

Phase Retrieval by Linear Algebra

Pengwen Chen ^{*} Albert Fannjiang [†] Gi-Ren Liu [‡]

December 12, 2016

arXiv:1607.07484

Abstract

The null vector method, based on a simple linear algebraic concept, is proposed as an initialization method for non-convex approaches to the phase retrieval problem. In the case of complex Gaussian random measurement matrices, a non-asymptotic error bound is derived, yielding an asymptotically accurate approximation, in the limit of large oversampling ratio, better than that of the spectral vector method. Numerical experiments show that the null vector method also has a superior performance for small to medium oversampling ratios.

1 Introduction

We consider the following phase retrieval problem: Let $A = [a_{ij}] \in \mathbb{C}^{n \times N}$ where a_{ij} are independently and identically distributed (i.i.d.), circularly symmetric complex standard Gaussian random variables (the $n \times N$ standard complex Gaussian ensemble). Let $x_0 \in \mathbb{C}^n$ and $y = A^*x_0$. Suppose we are given A and $b := |y|$ where $|y|$ denote the vector such that $|y|(j) = |y(j)|, \forall j$. The aim of phase retrieval is to find x_0 .

Clearly this is a nonlinear inversion problem. Simple dimension count shows that, for the solution to be unique in general, the number of (nonnegative) data N needs to be at least twice the number n of unknown (complex) components. There are many approaches to phase retrieval, the most efficient and effective, especially when the problem size is large, being fixed point algorithms (see [4,6,7] and references therein) and gradient-descent methods [2,3]. Phase retrieval has a wide range of applications, see [9] for a recent survey.

A key to the success of any non-convex methods is an effective initialization. The following observation motivates our approach based on a simple linear algebraic concept: Let I be a

^{*}Department of Applied Mathematics, National Chung Hsing University, Taiwan.

[†]Department of Mathematics, University of California, Davis, CA 95616, USA.

[‡]Department of Mathematics, National Cheng Kung University, Taiwan

subset of $\{1, \dots, N\}$ and I_c its complement such that $b(i) \leq b(j)$ for all $i \in I, j \in I_c$. In other words, $\{b(i) : i \in I\}$ are the “weak” signals and $\{b(j) : j \in I_c\}$ the “strong” signals. Let $|I|$ be the cardinality of the set I . We always assume $|I| \geq n$. Since $b(i) = |a_i^* x_0|, i \in I$, are small, $\{a_i\}_{i \in I}$ is a set of sensing vectors nearly orthogonal to x_0 . Denote the sub-column matrices consisting of $\{a_i\}_{i \in I}$ and $\{a_j\}_{j \in I_c}$ by A_I and A_{I_c} , respectively. For $|I| \geq n$, define the null vector as the singular vector for the least singular value of A_I :

$$x_{\text{null}} := \arg \min \{ \|A_I^* x\|^2 : x \in \mathbb{C}^n, \|x\| = \|x_0\| \} \quad (1)$$

which can be computed by various matrix methods.

The main goal of the paper is to prove a non-asymptotic error bound for x_{null} as an estimate of x_0 (Theorem 2.1 and Corollary 2.2) and given ample numerical evidence for the performance of the null vector method (Section 5).

2 Main results

Note that both x_{null} and the phase retrieval solution is at best uniquely defined up to a global phase factor. So a standard error metric must be phase-adjusted as in

$$\min_{\theta \in \mathbb{R}} \|e^{i\theta} x_{\text{null}} - x_0\| = \sqrt{2(\|x_0\|^2 - |x_0^* x_{\text{null}}|)}. \quad (2)$$

Alternatively, we can use the following error metric

$$\|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\| = \sqrt{2(\|x_0\|^4 - |x_0^* x_{\text{null}}|^2)} \quad (3)$$

where the left hand side is measured in the spectral norm.

The following non-asymptotic estimate is our main theoretical result.

Theorem 2.1. *Let A be drawn from the $n \times N$ standard complex Gaussian ensemble. Let $\sigma, \nu, \epsilon, \delta, t$ be any constants constrained as follows*

$$\sigma := \frac{|I|}{N} < 1, \quad \nu = \frac{n}{|I|} < 1, \quad \epsilon \in (0, 1), \quad \delta > 0, \quad t \in (0, \nu^{-1/2} - 1). \quad (4)$$

Then for any $x_0 \in \mathbb{C}^n$ and x_{null} given by (1) the following error bound

$$\|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\|^2 \leq \left(\left(\frac{2+t}{1-\epsilon} \right) \sigma + \epsilon (-2 \ln(1-\sigma) + \delta) \right) \frac{\|x_0\|^4}{(1 - (1+t)\sqrt{\nu})^2} \quad (5)$$

holds with probability at least

$$1 - 2 \exp\left(-\frac{1}{2} N \delta^2 e^{-\delta} |1-\sigma|^2\right) - \exp\left(-2 \frac{|I| \epsilon^2}{N}\right) - 2 \exp\left\{-\frac{c \epsilon t}{4} |I| \ln \frac{1}{\sigma}\right\} - 4e^{-nt^2} \quad (6)$$

with an absolute constant c .

The proof of Theorem 2.1 is given in Section 3.

To unpack the implications of Theorem 2.1, let us consider an asymptotic regime where the error bound (5) is arbitrarily small and the success probability bound (6) is arbitrarily close to one.

For the error bound to be small, we fix $\epsilon > 0, t > 0$ and let

$$\nu < (1+t)^{-2}/2 \tag{7}$$

which can be arbitrarily small.

Next we set $\delta = c_0\sigma$ where c_0 is a positive constant and let $\sigma \ll 1$. The error bound (5) becomes

$$\|x_0\|^{-2}\|x_0x_0^* - x_{\text{null}}x_{\text{null}}^*\| \leq c\sqrt{\sigma} \tag{8}$$

where c is a constant.

To ensure the success probability is close to one, we let $n \gg 1$ and $|I|^2/N \gg 1$. As a result, the second and third term in (6) are bounded from above by a term of the form

$$c_1 \exp(-c_2|I|^2/N) \tag{9}$$

for some constants c_1, c_2 . Moreover, since $|I| \gg |I|^2/N$, the third term is also bounded from above by a term like (9).

In summary, with $\epsilon > 0, t > 0$ fixed and arbitrary ν bounded by (7) we obtain the following estimate.

Corollary 2.2. *Under*

$$1 \ll n < |I| \ll N \ll |I|^2 \tag{10}$$

we have the error bound

$$\|x_0\|^{-2}\|x_0x_0^* - x_{\text{null}}x_{\text{null}}^*\| \leq c\sqrt{\frac{|I|}{N}} \tag{11}$$

with probability at least

$$1 - c_1 \exp(-c_2|I|^2/N) - 4e^{-nt^2/2}$$

for some constants c, c_1, c_2, t .

In other words, under the regime (10), the relative error (11) is arbitrarily small with probability exponentially (in n and $|I|^2/N$) close to 1.

Next, we describe a power method for computing the null vector.

3 The null vector algorithm

For a full rank A , let $A^* = QR$ be the QR-decomposition of A^* where Q is isometric and R is a full-rank, upper-triangular square matrix. Let $z = Rx$, $z_0 = Rx_0$ and $z_{\text{null}} = Rx_{\text{null}}$. Clearly, z_{null} is the null vector for the isometric phase retrieval problem $b = |Qz|$.

Denote the sub-column matrices of Q corresponding to the index sets I and I_c (the complement of I), respectively, by Q_I and Q_{I_c} .

Let

$$z_{\text{null}} := \arg \min \left\{ \sum_{i \in I} \|Q_I z\|^2 : \|z\| = \|b\| \right\}. \quad (12)$$

which is assumed to have the optimal global phase.

$$z_{\text{dual}} := \arg \max \left\{ \|Q_{I_c} x\|^2 : \|x\| = \|b\| \right\} \quad (13)$$

whose phase factor is optimally adjusted as z_{null} .

We have

$$\|Q_I^* z\|^2 + \|Q_{I_c}^* z\|^2 = \|b\|^2$$

and hence

$$z_{\text{null}} = z_{\text{dual}}. \quad (14)$$

Eq. (14) can be used to construct the null vector from $Q_{I_c}^* Q_{I_c}$ by the power method as in Algorithm 1 where $\mathbf{1}_c$ is the indicator function of I_c and z_{rand} denote random initialization whose pixels are given by, e.g. independent uniform random variables over $[0, 1]$.

Algorithm 1: The null vector

- 1 **Random initialization:** $z_1 = z_{\text{rand}}$
 - 2 **Loop:**
 - 3 **for** $k = 1 : k_{\text{max}} - 1$ **do**
 - 4 $z'_k \leftarrow Q^*(\mathbf{1}_c \odot Qz_k)$;
 - 5 $z_{k+1} \leftarrow z'_k / \|z'_k\|$
 - 6 **end**
 - 7 **Output:** $\hat{z}_{\text{dual}} = z_{k_{\text{max}}} \|b\| / \|z_{\text{rand}}\|$.
-

3.1 The spectral method

Here we compare the null vector method with the spectral method in [3] and [2].

The spectral vector method [2, 3, 8] is another linear algebraic method and uses the leading singular vector x_{spec} of $B^* = \text{diag}[b]A^*$ to approximate x_0 where

$$x_{\text{spec}} := \arg \max \{ \|B^*x\|^2 : x \in \mathbb{C}^n, \|x\| = \|x_0\| \}.$$

Algorithm 2: The spectral vector

- 1 **Random vector:** $x_1 = x_{\text{rand}}$
 - 2 **Loop:**
 - 3 **for** $k = 1 : k_{\text{max}} - 1$ **do**
 - 4 $x'_k \leftarrow A(|b|^2 \odot A^*x_k)$;
 - 5 $x_{k+1} \leftarrow [x'_k]_{\mathcal{X}} / \|[x'_k]_{\mathcal{X}}\|$;
 - 6 **end**
 - 7 **Output:** $x_{\text{spec}} = x_{k_{\text{max}}} \|x_0\| / \|x_{\text{rand}}\|$.
-

The key difference between Algorithms 1 and 2 is the different weights used in step 4 where the null vector method uses $\mathbf{1}_c$ and the spectral vector method uses $|b|^2$ (Algorithm 2). The truncated spectral vector method uses a still different weighting

$$x_{\text{t-spec}} = \arg \max_{\|x\|=1} \|A(\mathbf{1}_\tau \odot |b|^2 \odot A^*x)\| \quad (15)$$

where $\mathbf{1}_\tau$ is the characteristic function of the set

$$\{i : |A^*x(i)| \leq \tau \|b\|\} \quad (16)$$

with a thresholding parameter τ . Both γ of Algorithm 1 and τ of (15) can be optimized by tracking and minimizing the residual $\|b - |A^*x_k|\|$.

The performance guarantee for the spectral method (Theorem 4.1 of [8] and Theorem 3.3 of [3]) is weaker than Theorem 2.1. Specifically, for some constant C_1 and

$$N = \frac{C_1}{c^2} n \ln^3(n), \quad n \gg 1, \quad (17)$$

the spectral method initialization achieves the accuracy

$$\|x_0\|^{-2} \|x_0 x_0^* - x_{\text{spec}} x_{\text{spec}}^*\| \leq \sqrt{c},$$

where the positive constant c is arbitrary, with probability at least $1 - 4/N^2$ (Theorem 4.1 of [8]). In comparison, for some $C_1 > 0$ and arbitrary $C > 1$ and $|I| = Cn$, Corollary 2.2 implies that

$$\|x_0\|^{-2} \|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\| \leq C_1 (\ln n)^{-3/2}$$

with probability at least exponentially (in n) close to one.

Compared to the spectral method, the null vector method uses less information from the measured data (b versus I_c) but counterintuitively produces more accurate estimate of the object (see Section 5). Moreover, because the null vector method depends only on the choice of the index set I and not explicitly on b , the method is more stable to measurement noise.

4 Proof of Theorem 2.1

Let us begin with the following linear algebraic inequality.

Proposition 4.1. *There exists $x_\perp \in \mathbb{C}^n$ with $x_\perp^* x_0 = 0$ and $\|x_\perp\| = \|x_0\| = 1$ such that*

$$\|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\|^2 \leq \frac{2\|b_I\|^2}{\|A_I^* x_\perp\|^2}. \quad (18)$$

Proof. Since x_{null} is optimally phase-adjusted, we have

$$\beta := x_0^* x_{\text{null}} \geq 0 \quad (19)$$

and

$$x_0 = \beta x_{\text{null}} + \sqrt{1 - \beta^2} z \quad (20)$$

for some unit vector $z^* x_{\text{null}} = 0$. Then

$$x_\perp := -(1 - \beta^2)^{1/2} x_{\text{null}} + \beta z \quad (21)$$

is a unit vector satisfying $x_0^* x_\perp = 0$. Since x_{null} is a singular vector and z belongs in another singular subspace, we have

$$\begin{aligned} \|A_I^* x_0\|^2 &= \beta^2 \|A_I^* x_{\text{null}}\|^2 + (1 - \beta^2) \|A_I^* z\|^2, \\ \|A_I^* x_\perp\|^2 &= (1 - \beta^2) \|A_I^* x_{\text{null}}\|^2 + \beta^2 \|A_I^* z\|^2 \end{aligned}$$

from which it follows that

$$\begin{aligned} &\|A_I^* x_0\|^2 - (1 - \beta^2) \|A_I^* x_\perp\|^2 \\ &= \beta^2 \|A_I^* x_{\text{null}}\|^2 + (1 - \beta^2)^2 (\|A_I^* z\|^2 - \|A_I^* x_{\text{null}}\|^2) \geq 0. \end{aligned} \quad (22)$$

By (22), (3) and $\|b_I\| = \|A_I^* x_0\|$, we also have

$$\frac{\|b_I\|^2}{\|A_I^* x_\perp\|^2} \geq 1 - \beta^2 = \frac{1}{2} \|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\|^2. \quad (23)$$

□

In view of (18), we seek to give an upper bound on $\|b_I\|$ and a lower bound on $\|A_I^* x_\perp\|$ as follows.

Without loss of generality we may assume $\|x_0\| = 1$. Otherwise, we replace x_0, x_{null} by $x_0/\|x_0\|$ and $x_{\text{null}}/\|x_0\|$, respectively. Let $Q = [Q_1 \ Q_2 \ \cdots \ Q_n]$ be a unitary transformation where $Q_1 = x_0$ or equivalently $x_0 = Q e_1$ where e_1 is the canonical vector with 1 as the first entry and zero elsewhere. Since unitary transformations do not affect the covariance structure of Gaussian random vectors, the matrix $A^* Q$ is distributed as the standard complex Gaussian ensemble.

Proposition 4.2. *Let I be any set such that $b(i) \leq b(j)$ for all $i \in I$ and $j \in I_c = \{1, 2, \dots, N\} \setminus I$. For any unitary matrix Q , let $A' \in \mathbb{C}^{|I| \times (n-1)}$ be the sub-column matrix of A_I^*Q with its first column vector deleted. Then A' is distributed as the standard complex Gaussian ensemble.*

Proof. First note that $A_I^*Q = (A^*Q)_I$, the row submatrix of A^*Q indexed by I . As noted already, A^*Q is distributed as the standard complex Gaussian ensemble.

Since $x_0 = Qe_1$ and $b = |A^*Qe_1|$, I and I_c are entirely determined by the first column of A^*Q which is independent of the other columns of A^*Q . Consequently, the probability law of A' conditioned on the choice of I equals the probability law of A' for a fixed I . Therefore, A' is distributed as the standard complex Gaussian ensemble. \square

Let $\{\nu_i\}_{i=1}^{n-1}$ be the singular values of A' in the ascending order. For any $z \in \mathbb{C}^{n-1}$ the matrix

$$B' := A' \text{diag}(z/|z|)$$

has the same set of singular values as A' . Again, we adopt the convention that $z(j)/|z(j)| = 1$ when $z(j) = 0$. We have

$$\|A'z\| = \|B'z\|$$

and hence

$$\|A'z\| = (\|\Re(B')z\|^2 + \|\Im(B')z\|^2)^{1/2} \geq \sqrt{2} (\|\Re(B')z\| \wedge \|\Im(B')z\|).$$

The following statement gives the desired upper bound on $\|b_I\|$.

Proposition 4.3. *For any $\epsilon > 0, \delta > 0, t > 0$*

$$\|b_I\|^2 \leq |I| \left(\left(\frac{2+t}{1-\epsilon} \right) \frac{|I|}{N} + \epsilon \left(-2 \ln \left(1 - \frac{|I|}{N} \right) + \delta \right) \right)$$

with probability at least

$$1 - 2 \exp(-N\delta^2 e^{-\delta} |1 - \sigma|^2 / 2) - 2 \exp(-2\epsilon^2 |1 - \sigma|^2 \sigma^2 N) - Q \quad (24)$$

where Q has the asymptotic upper bound

$$2 \exp \left\{ -c \min \left[\frac{e^2 t^2 |I|^2}{16 N} (\ln \sigma^{-1})^2, \frac{et}{4} |I| \ln \sigma^{-1} \right] \right\}, \quad \sigma := \frac{|I|}{N} \ll 1.$$

The proof of Proposition 4.3 is given in Section 4.1.

The lower bound on $\|A_I^*x_\perp\|$ is given by the theory of Wishart matrices [10, 11]. The singular values $\{\nu_j^R\}_{j=1}^{n-1}, \{\nu_j^I\}_{j=1}^{n-1}$ (in the ascending order) of $\Re(B'), \Im(B')$ satisfy the probability

bounds that for every $t > 0$ and $j = 1, \dots, n-1$

$$\mathbb{P}\left(\sqrt{|I|} - (1+t)\sqrt{n} \leq \nu_j^R \leq \sqrt{|I|} + (1+t)\sqrt{n}\right) \geq 1 - 2e^{-nt^2/2}, \quad (25)$$

$$\mathbb{P}\left(\sqrt{|I|} - (1+t)\sqrt{n} \leq \nu_j^I \leq \sqrt{|I|} + (1+t)\sqrt{n}\right) \geq 1 - 2e^{-nt^2/2}. \quad (26)$$

If $x_\perp \perp x_0$, then $x_\perp = (0, z^\top)^\top$ with $z \in \mathbb{C}^{n-1}$. By Proposition 4.1 and (25)-(26), we have for some $z \in \mathbb{C}^{n-1}, \|z\| = 1$ that

$$\begin{aligned} \|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\| &\leq \frac{\|b_I\|}{\|\Re(B')|z|\| \wedge \|\Im(B')|z|\|} \\ &\leq \|b_I\| (\nu_{n-1}^R \wedge \nu_{n-1}^I)^{-1} \\ &\leq \|b_I\| (\sqrt{|I|} - (1+t)\sqrt{n})^{-1}. \end{aligned}$$

By Proposition 4.3, we obtain the desired bound (5). The success probability is at least the expression (24) minus $4e^{-nt^2/2}$.

Now let

4.1 Proof of Proposition 4.3

By the Gaussian assumption, $b(i)^2 = |a_i^* x_0|^2$ has a chi-squared distribution with the probability density $e^{-z/2}/2$ on $z \in [0, \infty)$ and the cumulative distribution

$$F(\tau) := \int_0^\tau 2^{-1} \exp(-z/2) dz = 1 - \exp(-\tau/2).$$

Let

$$\tau_* = -2 \ln(1 - |I|/N) \quad (27)$$

for which $F(\tau_*) = |I|/N$.

Define

$$\hat{I} := \{i : b(i)^2 \leq \tau_*\} = \{i : F(b^2(i)) \leq |I|/N\},$$

and

$$\|\hat{b}\|^2 := \sum_{i \in \hat{I}} b(i)^2.$$

Let

$$\{\tau_1 \leq \tau_2 \leq \dots \leq \tau_N\}$$

be the sorted sequence of $\{b(1)^2, \dots, b(N)^2\}$ in magnitude.

Proposition 4.4. (i) For any $\delta > 0$, we have

$$\tau_{|I|} \leq \tau_* + \delta \quad (28)$$

with probability at least

$$1 - \exp\left(-\frac{N}{2}\delta^2 e^{-\delta}|1 - |I|/N|^2\right) \quad (29)$$

(ii) For each $\epsilon > 0$, we have

$$|\hat{I}| \geq |I|(1 - \epsilon) \quad (30)$$

or equivalently,

$$\tau_{\lfloor |I|(1-\epsilon) \rfloor} \leq \tau_* \quad (31)$$

with probability at least

$$1 - 2 \exp(-4\epsilon^2|1 - |I|/N|^2|I|^2/N) \quad (32)$$

Proof. (i) Since $F'(\tau) = \exp(-\tau/2)/2$,

$$|F(\tau + \epsilon) - F(\tau)| \geq \epsilon/2 \exp(-(\tau + \epsilon)/2). \quad (33)$$

For $\delta > 0$, let

$$\zeta := F(\tau_* + \delta) - F(\tau_*)$$

which by (33) satisfies

$$\zeta \geq \frac{\delta}{2} \exp(-\frac{1}{2}(\tau_* + \delta)). \quad (34)$$

Let $\{w_i : i = 1, \dots, N\}$ be the i.i.d. indicator random variables

$$w_i = \chi_{\{b(i)^2 > \tau_* + \delta\}}$$

whose expectation is given by

$$\mathbb{E}[w_i] = 1 - F(\tau_* + \delta).$$

The Hoeffding inequality yields

$$\begin{aligned} \mathbb{P}(\tau_{|I|} > \tau_* + \delta) &= \mathbb{P}\left(\sum_{i=1}^N w_i > N - |I|\right) \\ &= \mathbb{P}\left(N^{-1} \sum_{i=1}^N w_i - \mathbb{E}[w_i] > 1 - |I|/N - \mathbb{E}[w_i]\right) \\ &= \mathbb{P}\left(N^{-1} \sum_{i=1}^N w_i - \mathbb{E}[w_i] > \zeta\right) \\ &\leq \exp(-2N\zeta^2). \end{aligned} \quad (35)$$

Hence, for any fixed $\delta > 0$,

$$\tau_{|I|} \leq \tau_* + \delta \quad (36)$$

holds with probability at least

$$\begin{aligned} 1 - \exp(-2N\zeta^2) &\geq 1 - \exp\left(-\frac{N\delta^2}{2}e^{-\tau_*-\delta}\right) \\ &= 1 - \exp\left(-\frac{N\delta^2}{2}e^{-\delta}|1 - |I|/N|^2\right) \end{aligned}$$

by (34).

(ii) Consider the following replacement

$$\begin{aligned} (a) \quad &|I| \longrightarrow \lceil |I|(1 - \epsilon) \rceil \\ (b) \quad &\tau_* \longrightarrow F^{-1}(\lceil |I|(1 - \epsilon) \rceil / N) \\ (c) \quad &\delta \longrightarrow F^{-1}(|I|/N) - F^{-1}(\lceil |I|(1 - \epsilon) \rceil / N) \\ (d) \quad &\zeta \longrightarrow F^{-1}(\tau_* + \delta) - F^{-1}(\tau_*) = |I|/N - \lceil |I|(1 - \epsilon) \rceil / N = \frac{\lfloor |I|\epsilon \rfloor}{N} \end{aligned}$$

in the preceding argument. Then (35) becomes

$$\mathbb{P}(\tau_{\lceil |I|(1-\epsilon) \rceil} > F^{-1}(|I|/N)) \leq \exp(-2N\zeta^2) = \exp\left(-\frac{2\lfloor |I|\epsilon \rfloor^2}{N}\right).$$

That is,

$$\tau_{\lceil |I|(1-\epsilon) \rceil} \leq \tau_*$$

holds with probability at least

$$1 - \exp(-2\lfloor |I|\epsilon \rfloor^2 / N).$$

□

Proposition 4.5. *For each $\epsilon > 0$ and $\delta > 0$,*

$$\frac{\|b_I\|^2}{|I|} \leq \frac{\|\hat{b}\|^2}{|\hat{I}|} + \epsilon(\tau_* + \delta) \quad (37)$$

with probability at least

$$1 - 2 \exp\left(-\frac{1}{2}\delta^2 e^{-\delta}|1 - |I|/N|^2 N\right) - 2 \exp\left(-2\epsilon^2|1 - |I|/N|^2 \frac{|I|^2}{N}\right). \quad (38)$$

Proof. Since $\{\tau_j\}$ is an increasing sequence, the function $T(m) = m^{-1} \sum_{i=1}^m \tau_i$ is also increasing. Consider the two alternatives either $|I| \geq |\hat{I}|$ or $|\hat{I}| \geq |I|$. For the latter,

$$\|b_I\|^2 / |I| \leq \|\hat{b}\|^2 / |\hat{I}|$$

due to the monotonicity of T .

For the former case $|I| \geq |\hat{I}|$, we have

$$\begin{aligned} T(|I|) &= |I|^{-1} \left(\sum_{i=1}^{|\hat{I}|} \tau_i + \sum_{i=|\hat{I}+1}^{|I|} \tau_i \right) \\ &\leq T(|\hat{I}|) + |I|^{-1}(|I| - |\hat{I}|)\tau_{|I|}. \end{aligned}$$

By Proposition 4.4 (ii) $|\hat{I}| \geq (1 - \epsilon)|I|$ and hence

$$T(|I|) \leq T(|\hat{I}|) + |I|^{-1}(|I| - |I|(1 - \epsilon))\tau_{|I|} = T(|\hat{I}|) + \epsilon\tau_{|I|}$$

with probability at least given by (32).

By Proposition 4.4 (i), $\tau_{|I|} \leq \tau_* + \delta$ with probability at least given by (29). \square

Continuing the proof of Proposition 4.3, let us consider the i.i.d. centered, bounded random variables

$$Z_i := \frac{N^2}{|I|^2} [b(i)^2 \chi_{\tau_*} - \mathbb{E}[b(i)^2 \chi_{\tau_*}]] \quad (39)$$

where χ_{τ_*} is the characteristic function of the set $\{b(i)^2 \leq \tau_*\}$. Note that

$$\mathbb{E}(b(j)^2 \chi_{\tau_*}) = \int_0^{\tau_*} 2^{-1} z \exp(-z/2) dz = 2 - (\tau_* + 2) \exp(-\tau_*/2) \leq 2|I|^2/N^2 \quad (40)$$

and hence

$$-2 \leq Z_i \leq \sup \left\{ \frac{N^2}{|I|^2} b(i)^2 \chi_{\tau_*} \right\} = \frac{N^2}{|I|^2} \tau_*. \quad (41)$$

Next recall the Bernstein-inequality.

Proposition 4.6. [11] *Let Z_1, \dots, Z_N be i.i.d. centered sub-exponential random variables. Then for every $t \geq 0$, we have*

$$\mathbb{P} \left\{ N^{-1} \left| \sum_{i=1}^N Z_i \right| \geq t \right\} \leq 2 \exp \left\{ -c \min(Nt^2/K^2, Nt/K) \right\}, \quad (42)$$

where c is an absolute constant and

$$K = \sup_{p \geq 1} p^{-1} (\mathbb{E}|Z_j|^p)^{1/p}.$$

Remark 4.7. For K we have the following estimates

$$\begin{aligned}
K &\leq \frac{2N^2}{|I|^2} \sup_{p \geq 1} p^{-1} (\mathbb{E} |b(i)^2 \chi_{\tau_*}|^p)^{1/p} \\
&\leq \frac{2N^2}{|I|^2} \tau_* \sup_{p \geq 1} p^{-1} (\mathbb{E} \chi_{\tau_*})^{1/p} \\
&\leq \frac{2N^2}{|I|^2} \tau_* \sup_{p \geq 1} p^{-1} (1 - e^{-\tau_*/2})^{1/p}.
\end{aligned} \tag{43}$$

The maximum of the right hand side of (43) occurs at

$$p_* = -\ln(1 - e^{-\tau_*/2})$$

and hence

$$K \leq \frac{2N^2}{|I|^2} \frac{\tau_*}{p_*} (1 - e^{-\tau_*/2})^{1/p_*}.$$

We are interested in the regime

$$\tau_* \asymp 2|I|/N \ll 1$$

which implies

$$p_* \asymp -\ln \frac{\tau_*}{2} \asymp \ln \frac{N}{|I|}$$

and consequently

$$K \leq \frac{4N}{e|I|} \left(\ln \frac{N}{|I|} \right)^{-1}, \quad \sigma = |I|/N \ll 1. \tag{44}$$

On the other hand, upon substituting the asymptotic bound (44) in the probability bound

$$Q = 2 \exp \left\{ -c \min(Nt^2/K^2, Nt/K) \right\}$$

of (42), we have

$$K \leq 2 \exp \left\{ -c \min \left[\frac{e^2 t^2}{16} (\ln \sigma^{-1})^2 |I|^2/N, \frac{et}{4} |I| \ln \sigma^{-1} \right] \right\}, \quad \sigma \ll 1.$$

The Bernstein inequality ensures that with high probability

$$\left| \frac{\|\hat{b}\|^2}{N} - \mathbb{E}(b^2(i) \chi_{\tau_*}) \right| \leq t \frac{|I|^2}{N^2}.$$

By (30) and (40), we also have

$$\begin{aligned}
\frac{\|\hat{b}\|^2}{|\hat{I}|} &\leq \mathbb{E}(b(i)^2 \chi_{\tau_*}) \frac{N}{|\hat{I}|} + t \frac{|I|^2}{|\hat{I}|N} \\
&\leq \left(\mathbb{E}(b(i)^2 \chi_{\tau_*}) \frac{N^2}{|I|^2} + t \right) \frac{|I|}{N} \\
&\leq \frac{2+t}{1-\epsilon} \cdot \frac{|I|}{N}
\end{aligned} \tag{45}$$

By Prop. 4.5, we now have

$$\|b_I\|^2 \leq |I| \left(\frac{\|\hat{b}\|^2}{|\hat{I}|} + \epsilon(\tau_* + \delta) \right)$$

with probability at least given by (6), which together with (45) and (27) complete the proof of Proposition 4.3.

5 Numerical experiments

In this section we test numerically the null vector method and the spectral vector method. Let $p_k, q_k, k = 1, \dots, n$ be independent standard normal random variables and define the following three types of signals x_0 :

$$\textbf{White noise} \quad x_0(t) = \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} (p_k + iq_k) e^{i2\pi k(t-1)/n}, \quad t = 0, 1, \dots, n-1; \tag{46}$$

$$\textbf{Low-pass signal} \quad x_0(t) = \sum_{k=-\frac{n}{8}}^{\frac{n}{8}-1} (p_k + iq_k) e^{i2\pi k(t-1)/n}, \quad t = 0, 1, \dots, n-1; \tag{47}$$

Randomly phased Phantom (RPP) signal: x_0 is the vector form of the $\lceil \sqrt{n} \rceil \times \lceil \sqrt{n} \rceil$ positive-valued Phantom with phase at each pixel being independently and uniformly distributed over $[0, 2\pi)$.

5.1 Convergence test

First we test the capability of the null vector method as a stand-alone method for phase retrieval under the regime (10) of Corollary 2.2. In particular, we would like to see how sharp the error bound (11) is.

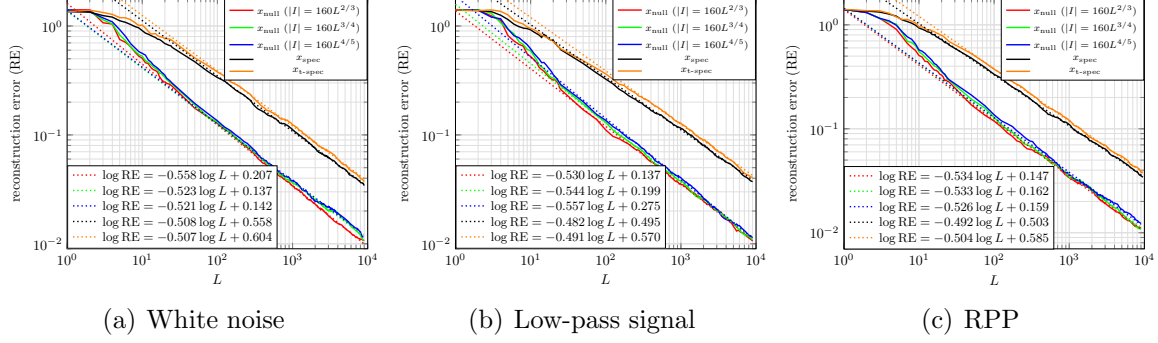


Figure 1: Log-log plot of relative error (RE) of the null vector method with $\alpha = 4/5$ (blue), $\alpha = 3/4$ (green), $\alpha = 2/3$ (red), the spectral (black) and the truncated spectral method (yellow) vs. $L \leq 10^4$. The legend at the top describes the results of linear regression.

To this end and for convenience of simulation, we let $N = Ln$ and $|I| = nL^\alpha$ with $n = 160$ and $\alpha \in (0.5, 1)$. The number L is called the oversampling ratio.

For $L_{\max} = 10^4$ we generate the largest measurement matrix $A_{\max} := [a_j]_{j=1, \dots, nL_{\max}}$ and the corresponding measurements $b_{\max} := |A_{\max}^* x_0|$. For $L < L_{\max}$ the measurement matrix A consists of the first nL columns of A_{\max} and b the first nL components of b_{\max} .

In Fig. 1 we plot the logarithm of the relative error (RE)

$$\text{RE} := \|x_0\|^{-2} \|x_0 x_0^* - \hat{x} \hat{x}^*\|_2 \quad (48)$$

where $\hat{x} = x_{\text{null}}, x_{\text{spec}}$ or $x_{\text{t-spec}}$, as a function of $\log L$. We use the data points for $L \geq 25$ to estimate the slope and the intercept of the linear regression lines (dotted lines). The estimated slope for the null vector method is less than -0.5 for $\alpha = 2/3, 3/4, 4/5$ and all three test signals, indicating overestimation of (11), and the estimated slope is also less than those for the spectral method and its truncated version. For every L , the error of the null vector method is significantly smaller than that of the spectral method. Here the thresholding parameter in (16) is chosen to remove the top 1% (in magnitude) among the components of b . In this test, however, the truncation does not improve the performance of the spectral method.

5.2 Initialization test

In practice, the oversampling ratio L is usually small and hence we can not expect the null vector method to produce an accurate estimate. According to [1], the phase retrieval map $\mathcal{M}(x) = |A^* x|$ is injective for *generic* complex-valued A^* if $N \geq 4n - 2$. So below we consider the borderline case $L = 4$.

For computational consistency, we enforce the following nested structure in the test signal and measurement matrix. For $n \in [n_{\min}, n_{\max}]$, a test signal $x_0^{(n_{\max})} \in \mathbb{C}^{n_{\max}}$ is generated in

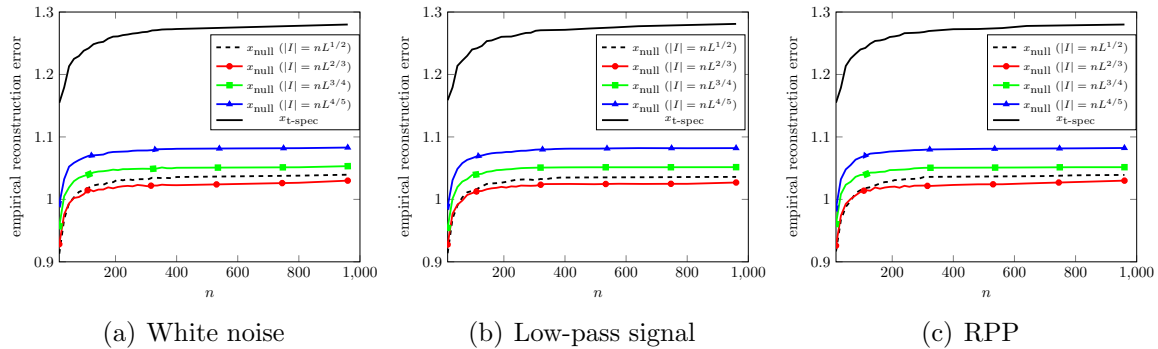


Figure 2: RE vs the dimension n of the object vector with the oversampling ratio $L = 4$.

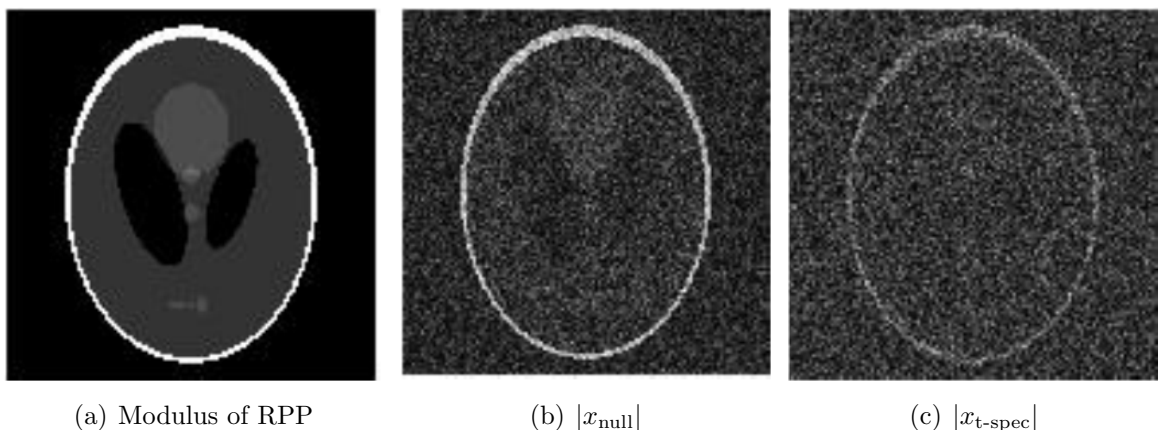


Figure 3: The initialization for (a) the original 120×120 RPP by (b) the null vector method and (c) the truncated spectral method. REs corresponding to (b) and (c) are 1.0270 and 1.3063, respectively.

advance and fixed for the rest of trials. The test signal x_0 of length $n < n_{\max}$ is a sub-vector of $x_0^{(n_{\max})}$:

$$x_0(j) = x_0^{(n_{\max})}(j), \quad j = 1, 2, \dots, n. \quad (49)$$

For $L = 4$, the reconstruction error is highly fluctuating and to smooth out the error curves we average the reconstruction errors from many trials. We randomly generate J complex Gaussian matrices $\{A_{\max}^{(j)}\}_{j=1,2,\dots,J} \subset \mathbb{C}^{n_{\max} \times Ln_{\max}}$, where the superscript j is the index of the realization. For $n < n_{\max}$, $A^{(j)}$ is the upper left $n \times nL$ submatrix of $A_{\max}^{(j)}$ for each $j = 1, \dots, J$.

Fig. 2 shows the average REs (with $J = 10^4$) versus n for $L = 4$ and various α . For the null vector method, the simple rule of using the median value $|I| = 0.5N$ is effective for small L (corresponding to $\alpha = 1/2$ for $L = 4$). The spectral method produces poor results which are not shown. Here the truncation trick improves the performance of the spectral method. We see that the null vector method significantly outperforms the truncated spectral method for

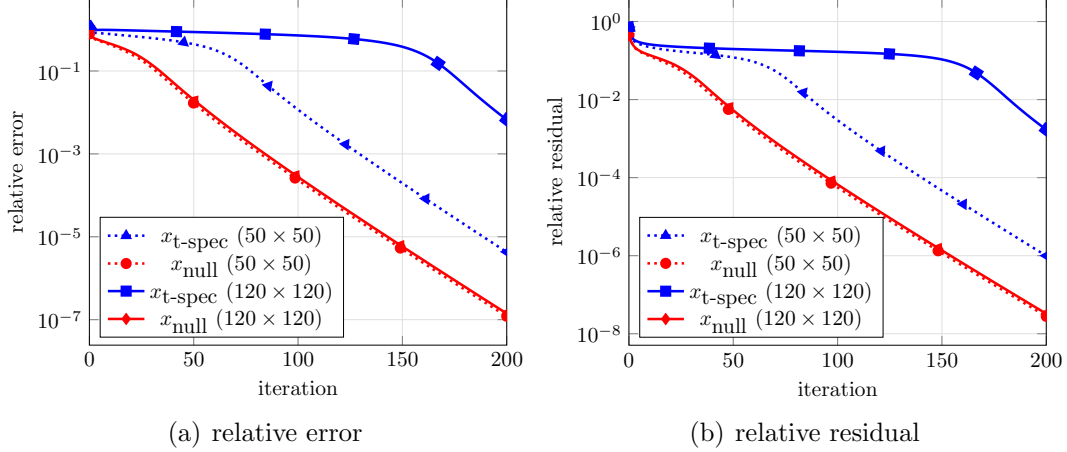


Figure 4: AP reconstruction errors with the initializations x_{null} and $x_{t\text{-spec}}$ for the 50×50 and 120×120 RPPs.

small L .

To better visualize the result of an individual trial, we show the moduli of the original 120×120 RPP, the truncated spectral and the null vector reconstructions in Fig 3. We then use these results as initialization for the alternating projection (AP) or error reduction (ER) [6, 7] and plot the REs, on the logarithmic scale, of the AP in Fig. 4. For the reader's convenience, we describe the version of AP that is implemented.

Step 1: Perform the QR factorization $A^* = QR$ where $Q \in \mathbb{C}^{N \times n}$ is isometric and $R \in \mathbb{C}^{n \times n}$ is upper triangular. For an i.i.d. Gaussian matrix A^* , R is invertible almost surely.

Step 2: Let $\xi = Rx$ and rewrite $|A^*x| = b$ as $|Q\xi| = b$.

Step 3: Perform the iteration

$$\xi^{k+1} = Q^* \left(b \odot \frac{Q\xi^k}{|Q\xi^k|} \right) \quad (50)$$

for $k = 0, 1, 2, \dots$ with $\xi^0 = Rx_{\text{null}}$ or $Rx_{t\text{-spec}}$. Let $x^k = R^{-1}\xi^k$.

Clearly, all iteration processes exhibited in Fig. 4 eventually enter the geometric convergence regime with the null vector initialization entering at much earlier times. More importantly, the null vector method does so essentially independent of the dimension of the object vector while the results of the truncated spectral method are sensitive to the dimension of the object vector .

6 Conclusion and discussion

We have proposed an approximate method (the null vector method) for solving the phase retrieval problem and given a performance guarantee in the case of independent Gaussian measurement (Theorem 2.1). We have further identified an asymptotic regime for which the reconstruction error tends to zero with probability exponentially close to one (Corollary 2.2). This is an improvement over the performance guarantee for the spectral method and its truncated version.

In numerical experiments, the null vector method significantly outperforms the spectral method for various test signals and exhibits a power-law behavior, in the limit of large oversampling ratio, that is consistent with, but more optimistic than, our theoretical prediction (Corollary 2.2). It remains an open problem as to what the optimal scaling law of the reconstruction error should be.

Numerical experiments further show that, in the opposite regime of small oversampling ratio, the null vector method also has a superior performance as initialization for iterative phase retrieval algorithms.

In the case of phase retrieval with coded diffraction patterns, the null vector method continues to perform well regardless of structured measurement matrices inherent in this context [5]. Also, because the null vector method depends only on the choice of the index set I and not explicitly on b , it is more stable to measurement noise.

Acknowledgements. Research of P. Chen is supported in part by the grant 103-2115-M-005-006-MY2 from Ministry of Science and Technology, Taiwan, and US NIH grant U01-HL-114494. Research of A. Fannjiang is supported in part by US National Science Foundation grant DMS-1413373 and Simons Foundation grant 275037.

References

- [1] R. Balan, P. Casazza and D. Edidin, “On signal reconstruction without phase,” *Appl. Comput. Harmon. Anal.* **20**, 345–356 (2006).
- [2] E.J. Candès and Y. Chen, “Solving random quadratic systems of equations is nearly as easy as solving linear systems.” arXiv:1505.05114, 2015.
- [3] E.J. Candès, X. Li and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: theory and algorithms,” *IEEE Trans Inform. Th.* **61**(4), 1985–2007 (2015).
- [4] P. Chen and A. Fannjiang, “Phase retrieval with a single mask by the Douglas-Rachford algorithm,” *Appl. Comput. Harmon. Anal.* (2016) <http://dx.doi.org/10.1016/j.acha.2016.07.003>.

- [5] P. Chen, A. Fannjiang and G.-R. Liu, “Phase retrieval with one or two coded diffraction patterns by alternating projection with the null initialization,” [arxiv:1510.07379](https://arxiv.org/abs/1510.07379).
- [6] J. R. Fienup, “Phase retrieval algorithms: a comparison,” *Appl. Opt.* **21**, 2758-2769 (1982).
- [7] R. W. Gerchberg and W. O. Saxton, “A practical algorithm for the determination of the phase from image and diffraction plane pictures,” *Optik* **35**, 237-246 (1972).
- [8] P. Netrapalli, P. Jain, S. Sanghavi, “Phase retrieval using alternating minimization,” *IEEE Trans. Signal Proc.* **63** (2015), pp. 4814-4826.
- [9] Y. Shechtman, Y.C. Eldar, O. Cohen, H.N. Chapman, M. Jianwei and M. Segev, “Phase retrieval with application to optical imaging: A contemporary overview,” *IEEE Mag. Signal Proc.* **32**(3) (2015), 87 - 109.
- [10] J. A. Tropp, *An Introduction to Matrix Concentration Inequalities*, [arXiv:1501.01571](https://arxiv.org/abs/1501.01571).
- [11] R. Vershynin. “Introduction to the non-asymptotic analysis of random matrices.” In *Compressed sensing*, pages 210-268. Cambridge Univ. Press, Cambridge, 2012.