

PHASE RETRIEVAL BY LINEAR ALGEBRA*

PENGWEN CHEN[†], ALBERT FANNJIANG[‡], AND GI-REN LIU[§]

Abstract. The null vector method, based on a simple linear algebraic concept, is proposed as an initialization method for nonconvex approaches to the phase retrieval problem. For the stylized measurement with random complex Gaussian matrices, a nonasymptotic error bound is derived, stronger than that of the spectral vector method. Numerical experiments show that the null vector method also has a superior performance for the realistic measurement of coded diffraction patterns in coherent diffractive imaging.

Key words. phase retrieval, coded diffraction patterns, null initialization

AMS subject classifications. 49K35, 05C70, 90C08

DOI. 10.1137/16M1107747

1. Introduction. We consider the following phase retrieval problem: Let $A = [a_{ij}] \in \mathbb{C}^{n \times N}$ be a full-rank matrix. Let $x_0 \in \mathbb{C}^n$ and $y_0 = A^*x_0$. Suppose we are given A and $b := |y_0|$, where $|y_0|$ denotes the modulus vector with $|y|(j) = |y(j)|$ for all j . The aim of phase retrieval is to find x_0 .

For this 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 [3, 4, 6, 7] and references therein) and gradient-descent methods [1, 2]. Phase retrieval has a wide range of applications, including coherent diffractive imaging where A^* is a Fourier-like matrix and b a set of diffraction patterns (see [9] for a recent survey).

A key to the success of any nonconvex method is an effective initialization. The following consideration motivates our approach:

First we reformulate the phase retrieval problem in the *isometric* form. For a full rank A , let $A^* = QR$ be the QR-decomposition of A^* , where Q is isometric and R is an invertible upper-triangular square matrix. Let $z = Rx$ and $z_0 = Rx_0$. The phase retrieval problem is equivalent to finding a solution z_0 to $b = |Qz|$ and then recovering $x_0 = R^{-1}z_0$.

Now sort the signals in terms of their magnitudes and apply a threshold (to be determined) to separate the “weak” signals from the “strong” signals. Let $I \subset \{1, \dots, N\}$ be the support set of the weak signals and I_c its complement such that $b(i) \leq b(j)$ for all $i \in I, j \in I_c$. In other words, $\{b(i) : i \in I_c\}$ are the strong signals.

*Received by the editors December 13, 2016; accepted for publication (in revised form) by J. Tropp June 12, 2017; published electronically August 17, 2017.

<http://www.siam.org/journals/simax/38-3/M110774.html>

Funding: The work of the first author was supported in part by grant 103-2115-M-005-006-MY2 from the Ministry of Science and Technology, Taiwan, and U.S. NIH grant U01-HL-114494. The work of the second author was supported in part by U.S. National Science Foundation grant DMS-1413373 and Simons Foundation grant 275037.

[†]Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan (pengwen@nchu.edu.tw).

[‡]Corresponding author. Department of Mathematics, University of California, Davis, CA 95616 (fannjiang@math.ucdavis.edu).

[§]Department of Mathematics, National Cheng Kung University, Tainan, Taiwan (girenliu@mail.ncku.edu.tw).

Let Q_I and Q_{I_c} be the sub-row matrices of Q corresponding to the index sets I and I_c , respectively. 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. Clearly, $A_I^* = Q_I R$ and $A_{I_c}^* = Q_{I_c} R$. Let $b_I = |A_I^* x_0|$ and $b_{I_c} = |A_{I_c}^* x_0|$.

Let $|I|$ be the cardinality of the set I . We always assume $|I| \geq n$ so that A_I^* and Q_I have a trivial null space and hence preserve the information of x_0 .

Since $b_I = |Q_I z_0|$ is “weak,” the rows of Q_I are nearly orthogonal to z_0 . It is then natural to “linearize” the weak components of the phase retrieval problem and formulate it as the variational principle

$$(1) \quad x_{\text{null}} = R^{-1} z_{\text{null}}, \quad z_{\text{null}} \in \arg \min \{ \|Q_I z\|^2 : z \in \mathbb{C}^n, \|z\| = \|b\| \}$$

(which may have more than one minimizer) or, equivalently,

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

In view of the isometry property

$$(3) \quad \|z\|^2 = \|Q_I z\|^2 + \|Q_{I_c} z\|^2 = \|b\|^2,$$

minimizing $\|Q_I z\|^2$ is equivalent to maximizing $\|Q_{I_c} z\|^2$ over $\{z : \|z\| = \|b\|\}$. Therefore an equivalent alternative variational principle for the null vector is

$$(4) \quad x_{\text{null}} \in \arg \max \{ \|A_{I_c}^* x\|^2 : x \in \mathbb{C}^n, \|Rx\| = \|b\| \}.$$

Now (4) gives rise to the power method for constructing the null vector as in Algorithm 1.

Algorithm 1: The null vector

- 1 **Input:** QR-decomposition of A^* , I_c , $\|b\|$.
 - 2 **Initialization:** z_1
 - 3 **for** $k = 1, 2, 3, \dots$ **do**
 - 4 $z'_k \leftarrow Q^*(\mathbf{1}_c \odot Qz_k)$, where $\mathbf{1}_c$ is the indicator function of I_c
 - 5 $z_{k+1} \leftarrow z'_k / \|z'_k\|$
 - 6 **until** $\|z_{k+1} - z_k\|$ is sufficiently small.
 - 7 **end**
 - 8 **Output:** $x_{\text{null}} = \|b\| R^{-1} z_{\text{null}}$, $z_{\text{null}} = z_{k+1}$.
-

We can enhance the result of Algorithm 1 with the full information of b by the following procedure:

$$(5) \quad \tilde{x}_{\text{null}} = R^{-1} \tilde{z}_{\text{null}}, \quad \tilde{z}_{\text{null}} = Q^* \left[b \odot \frac{Qz_{\text{null}}}{|Qz_{\text{null}}|} \right].$$

The idea in (5) is to enhance the transform phase inherent in z_{null} (i.e., phase retrieval) with the measured information of the transform magnitude.

The key to the performance of Algorithm 1 hinges on the threshold $|I|$. To characterize, as precisely as we can, the conditions for a proper choice of $|I|$ in relation to N and n , we assume independent and identically distributed (i.i.d.) Gaussian measurements for a lower technical barrier of the proof as in [1, 2, 8] and other phase retrieval literature seeking performance guarantee. The randomness assumption is

in line with the empirical fact that increased randomness in the measurement matrix enhances the performance of Algorithm 1. For a more realistic setup, such as coherent diffractive imaging, this suggests introducing disorder into the structured, deterministic measurements (e.g., random masking).

To the end of proving the theoretical bound, we consider the following simplified version of the null vector:

$$(6) \quad \hat{x}_{\text{null}} \in \arg \min \{ \|A_I^* x\|^2 : x \in \mathbb{C}^n, \|x\| = \|x_0\| \},$$

which has a behavior similar to x_{null} when R is close to a scalar matrix. This is so when the oversampling ratio L of the i.i.d. Gaussian matrix is large or when the measurement matrix is isometric (hence $R = I$) as for coded diffraction patterns. Like x_{null} , \hat{x}_{null} can also be efficiently computed by the power method by iterating $\lambda - A_I A_I^*$, where the constant λ is chosen to be around the leading singular value of A_I .

Our theoretical analysis (Theorem 2.1) leads to a nonasymptotic error bound for \hat{x}_{null} as an estimate of x_0 and forms the basic guideline for the choice of $|I|$:

$$n < |I| < N < |I|^2.$$

In particular, this can be met for any oversampling ratio $L = N/n$ by the rule

$$(7) \quad |I| = \lceil nL^\alpha \rceil = \lceil n^{1-\alpha} N^\alpha \rceil, \quad \alpha \in [0.5, 1),$$

which yields the (relative) error bound $\mathcal{O}(L^{(\alpha-1)/2})$, with probability exponentially (in n) close to 1, which achieves the asymptotic minimum at $\alpha = 1/2$ (the geometric mean rule) for all n and N (Corollary 2.2 and Remark 2.1).

For phase retrieval with randomly coded diffraction patterns where the oversampling ratio L is small and the measurement matrix is isometric (hence $\hat{x}_{\text{null}} = x_{\text{null}}$), we demonstrate the superior performance of the null vector with the geometric mean rule ($\alpha = 1/2$)

$$(8) \quad |I| = \lceil nL^{1/2} \rceil = \lceil \sqrt{nN} \rceil$$

in section 5.2. Other, more ad hoc rules have also been found to perform well with coded diffraction patterns [4].

2. Nonasymptotic error bound. In this section, we present the nonasymptotic error bound. In addition to theoretical interest, the main purpose is practical: to characterize the conditions for a proper choice of $|I|$ in relation to N and n , through probabilistic analysis.

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

$$(9) \quad \min_{\theta \in \mathbb{R}} \|e^{i\theta} x_{\text{null}} - x_0\| = \sqrt{2(\|x_0\|^2 - |x_0^* x_{\text{null}}|)}.$$

Alternatively, we can use the error metric

$$(10) \quad \|x_0 x_0^* - x_{\text{null}} x_{\text{null}}^*\| = \sqrt{2(\|x_0\|^4 - |x_0^* x_{\text{null}}|^2)},$$

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

THEOREM 2.1. *Let $A = [a_{ij}] \in \mathbb{C}^{n \times N}$, where a_{ij} are i.i.d. circularly symmetric complex standard Gaussian random variables. Let $\sigma, \nu, \epsilon, \delta, t$ be any constants constrained as follows:*

$$(11) \quad \sigma := \frac{|I|}{N} < 1, \quad \nu = \frac{n}{|I|} < 1, \quad \epsilon \in (0, 1), \quad \delta > 0, \quad t \in (0, \nu^{-1/2} - 1).$$

Then for any $x_0 \in \mathbb{C}^n$ and \hat{x}_{null} given by (6) the error bound

$$(12) \quad \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\|^2 \leq \left[\left(\frac{2+t}{1-\epsilon} \right) \sigma + \epsilon (-2 \ln(1-\sigma) + \delta) \right] \frac{\|x_0\|^4}{(1 - (1+t)\sqrt{\nu})^2}$$

holds with probability at least

$$(13) \quad 1 - 2 \exp \left[-\frac{1}{2} N \delta^2 e^{-\delta} |1 - \sigma|^2 \right] - \exp \left[-2 \frac{|I| \epsilon^2}{N} \right] - 2 \exp \left[-\frac{cet}{4} |I| \ln \frac{1}{\sigma} \right] - 4e^{-nt^2/2}$$

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 a regime where the error bound (12) is arbitrarily small and the success probability bound (13) is exponentially close to one.

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

$$(14) \quad \nu < (1+t)^{-2}/2,$$

which can be arbitrarily small.

Next we set $\delta = c_0 \sigma$, where c_0 is a positive constant. The error bound (12) becomes

$$(15) \quad \|x_0\|^{-2} \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\| \leq c \sqrt{\sigma},$$

where c is a constant.

The second, third, and fourth terms in (13) are bounded from above by a term of the form

$$(16) \quad c_1 \exp [-c_2 |I|^2/N]$$

for some constants c_1, c_2 .

In summary, with $\epsilon > 0, t > 0$ fixed and arbitrary ν bounded by (14), we obtain the following estimate.

COROLLARY 2.2. *Under*

$$(17) \quad 1 < n < |I| < N < |I|^2$$

the error bound

$$(18) \quad \|x_0\|^{-2} \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\| \leq c \sqrt{\frac{|I|}{N}}$$

holds 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 .

Remark 2.1. In the case of large oversampling ratio $L = N/n \gg 1$, the relative error (18) is small with probability exponentially (in n and $|I|^2/N$) close to 1 if

$$(19) \quad 1 \ll n < |I| \ll N \ll |I|^2.$$

In particular, for

$$(20) \quad |I| = \lceil nL^\alpha \rceil, \quad \alpha \in [0.5, 1),$$

the error bound (18) becomes

$$(21) \quad \|x_0\|^{-2} \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\| \leq \frac{c}{L^{(1-\alpha)/2}}.$$

Our numerical test of (21) in section 5 confirms the scaling behavior but with an exponent slightly greater than 0.5 (Figure 1).

3. Comparison with the spectral methods. The spectral method [1, 2, 8] is another linear algebraic approach to initialization which uses the leading singular vector x_{spec} of $\text{diag}[b]A^*$:

$$(22) \quad x_{\text{spec}} \in \arg \max \{ \|\text{diag}[b]A^*x\|^2 : x \in \mathbb{C}^n, \|x\| = \|x_0\| \}.$$

The power method for computing (22) is given in Algorithm 2.

Algorithm 2: The spectral vector

1 **Input:** $A, b, \|x_0\|$.
 2 **Initialization:** x_1
 3 **for** $k = 1, 2, 3, \dots$ **do**
 4 $x'_k \leftarrow A(|b|^2 \odot A^*x_k)$;
 5 $x_{k+1} \leftarrow x'_k / \|x'_k\|$;
 6 **until** $\|x_{k+1} - x_k\|$ is sufficiently small.
 7 **end**
 8 **Output:** $\hat{x}_{\text{spec}} = x_k \|x_0\| / \|x_1\|$.

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 $|b|^2$. The truncated spectral vector method uses a still different weighting,

$$(23) \quad x_{\text{t-spec}} \in \arg \max_{\|x\|=\|x_0\|} \|A(\mathbf{1}_\tau \odot |b|^2 \odot A^*x)\|,$$

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

$$(24) \quad \left\{ i : b(i) \leq \tau \frac{\|b\|_2}{\sqrt{N}} \right\}$$

for some thresholding parameter τ .

The available performance guarantee for the spectral method is weaker than Theorem 2.1. For example, according to Theorem 4.1 of [8], for any given $c_0 > 0$ and

$$(25) \quad N = \frac{C_1}{c_0^2} n \ln^3 n, \quad n \gg 1,$$

with some constant C_1 independent of c_0 , the spectral method achieves the accuracy

$$(26) \quad \|x_0\|^{-2} \|x_0 x_0^* - x_{\text{spec}} x_{\text{spec}}^*\| \leq \sqrt{c_0},$$

with probability at least $1 - 4/N^2$, while, according to Theorem 3.3 of [2], the same estimate (26) holds with probability at least $1 - \mathcal{O}(n^{-2})$ for

$$(27) \quad N = C_2 n \ln n, \quad n \gg 1,$$

with a sufficiently large C_2 depending on c_0 .

In comparison, for $N = Cn \ln n$ and $|I| = Cn$ with any $C > 0$, Corollary 2.2 implies that

$$\|x_0\|^{-2} \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\| \leq \frac{c}{\sqrt{\ln n}},$$

with probability exponentially (in n) close to one.

As shown by the numerical tests in section 5, the null vector method counter-intuitively produces more accurate estimate of the signal by using less information (I_c vs. b). Moreover, because the null vector method depends only on the support set I , and not explicitly on b , the method is more stable to measurement noise (Figure 3).

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*

$$(28) \quad \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\|^2 \leq \frac{2\|b_I\|^2}{\|A_I^* x_{\perp}\|^2}.$$

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

$$(29) \quad \beta := x_0^* x_{\text{null}} \geq 0$$

and

$$(30) \quad x_0 = \beta \hat{x}_{\text{null}} + \sqrt{1 - \beta^2} z$$

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

$$(31) \quad x_{\perp} := -(1 - \beta^2)^{1/2} \hat{x}_{\text{null}} + \beta z$$

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

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

from which it follows that

$$(32) \quad \begin{aligned} &\|A_I^* x_0\|^2 - (1 - \beta^2) \|A_I^* x_{\perp}\|^2 \\ &= \beta^2 \|A_I^* \hat{x}_{\text{null}}\|^2 + (1 - \beta^2)^2 (\|A_I^* z\|^2 - \|A_I^* \hat{x}_{\text{null}}\|^2) \geq 0 \end{aligned}$$

since $\|A_I^* z\|^2 \geq \|A_I^* \hat{x}_{\text{null}}\|^2$ by the variational principle (6). By (32), (10), and $\|b_I\| = \|A_I^* x_0\|$, we also have

$$(33) \quad \frac{\|b_I\|^2}{\|A_I^* x_{\perp}\|^2} \geq 1 - \beta^2 = \frac{1}{2} \|x_0 x_0^* - \hat{x}_{\text{null}} \hat{x}_{\text{null}}^*\|^2. \quad \square$$

In view of (28), we seek to give an upper bound on $\|b_I\|$ and lower bound on $\|A_I^* x_\perp\|$ as follows.

Without loss of the generality we may assume $\|x_0\| = 1$. Otherwise, we replace $x_0, \hat{x}_{\text{null}}$ by $x_0/\|x_0\|$ and $\hat{x}_{\text{null}}/\|x_0\|$, respectively. Let $U = [U_1 \ U_2 \ \cdots \ U_n]$ be a unitary transformation where $U_1 = x_0$, or equivalently $x_0 = Ue_1$, 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^*U 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 U , let $A' \in \mathbb{C}^{I \times (n-1)}$ be the sub-column matrix of A_I^*U with its first column vector deleted. Then A' is distributed as the standard complex Gaussian ensemble.*

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

Since $x_0 = Ue_1$ and $b = |A^*Ue_1|$, I and I_c are entirely determined by the first column of A^*U , which is independent of the other columns of A^*U . 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 result on order statistics 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

$$(34) \quad 1 - 2 \exp(-N\delta^2 e^{-\delta} |1 - \sigma|^2 / 2) - 2 \exp(-2\epsilon^2 |1 - \sigma|^2 \sigma^2 N) - Q,$$

where Q has the asymptotic upper bound

$$(35) \quad 2 \exp \left\{ -c \min \left[\frac{e^2 t^2}{16} \frac{|I|^2}{N} (\ln \sigma^{-1})^2, \frac{et}{4} |I| \ln \sigma^{-1} \right] \right\}, \quad \sigma := \frac{|I|}{N} \ll 1.$$

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 $j = 1, \dots, n - 1$

$$(36) \quad \mathbb{P}\left(\sqrt{|I|} - (1+t)\sqrt{n} \leq \nu_j^R \leq \sqrt{|I|} + (1+t)\sqrt{n}\right) \geq 1 - 2e^{-nt^2/2},$$

$$(37) \quad \mathbb{P}\left(\sqrt{|I|} - (1+t)\sqrt{n} \leq \nu_j^I \leq \sqrt{|I|} + (1+t)\sqrt{n}\right) \geq 1 - 2e^{-nt^2/2}.$$

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 (36)–(37), we have for some $z \in \mathbb{C}^{n-1}, \|z\| = 1$ that

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

By Proposition 4.3, we obtain the desired bound (12). The success probability is at least the expression (34) minus $4e^{-nt^2/2}$.

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

$$(38) \quad \tau_* = -2 \ln(1 - |I|/N)$$

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

Proposition 4.3 calls for study of order statistics for i.i.d. chi-squared random variables.

Define

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

and

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

Next we show that \hat{I} is a good approximation of I .

PROPOSITION 4.4. *Let τ_* be given by (38) and $\{\tau_1 \leq \tau_2 \leq \dots \leq \tau_N\}$ the sorted sequence of $\{b(1)^2, \dots, b(N)^2\}$ in magnitude.*

(i) *For any $\delta > 0$, we have*

$$(39) \quad \tau_{|I|} \leq \tau_* + \delta$$

with probability at least

$$(40) \quad 1 - \exp\left(-\frac{N}{2} \delta^2 e^{-\delta} |1 - |I|/N|^2\right).$$

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

$$(41) \quad |\hat{I}| \geq |I|(1 - \epsilon)$$

or, equivalently,

$$(42) \quad \tau_{\lfloor |I|(1-\epsilon) \rfloor} \leq \tau_*$$

with probability at least

$$(43) \quad 1 - 2 \exp(-4\epsilon^2 |1 - |I|/N|^2 |I|^2 / N).$$

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

$$(44) \quad |F(\tau + \epsilon) - F(\tau)| \geq \epsilon/2 \exp(-(\tau + \epsilon)/2).$$

For $\delta > 0$, let

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

which by (44) satisfies

$$(45) \quad \zeta \geq \frac{\delta}{2} \exp\left(-\frac{1}{2}(\tau_* + \delta)\right).$$

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

$$(46) \quad \begin{aligned} \mathbb{P}(\tau_{|I|} > \tau_* + \delta) &= \mathbb{P}\left(\sum_{i=1}^N w_i > N - |I|\right) \\ &= \mathbb{P}\left(N^{-1} \sum_{i=1}^N w_i - \mathbb{E}[w_i] > 1 - |I|/N - \mathbb{E}[w_i]\right) \\ &= \mathbb{P}\left(N^{-1} \sum_{i=1}^N w_i - \mathbb{E}[w_i] > \zeta\right) \\ &\leq \exp(-2N\zeta^2). \end{aligned}$$

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

$$(47) \quad \tau_{|I|} \leq \tau_* + \delta$$

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 (45).

(ii) Consider the following replacements in the preceding argument:

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

Then (46) becomes

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

That is,

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

holds with probability at least

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

The next proposition says that the two ratios $\|b_I\|^2/|I|$ and $\|\hat{b}\|^2/|\hat{I}|$ are close to each other.

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

$$(48) \quad \frac{\|b_I\|^2}{|I|} \leq \frac{\|\hat{b}\|^2}{|\hat{I}|} + \epsilon(\tau_* + \delta)$$

with probability at least

$$(49) \quad 1 - 2 \exp\left(-\frac{1}{2}\delta^2 e^{-\delta} |1 - |I|/N|^2 N\right) - 2 \exp(-2\epsilon^2 |1 - |I|/N|^2 |I|^2 / N).$$

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 (43).

By Proposition 4.4(i), $\tau_{|I|} \leq \tau_* + \delta$ with probability at least given by (40). □

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

$$(50) \quad Z_i := \frac{N^2}{|I|^2} [b(i)^2 \chi_{\tau_*} - \mathbb{E}[b(i)^2 \chi_{\tau_*}]],$$

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

$$(51) \quad \mathbb{E}(b(j)^2 \chi_{\tau_*}) = \int_0^{\tau_*} 2^{-1} z \exp(-z/2) dz = 2 - (\tau_* + 2) \exp(-\tau_*/2) \leq 2|I|^2/N^2,$$

and hence

$$(52) \quad -2 \leq Z_i \leq \sup \left\{ \frac{N^2}{|I|^2} b(i)^2 \chi_{\tau_*} \right\} = \frac{N^2}{|I|^2} \tau_*.$$

Now recall the Bernstein inequality.

PROPOSITION 4.6 (see [11]). *Let Z_1, \dots, Z_N be i.i.d. centered subexponential random variables. Then for every $t \geq 0$ we have*

$$(53) \quad \mathbb{P} \left\{ N^{-1} \left| \sum_{i=1}^N Z_i \right| \geq t \right\} \leq 2 \exp \left\{ -c \min(Nt^2/K^2, Nt/K) \right\},$$

where c is an absolute constant and

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

Remark 4.1. We have the following explicit estimates for the constant K :

$$(54) \quad \begin{aligned} K &\leq \frac{2N^2}{|I|^2} \sup_{p \geq 1} p^{-1} (\mathbb{E}|b(i)^2 \chi_{\tau_*}|^p)^{1/p} \\ &\leq \frac{2N^2}{|I|^2} \tau_* \sup_{p \geq 1} p^{-1} (\mathbb{E}\chi_{\tau_*})^{1/p} \\ &\leq \frac{2N^2}{|I|^2} \tau_* \sup_{p \geq 1} p^{-1} (1 - e^{-\tau_*/2})^{1/p}. \end{aligned}$$

The maximum of the right-hand side of (54) 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

$$(55) \quad K \leq \frac{4N}{e|I|} \left(\ln \frac{N}{|I|} \right)^{-1}, \quad \sigma = |I|/N \ll 1.$$

On the other hand, upon substituting the asymptotic bound (55) in the expression

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

on the right side of the probability bound (53), we have

$$(56) \quad Q \leq 2 \exp \left\{ -c \min \left[\frac{e^2 t^2}{16} (\ln \sigma^{-1})^2 |I|^2/N, \frac{et}{4} |I| \ln \sigma^{-1} \right] \right\}, \quad \sigma \ll 1.$$

Now we estimate $\|\hat{b}\|^2/|\hat{I}|$ (Proposition 4.5) by the Bernstein inequality with (56). 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 (41) and (51), we also have

$$(57) \quad \begin{aligned} \frac{\|\hat{b}\|^2}{|\hat{I}|} &\leq \mathbb{E}(b(i)^2\chi_{\tau_*}) \frac{N}{|\hat{I}|} + t \frac{|I|^2}{|\hat{I}|N} \\ &\leq \left(\mathbb{E}(b(i)^2\chi_{\tau_*}) \frac{N^2}{|I|^2} + t \right) \frac{|I|}{N} \\ &\leq \frac{2+t}{1-\epsilon} \cdot \frac{|I|}{N}. \end{aligned}$$

By Proposition 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 (34)–(35), which together with (57) and (38) completes the proof of Proposition 4.3.

5. Numerical experiments. In this section we test numerically the null vector method and compare the performance with 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 :

- White noise:

$$(58) \quad x_0(t) = \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} (p_k + iq_k) e^{i2\pi k(t-1)/n}, \quad t = 0, 1, \dots, n-1.$$

- Low-pass signal:

$$(59) \quad x_0(t) = \sum_{k=-\frac{n}{8}}^{\frac{n}{8}-1} (p_k + iq_k) e^{i2\pi k(t-1)/n}, \quad t = 0, 1, \dots, n-1.$$

- Randomly phased phantom (RPP): x_0 is the vectorized version of the phantom (Figure 2(a)) with phase at each pixel being i.i.d. over $[0, 2\pi)$.

To make a fair comparison with the spectral vector method, we normalize the null vector as computed by Algorithm 1 so that $\|x_{\text{null}}\| = \|x_0\|$, as assumed in Algorithm 2.

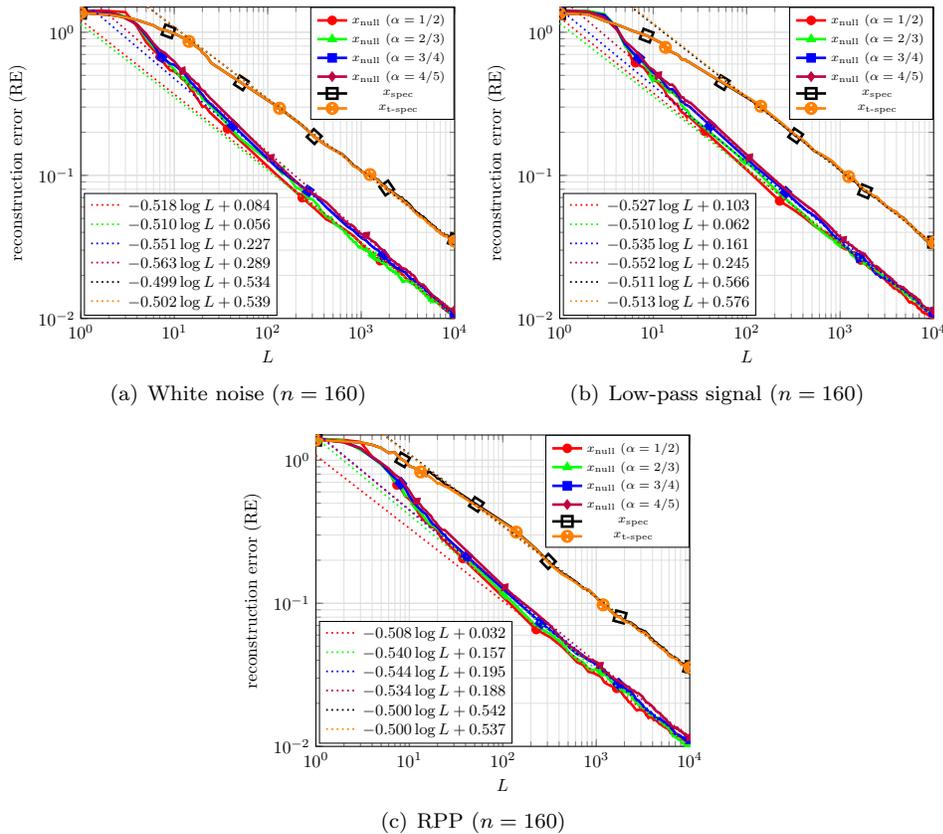


FIG. 1. Log-log plot of relative error (RE) of the null vector method with $\alpha = 1/2$ (red), $\alpha = 2/3$ (green), $\alpha = 3/4$ (blue), $\alpha = 4/5$ (purple), the spectral (black), and the truncated spectral method (yellow) vs. $L \leq 10^4$. (Color available online.)

5.1. Convergence test for the Gaussian measurements. First we test the scaling behavior (19)–(20) as predicted by Corollary 2.2.

Figure 1 is the log-log plot of the relative error (RE)

$$(60) \quad \text{RE} := \|x_0\|^{-2} \|x_0 x_0^* - \hat{x} \hat{x}^*\|_2,$$

where $\hat{x} = x_{\text{null}}$, x_{spec} or $x_{\text{t-spec}}$ as L varies. 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 slightly more negative than -0.5 (and that for the spectral methods) for all α and signals tested (see the legend and caption of Figure 1). We set, according to [1], the thresholding parameter $\tau = 3$ in (24).

5.2. Initialization with randomly coded diffraction patterns. Next we test the performance of the null vector and the spectral vector methods for randomly coded diffraction patterns (RCDPs).

Let the mask function $\{\mu(\mathbf{k})\}$ be continuously and independently distributed. Each RCDP is the squared modulus of the 2-dim discrete-space Fourier transform

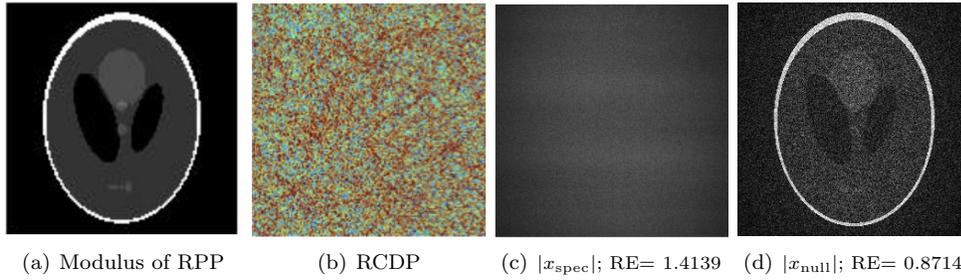


FIG. 2. Noiseless estimation by (d) Algorithm 1 with $|I| = \lceil \sqrt{nN} \rceil$ and (c) Algorithm 2 for (a) 256×256 RPP. (b) is an example of RCDP, where the color scheme represents the intensity of diffraction pattern.

indexed by \mathbf{w} ,

$$(61) \quad \sum_{\mathbf{k}} x_0(\mathbf{k}) \mu(\mathbf{k}) e^{-i2\pi \mathbf{k} \cdot \mathbf{w}}, \quad \mathbf{k} \in [1, \sqrt{n}]^2 \subset \mathbb{Z}^2,$$

with the sampling set

$$(62) \quad \mathbf{w} \in \mathcal{S} = \left\{ (w_1, w_2) \in [0, 1]^2 \mid w_j = 0, \frac{1}{2\sqrt{n}+1}, \frac{2}{2\sqrt{n}+1}, \dots, \frac{2\sqrt{n}}{2\sqrt{n}+1} \right\}.$$

As first proved in [5], the phase retrieval solution is unique almost surely, up to a constant phase factor, with two independently and continuously distributed RCDPs.

In the following simulations, we use two independent masks, $\{\mu_1(\mathbf{k})\}, \{\mu_2(\mathbf{k})\}$, each of which is generated by i.i.d. uniform random variables over the unit circle on the complex plane. Figure 2(b) is an example of such an RCDP with RPP (a) as the object.

Let Φ denote the 2-dimensional discrete-space Fourier transform and define the measurement matrix as

$$(63) \quad A^* = a \begin{bmatrix} \Phi \text{diag}\{\mu_1\} \\ \Phi \text{diag}\{\mu_2\} \end{bmatrix}.$$

With a proper normalization constant a , A^* is isometric. In view of (62) and (63), $|\mathcal{S}| = (2\sqrt{n} - 1)^2$, and hence $N = 2|\mathcal{S}| \approx 8n$ (equivalently $L \approx 8$) for large n .

Figure 2 shows the noiseless reconstruction by the spectral vector and the null vector with $|I| = \lceil \sqrt{nN} \rceil$. The spectral vector (Figure 2(c)) does not yield a meaningful estimate, while the null vector has a decent visual quality (Figure 2(d)). We are unable to improve the result in Figure 2(c) by truncation (24) with any $\tau \in [0.5, 3]$.

To demonstrate the noise stability of the null vector method, we add i.i.d. complex Gaussian noise to A^*x_0 to obtain the noisy data b . Figure 3 shows the results of Algorithm 1 with $|I| = \lceil \sqrt{nN} \rceil$ for 5 ~ 20% noise-to-signal ratio (NSR) defined as

$$\text{NSR} = \frac{\|b - |A^*x_0|\|}{\|A^*x_0\|}.$$

The noise stability of the null vector manifests in the small ratio (less than 1.5) of the increase in RE to the increase in NSR in Figure 3.

As remarked in section 1, the null vector estimation can be further improved by incorporating the full data b as in (5). The reduction in RE by step (5) is as follows: 0.8714 \rightarrow 0.6893 (NSR 0%), 0.8780 \rightarrow 0.7115 (NSR 5%), 0.9173 \rightarrow 0.7520 (NSR 10%), 0.9774 \rightarrow 0.8314 (NSR 15%), and 1.0797 \rightarrow 0.9726 (NSR 20%).

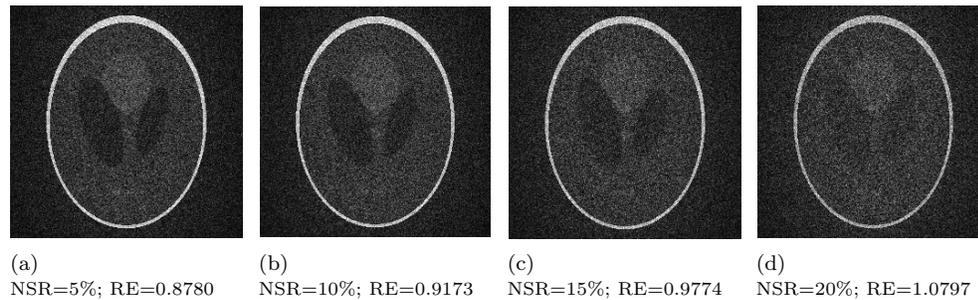


FIG. 3. Noisy estimation by Algorithm 1 with $|I| = \lceil \sqrt{nN} \rceil$ at various NSRs.

6. Conclusion. We have proposed a simple and efficient estimation for phase retrieval and given a performance guarantee for the case of random Gaussian measurements (Theorem 2.1 and Corollary 2.2).

Our analysis predicts a scaling behavior for the error bound consistent with our numerical results, which suggest a universal power law in the limit of large oversampling ratios as long as the choice of $|I|$ lies in the admissible range prescribed in (7).

In the realistic case of coded diffraction patterns of small oversampling ratio, the null vector method continues to perform well and is stable to measurement noise (Figure 3). The effectiveness of the null vector as initialization for nonconvex iterative algorithms such as Alternating Projections and the Wirtinger flow has been demonstrated in [4].

REFERENCES

- [1] E. J. CANDÈS AND Y. CHEN, *Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems*, preprint, <https://arxiv.org/abs/1505.05114>, 2015.
- [2] E. J. CANDÈS, X. LI, AND M. SOLTANOLKOTABI, *Phase retrieval via Wirtinger flow: theory and algorithms*, IEEE Trans. Inform. Theory, 61 (2015), pp. 1985–2007.
- [3] P. CHEN AND A. FANNJIANG, *Phase retrieval with a single mask by the Douglas-Rachford algorithm*, Appl. Comput. Harmonic Anal., 2016, <https://doi.org/10.1016/j.acha.2016.07.003>.
- [4] P. CHEN, A. FANNJIANG, AND G. LIU, *Phase retrieval with one or two coded diffraction patterns by alternating projection with the null initialization*, J. Fourier Anal. Appl., 2017, <https://doi.org/10.1007/s00041-017-9536-8>.
- [5] A. FANNJIANG, *Absolute uniqueness of phase retrieval with random illumination*, Inverse Problems, 28 (2012), 075008.
- [6] J. R. FIENUP, *Phase retrieval algorithms: A comparison*, Appl. Opt., 21 (1982), pp. 2758–2769.
- [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 (1972), pp. 237–246.
- [8] P. NETRAPALLI, P. JAIN, AND S. SANGHAVI, *Phase retrieval using alternating minimization*, IEEE Trans. Signal Process., 63 (2015), pp. 4814–4826.
- [9] Y. SHECHTMAN, Y. C. ELДАР, O. COHEN, H. N. CHAPMAN, M. JIANWEI, AND M. SEGEV, *Phase retrieval with application to optical imaging: A contemporary overview*, IEEE Mag. Signal Process., 32 (2015), pp. 87–109.
- [10] J. A. TROPP, *An Introduction to Matrix Concentration Inequalities*, preprint, <https://arxiv.org/abs/1501.01571>, 2015.
- [11] R. VERSHYNIN, *Introduction to the non-asymptotic analysis of random matrices*, in Compressed Sensing, Cambridge University Press, Cambridge, UK, 2012, pp. 210–268.