpha x^2}{|x|^2} \odot (L^{-1})^*x \odot |(L^{-1})^*x| \right) \| \leq \| C^*x \| = \| x \|,
\]
as \( C^* \) is isometric. The equality holds if and only if
\[
\left( \frac{\alpha x^2}{|x|^2} \odot (L^{-1})^*x \odot |(L^{-1})^*x| \right) = i\delta,
\]
where components of \( i\delta \) are either 1 or \(-1\). So we have
\[
\left( \frac{(L^{-1})^*x}{|(L^{-1})^*x|} \right) = \delta \odot \left( \frac{\alpha x^2}{|x|^2} \right).
\]
\[\Box\]

From the uniqueness of the magnitude retrieval for sampling vectors \( \{a_i\} \) (assumption 1), we have \( x = \pm \frac{L^*\alpha x^2}{|Lx|^2} \). By the result \( \sigma_2 = \max \{ \| \text{Im}(B^*u) \| : u \in C^n, u \perp iG^{-1}(\xi_1), \| u \| = 1 \} \) \[28\], thus \( \sigma_2 < 1 \).

In a nutshell, we have shown the local convergence of GPS and its robust version RGPS for phase retrieval problem (1). Though we have not given the attraction radius around the solution, it seems that GPS/RGPS shows global convergence starting from a random initialization when the ratio \( m/n \) is large enough for Gaussian phase retrieval in all our numerical experiments. At the process of preparing this manuscript, a close work on the global convergence on alternating minimization (a.k.a. alternating projection) for Gaussian phase retrieval is uploaded to arXiv \[29\]. The requirement is \( m/\log^2m \geq Mn^{1/2}\log^{1/2}n \) as \( n, m \to \infty \). Its argument may be helpful for the proof of global convergence of GPS/RGPS. Although global convergence is possible in the limit regime, AP performs worse than GPS/RGPS in our Gaussian phase retrieval setting.

5. Numerical experiments

In this section, we use various tests to demonstrate the performance of GPS/RGPS applied to phase retrieval problem (1) with and without prior information. To measure the reconstruction
quality, we define the relative error between the reconstruction \( x \) and the optimal solution \( x^\#$ as follows

\[
\text{rel.err}(x, x^\#) = \frac{\text{dist}(x, x^\#)}{\|x\|^2} = \min_{\alpha} \frac{\|x - \alpha x^\#\|}{\|x^\#\|}.
\] (17)

It is easy to see that the optimal \( \alpha \) is given by \( \frac{x^\#}{\|x^\#\|} \).

We first explore the numerical phase transition of GPS/RGPS applied to Gaussian phase retrieval without prior information. Then we add sparsity prior based on \( \ell_0 \) and \( \ell_1 \) norms respectively. We also consider more challenging and practical non-Gaussian transmission examples and (oversampling) Fourier phase retrieval. Current nonconvex solvers degrade significantly for non-Gaussian measurements without the help of randomness. With the flexibility and easiness of imposing prior information, we include nonnegativity condition in GPS/RGPS for transmission datasets. We show that the reconstruction is better than existing solvers, which have no straightforward way to add nonnegativity information. For Fourier phase retrieval, we consider three additional prior information: nonnegativity, rectangular support and total variation (TV) regularization. Demonstrates experiment superior performance of GPS/RGPS.

5.1. Synthetic Gaussian phase retrieval

Recently Gaussian phase retrieval is the most popular model problem of phase retrieval in the literature. We first compare GPS and RGPS to existing solvers for Gaussian phase retrieval without prior information. For Gaussian phase retrieval, problem (1) is called real and complex cases if \( a_i \)'s and the unknown \( x \) belong to \( \mathbb{R}^n \) and \( \mathbb{C}^n \) respectively. For the two cases, the length of the signal \( n \) is set to 400. We generate real/complex Gaussian signals at random, and draw the sampling vectors \( a_i \sim \mathcal{N}(0, I) \) and \( a_i \sim \mathcal{CN}(0, I) = \mathcal{N}(0, I/2) + i\mathcal{N}(0, I/2) \) for real and complex cases respectively. We compare GPS/RGPS and Douglas–Rachford (DR) and robust Douglas–Rachford (RDR) with other nonconvex optimization algorithms, such as RAF [10], TAF [9], TWF [7] and WirtFlow [8], all of which are gradient flow based and need to tune an optimal step size. We use the default configuration of the parameters described in the corresponding references. Note that here we only consider the nonconvex approaches for comparison, as the comparison of nonconvex solvers to SDR PhaseLift approach can be found in [18]. All tests are conducted with Monte-Carlo simulations. We do not explore the optimal \( t \) for RDR and RGPS and set \( t = 0.1 \) throughout all of our experiments. We set the maximum number of iterations to 5000 and the relative error tolerance to \( 1e-3 \) for each method. When either one of the two criteria satisfies, the iteration is stopped.

Phase transition. For Gaussian phase retrieval, the benchmark test 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 can be located by the algorithm for problem (1). When \( m/n \) is below the phase transition, there exists at least an instance for which the algorithm failed to find the true solution. To numerically test the phase transition, for a fixed signal with length \( n \), we test an array of different measurement sizes \( m = \{1n, 1.1n, \ldots, 5n\} \). For each pair of \( n \) and \( m \), we solve 30 randomly generated problems and calculate the successful recovery rate. A recovery is successful if the relative error is below \( 1e-3 \).

Since all of the compared algorithms are nonconvex, their performance depends on the initialization crucially. First we start these methods from the point returned by the reweighted maximal correlation method [10]. The initial guess is close to the optimal solution if the number of measurement \( m = O(n) \). We compute the recovery rate and plot the curves of recovery rate in figure 1(a) and (b) for real and complex cases respectively. We also explore the effect of
the initialization on these nonconvex methods, i.e., we start all algorithms from a random point. Phase transition in this case is plotted in figure 1(a) and (b). It shows GPS/RGPS and DR/RDR are much less sensitive to initialization than other gradient flow based solvers. For example, performance of Wirtinger flow depends on the quality of initialization, as the number of measurement does not fall into the range of global convergence. Compared to DR and RDR, GPS and RGPS show sharper phase transition, in particular for small number of measurement. This demonstrates that GPS/RGPS are more likely to escape the stagnation point than DR/RDR. GPS shows the sharpest phase transition among all methods, successfully solving problem (1) with $m = 1.7n$ and $m = 2.7n$ measurements for real and complex cases respectively. This phase transition is also better than the incremental nonconvex approach for PhaseLift [18].

For phase transition, compared with GPS, RGPS shows inferior performance and requires slightly more measurements for successful recovery, as shown in figure 1. We list the average number of iteration of GPS/RGPS and DR/RDR in table 1. For the four methods, their residual curves fall into two stages, locating the contraction basin and convergence to the solution. There are two different features of the residual curves of GPS/DR and RGPS/RDR. GPS/DR locate the contraction basin faster than their counterparts while RGPS/RDR converge faster than GPS/DR once in the contraction basin. Due to these different behaviors, the iteration

---

**Figure 1.** Recovery rate for real and complex Gaussian phase retrieval. The top row shows results from the specific initialization [10]. The bottom row shows results from random initialization.
Table 1. Comparison of average number of iteration from 30 repeat experiments.

| m/n | Real | Comp |
|-----|------|------|
|     | GPS  | RGPS | DR   | RDR  | GPS  | RGPS | DR   | RDR  |
| 1.5 | 4852 | 5000 | 4976 | 5000 | 4704 | 5000 | 4837 | 5000 |
| 1.6 | 3784 | 5000 | 4626 | 5000 | 3986 | 5000 | 3971 | 5000 |
| 1.7 | 1363 | 3842 | 972  | 4211 | 2687 | 4341 | 2514 | 4699 |
| 1.8 | 825  | 1178 | 504  | 1421 | 1999 | 3899 | 2027 | 3157 |
| 1.9 | 1131 | 444  | 1398 | 407  | 1566 | 1847 | 1653 | 1853 |
| 2.0 | 956  | 111  | 573  | 144  | 1469 | 1584 | 1382 | 1543 |
| 2.1 | 889  | 96   | 698  | 109  | 1299 | 904  | 1207 | 911  |
| 2.2 | 659  | 73   | 845  | 73   | 1205 | 599  | 1200 | 691  |
| 2.3 | 1000 | 68   | 937  | 63   | 1248 | 411  | 1080 | 605  |
| 2.4 | 854  | 63   | 933  | 56   | 1011 | 370  | 1051 | 498  |
| 2.5 | 742  | 61   | 699  | 56   | 1115 | 339  | 1219 | 396  |
| 2.6 | 1182 | 61   | 1021 | 52   | 1460 | 316  | 1224 | 298  |
| 2.7 | 894  | 58   | 1081 | 49   | 1507 | 274  | 1228 | 278  |
| 2.8 | 1000 | 56   | 1100 | 50   | 1234 | 258  | 1286 | 257  |
| 2.9 | 458  | 56   | 839  | 49   | 1067 | 248  | 1198 | 249  |
| 3.0 | 706  | 54   | 439  | 45   | 1258 | 240  | 1117 | 238  |

number for robust versions behave differently as the number of measurement changes. When $m \geq 1.9n$ and $m \geq 3.1n$ for real and complex cases respectively, the average number of iteration of RGPS and RDR is smaller than that of their counterparts. Within 5000 iterations, all of them find the solution, while GPS and DR oscillate around the solution, which leads to more iterations in total. However, when $m$ is smaller than the above critical point, the robust versions take more iterations and performs inferior to their counterparts in computation cost and phase transition. Within 5000 iterations, GPS and DR locate the basin of solution and converge, while RGPS and RDR do not locate the basin of solution. Based on these behaviors, we can combine GPS and RGPS to benefit from both of their advantages.

**Noisy case.** When measurements are corrupted by noise, RGPS and RDR should outperform GPS and DR. We consider Gaussian noise $\epsilon_\text{noise}$. To specify the effect of noise level, we generate noisy real and complex measurements with different SNR levels at $\{10, 15, \ldots, 50\}$. We set the number of measurement $m$ to $2n$ and $3n$ for real and complex cases respectively and the maximum number of iterations to 200. All test algorithms start from random initializations, since all initialization algorithms degrades significantly in the noisy cases for small ratio $m/n$. The behavior of relative error vs iteration number of a typical run for noisy measurement is plotted in figure 2(a) and (b). The oscillatory behavior of GPS/DR is clearly shown while their robust versions are much more monotone. The relative recovery error (in dB) vs noise level (SNR) is plotted in figure 2(c) and (d). Other gradient flow based non-convex methods fail to locate a solution of relative error blow unity starting from a random initialization. Further experiments suggest that when the output from RGPS or RDR, are input to these gradient flow based non-convex methods, a better solution can be located. Thus, one can utilize a hybrid approach: first locate an approximate recovery using RGPS, then feed it to another state-of-the-art method to further refine the recovery.
5.2. Sparse phase retrieval

In this section, we consider the phase retrieval problem with additional prior information of sparsity and nonnegativity of the solution. Function $g$ can take $\ell_1$ and $\ell_0$ respectively to promote sparsity in convex and nonconvex forms. The length of the signal is still 400. We set the known support to $J = [0, \text{floor}(n/2)]$ and the sparsity $s$ to 10, 20, 30. The number of measurement $m$ is set to $n$.

With the sparsity-promoting term $g$, $x$ can be updated by the following simple operations for the $\ell_1$ and $\ell_0$ functions respectively:

$$
x^{k+1} = \begin{cases} 
\max \left( x^k + \frac{1}{2} \lambda^k - p, 0 \right) & g = \ell_1 \\
\text{hardthreshold} \left( x^k + \frac{1}{2} \lambda^k \right) & g = \ell_0
\end{cases}
$$

where $p$ is the parameter for soft thresholding and the hard thresholding operator is just keeping the leading $s$ largest entries. For $\ell_1$ case, the recovery depends on the parameter $p$, which we set to 10, 20, 30 for the sparsity level $s = 10, 20, 30$ respectively. We report the successful recovery
rate and the number of iterations in table 2. The performance based on $\ell_1$ is much better than that based on $\ell_0$.

5.3. Real transmission dataset

Here we test RGPS for one type of practical measurement, where the measurement matrix $A^*$ is a real transmission matrix provided by Phasepack [30]. It is used to benchmark the performance of various phase retrieval algorithms. For the transmission matrix dataset, the rows of the measurement matrix 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 $A_{\text{prVAMP}}.mat$ of an image with a resolution ($16 \times 16$). The measurement matrix can be cut off to a smaller size by only loading the more accurate rows. We set the cut-off residual bound to 0.04, which allows the ratio $m/n$ to reach 6. Both the maximum number of iteration of RGPS and RAF are set to 3000.

The signal is of length $n = 256$. We randomly collect the sampling matrix from $A^*$ with the number of measurement $m = 2n, 3n, 4n, 5n, 6n$. For each pair $(m, n)$, we repeatedly solve each problem 10 times from different initializations. When applying RGPS, we impose both real-valuedness and nonnegativity prior, while we only impose real-valuedness for RAF, as nonnegativity can not be easily imposed in RAF. The average reconstruction error\footnote{We calculate the error by the algorithm provided in the paper [30].} is listed in table 3. Although different initialization leads to different reconstruction, RGPS outperforms RAF in all cases, in particular when the ratio $m/n$ is below 6. The best reconstructions of RGPS and RAF for each pair $(m, n)$ are plotted in figure 3. The subcaption below each image denotes the relative error of reconstruction.

5.4. Incorporating TV regularization

Our framework can also include the TV regularization to improve the reconstruction by popular HIO for oversampling Fourier phase retrieval. We first solve the phase retrieval problem without TV regularization and obtain a solution as the initialization, and then refine it by adding a TV minimization term to improve the reconstruction quality. In this case, we apply RGPS to

| Table 2. Recovery from 10 repeated experiments with sparsity prior information. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | Recov. rate     | Iteration       |
|                | 10              | 20              | 30              | 10              | 20              | 30              |
| $\ell_1$       | 1               | 1               | 0.8             | 787             | 897             | 2223            |
| $\ell_0$       | 0.7             | 0.6             | 0.5             | 1994            | 2594            | 3169            |

| Table 3. Relative reconstruction error averaged from 10 repeated experiments from different random initializations. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Sampled. $m/n$  | $m = 2n$        | $m = 3n$        | $m = 4n$        | $m = 5n$        | $m = 6n$        |
| RGPS            | 0.6245          | 0.6670          | 0.6927          | 0.6666          | 0.6227          |
| RAF             | 0.8253          | 0.8224          | 0.8156          | 0.8195          | 0.6357          |
Inverse Problems 36 (2020) 055003 L Ji and H Zhao

Figure 3. Reconstruction from RGPS (1st row) and RAF (2nd row) using transmission datasets of different sizes (from left to right, \( m = 2n, 3n, 4n, 5n, 6n \)) with relative error in subcaptions.

solve the following problem

\[
\begin{align*}
\min_{x,y_1,y_2} & \quad f_1(y_1) + f_2(y_2) + g(x) \\
\text{s.t.} & \quad A^*x = y_1 \\
& \quad Dx = y_2
\end{align*}
\]

(18)

where matrix \( D \) corresponds to the total variation linear operator which comprises of horizontal and vertical differences of \( x \). The oversampling measurement means a known rectangular support \( S \) of the solution which provides another constraint. Two sets of synthetic data are generated by the cameraman image and caffeine molecule [16], which are both of size \( 128 \times 128 \). We first pad the image of size \( 128 \times 128 \) to size of \( 256 \times 256 \) with zeros and the region of size \( 128 \times 128 \) is used as the known support set \( S \). The prior information is \( x \in \mathbb{R}^{n_1 \times n_2} \cap S \) and TV of the image should be small. We express (18) in standard form, where \( f(y) = f_1(y_1) + f_2(y_2) = \|y_1\|_1 + \|y_2\|_1 \), \( g(x) = \|x\|_\text{TV} \) and the graph set is \( C = \{(x,y_1,y_2)|A^*x = y_1, Dx = y_2\} \).

For Fourier phase retrieval, \( A^* \) is isometric. By exploiting the structure of TV operator, the graph projection step \( \Pi_C(c,d_1,d_2) \) of GPS is computed by solving

\[
(2I + DD^*)x = c + Ad_1 + D^*d_2.
\]

The conjugate gradient method is used to solve the above linear system for its matrix-free property. The robust GPS can be implemented accordingly for noisy measurements. Without TV regularization, RGPS produces about the same reconstruction quality as HIO. However, HIO is more efficient. Again we use a hybrid approach: first run HIO (\( \beta = 1 \)) by setting the number of iteration to 1000 and then feed the output to RGPS and run 30 iterations to solve (18). To investigate the effect of noise, we consider four noise levels with SNR being, \( \infty \) (noiseless case), 30, 40 and 50. For each case, we run HIO + RGPS 10 times from different initializations. The average relative error is listed in table 4. And the best outputs of HIO and HIO + RGPS for
Table 4. Average relative error of 10 experiments starting from random initializations for Fourier phase retrieval with/without TV minimization at different noise level.

| SNR | Cameraman | Molecule |
|-----|-----------|----------|
|     | HIO       | HIO + RGPS |
|     |           |           |
| ∞   | 0.0728    | 0.0289    |
| 30  | 0.1721    | 0.0752    |
| 40  | 0.1012    | 0.0408    |
| 50  | 0.0798    | 0.0343    |

Figure 4. Reconstruction of Fourier phase retrieval by HIO (1st and 3rd row) and HIO + RGPS (2nd and 4th row). From left to right, the noise level SNRs are ∞ (noiseless), 30, 40, 50 respectively.

cameraman and molecule at different noise levels are depicted in figure 4, where the subcaption gives the relative error. It is obvious that after the refinement of TV-minimization by RGPS, the quality of reconstruction is better. Note that the relative error is calculated after possible shift and mirror-reflection. There may be some misalignment of the molecule, so the relative error
may increase as SNR increases. Note that the comparison between HIO and HIO + RGPS makes sense for all cases, as they share the same alignment.

6. Conclusion

We proposed a unified framework for phase retrieval with prior information via graph projection splitting (GPS) and robust GPS (RGPS). Current solvers only work for either isometric Fourier measurements with special prior information or general measurements without prior information, while our framework allows general measurement matrix and prior information simultaneously. GPS is motivated by the splitting formulation and variable-stacking. By introducing the splitting and graph projection, GPS can flexibly incorporate additional prior information about the solution and each resulting subproblem can be solved easily. RGPS and robust Douglas–Rachford (RDR) for phase retrieval without prior information for noisy measurements are also proposed. Advantages of GPS and RGPS over existing gradient flow-based methods include graph projection step and no line search. We show local convergence of GPS and RGPS for noiseless case without prior. For noisy case, we characterize the reconstruction error around the solution.

For Gaussian phase retrieval without prior information, compared to other existing methods, GPS shows the sharpest phase transition and RGPS shows more stable reconstruction in various numerical experiments. RGPS outperforms GPS when the number of measurement is large enough. The performance of GPS and RGPS seem less dependent on the initialization than other gradient flow-based nonconvex solvers. RGPS also outperforms RAF for transmission measurement data especially when the number of measurements is small. It can also refine the reconstruction of HIO when TV regularization is added. The inclusion of TV regularization into oversampling Fourier phase retrieval is new and improves the reconstruction quality.

Acknowledgments

J L 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. Hongkai Zhao would like to thank the summer visitor program at CSRC.

ORCID iDs

Ji Li https://orcid.org/0000-0001-5244-0778

References

[1] Millane R P 1990 Phase retrieval in crystallography and optics J. Opt. Soc. Am. A 7 394–411
[2] Misell D L 1973 A method for the solution of the phase problem in electron microscopy J. Phys. D: Appl. Phys. 6 L6
[3] Shechtman Y, Eldar Y C, Cohen O, Chapman H N, Miao J and Segev M 2015 Phase retrieval with application to optical imaging: a contemporary overview IEEE Signal Process. Mag. 32 87–109
[4] Kuznetsova T I 1988 On the phase retrieval problem in optics Sov. Phys. Usp. 31 364
[5] Fienup J R 1982 Phase retrieval algorithms: a comparison Appl. Opt. 21 2758–69
[6] Candès E J, Strohmer T and Voroninski V 2012 PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming Commun. Pure Appl. Math. 66 1241–74
[7] Chen Y and Candès E J 2017 Solving random quadratic systems of equations is nearly as easy as solving linear systems Commun. Pure Appl. Math. 70 822–83
[8] Candès E J, Li X and Soltanolkotabi M 2015 Phase retrieval via wirtinger flow: theory and algorithms IEEE Trans. Inf. Theory 61 1985–2007
[9] Wang G, Giannakis G B and Eldar Y C 2017 Solving systems of random quadratic equations via truncated amplitude flow IEEE Trans. Inf. Theory 64 773–94
[10] Wang G, Giannakis G B, Saad Y and Chen J 2017 Solving Almost All Systems of Random Quadratic Equations (arXiv:1705.10407)
[11] Netrapalli P, Jain P and Sanghavi S 2015 Phase retrieval using alternating minimization IEEE Trans. Signal Process. 63 4814–26
[12] Li J, Zhou T and Wang C 2018 On global convergence of gradient descent algorithms for generalized phase retrieval problem J. Comput. Appl. Math. 329 202–22
[13] Sun J, Qu Q and Wright J 2018 A geometric analysis of phase retrieval Found. Comput. Math. 18 1131–98
[14] Soldovieri F, Angelo L, D’Elia G and Pierri R 2005 Global convergence of phase retrieval by quadratic approach IEEE Trans. Antennas Propag. 53 3135–41
[15] Candès E J, Eldar Y C, Strohmer T and Voroninski V 2015 Phase retrieval via matrix completion SIAM Rev. 57 225–51
[16] Waldspurger I, d’Aspremont A and Mallat S 2015 Phase recovery, maxcut and complex semidefinite programming Math. Program. 149 47–81
[17] Dhifallah O, Thampoulidis C and Lu Y M 2017 Phase retrieval via linear programming: fundamental limits and algorithmic improvements 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton) (IEEE) pp 1071–7
[18] Li J, Cai J-F and Zhao H 2018 Scalable Incremental Nonconvex Optimization Approach for Phase Retrieval from Minimal Measurements Scalable Incremental Nonconvex Optimization Approach for Phase Retrieval (arXiv:1807.05499)
[19] Russell Luke D 2004 Relaxed averaged alternating reflections for diffraction imaging Inverse Problems 21 37–50
[20] Russell Luke D, Bauschke H H and Combettes P L 2003 Hybrid projection–reflection method for phase retrieval J. Opt. Soc. Am. A 20 1025–34
[21] Bucci O M, D’Elia G, Leone G and Pierri R 1990 Far-field pattern determination from the near-field amplitude on two surfaces IEEE Trans. Antennas Propag. 38 1772–9
[22] Bucci O M, D’Elia G, Mazzarella G and Panariello G 1994 Antenna pattern synthesis: a new general approach Proc. IEEE 82 358–71
[23] Li J and Zhou T 2017 On relaxed averaged alternating reflections (raar) algorithm for phase retrieval with structured illumination Inverse Problems 33 025012
[24] Wen Z, Yang C, Liu X and Marchesini S 2012 Alternating direction methods for classical and ptychographic phase retrieval Inverse Problems 28 115010
[25] Candès E J, Li X and Soltanolkotabi M 2015 Phase retrieval from coded diffraction patterns Appl. Comput. Harmon. Anal. 39 277–99
[26] Hayes M H 1982 The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform IEEE Trans. Acoust., Speech, Signal Process. 30 140–54
[27] Kreutz-Delgado K 2009 The Complex Gradient Operator and the Cr-Calculus (arXiv preprint arXiv:0906.4835)
[28] Chen P and Albert F 2018 Fourier phase retrieval with a single mask by douglas–rachford algorithms Appl. Comput. Harmon. Anal. 44 665–99
[29] Zhang T 2020 Phase Retrieval by Alternating Minimization with Random Initialization IEEE Transactions on Information Theory
[30] Chandra R, Zhong Z, Hontz J, McCulloch V, Studer C and Goldstein T 2017 Phasepack: a phase retrieval library 2017 51st Asilomar Conf. on Signals, Systems, and Computers (IEEE) pp 1617–21