Article

A Composite Initialization Method for Phase Retrieval

Qi Luo [1] , **Shijian Lin** [2] **, Hongxia Wang** [1,]*

1   Department of Mathematics, National University of Defense Technology, Changsha 410000, China;
     luoqi10@nudt.edu.cn
2   College of Meteorology and Oceanography, National University of Defense Technology,
     Changsha 410000, China; linshijian10@nudt.edu.cn
*   Correspondence: wanghongxia@nudt.edu.cn

Abstract: Phase retrieval is a classical inverse problem with respect to recovering a signal from a system of phaseless constraints. Many recently proposed methods for phase retrieval such as PhaseMax and gradient-descent algorithms enjoy benign theoretical guarantees on the condition that an elaborate estimate of true solution is provided. Current initialization methods do not perform well when number of measurements are low, which deteriorates the success rate of current phase retrieval methods. We propose a new initialization method that can obtain an estimate of the original signal with uniformly higher accuracy which combines the advantages of the null vector method and maximal correlation method. The constructed spectral matrix for the proposed initialization method has a simple and symmetrical form. A lower error bound is proved theoretically as well as verified numerically.

Keywords: phase retrieval; spectral method; non-convex reconstruction

1. Introduction

Phase retrieval (PR) has many applications in science and engineering, which include X-ray crystallography [1], molecular imaging [2], biological imaging [3], and astronomy [4]. Mathematically, phase retrieval is a problem about finding a signal $x \in \mathbb{R}^n$ or $\mathbb{C}^n$ satisfying the following phaseless constraints:

$$b_i = |\langle a_i, x \rangle| + \epsilon_i, \quad i = 1, \cdots, m. \tag{1}$$

where $b_i \in \mathbb{R}$ is the observed measurement, $a_i \in \mathbb{R}^n$ or $\mathbb{C}^n$ is the measuring vector, and ϵ_i denotes noise. In optics and quantum physics, a_i is related to the Fourier transform or the Fractional Fourier transform [5]. To remove the theoretical barrier, recent works also focus on the generic measurement vectors, namely $a_i \in \mathbb{R}$ independently sampled from $\mathcal{N}(0, I_{n \times n})$.

Current methods for solving phase retrieval can be generally divided into two groups: convex and non-convex approaches. Convex methods, e.g., PhaseLift [6] and PhaseCut [7], utilize a semidefinite relaxation to transfer the original problem into a convex one. Nevertheless, the heavy computational cost of these methods hinders its application to practical scenarios. On the other hand, non-convex methods including the Gerchberg–Saxton (GS) [8,9], the (truncated) Wirtinger flow (WF/TWF) [10,11], and the (truncated) amplitude flow (AF/TAF) have a significant advantage in lower computational cost and sampling complexity (i.e., the minimal required m/n to achieve successful recovery) [12]. More recently, two variants of AF using the reweighting [13] and the smoothing technique [14,15] have been proposed to further lower sampling complexity. However, for the mentioned nonconvex methods, the rate of successful reconstruction relies upon elaborate initialization. Moreover, an initialization point nearby the true solution is also a necessity for convergency property in theoretical analysis.

1.1. Prior Art

Current initialization methods for PR problem include spectral method [16], the orthogonality promoting method [12], null vector method [17], and the (weighted) maximal correlation method [13]. These methods first construct a *spectral matrix* in the form of

$$\boldsymbol{M} = \sum_{i=1}^{m} w_i a_i a_i^{\mathrm{H}},$$

where the weight w_i is a positive real number. The estimate is then given as the (scaled) eigenvector corresponding to the largest or smallest eigenvalue of $\boldsymbol{M}$. The spectral method and the (weighted) maximal correlation method fall into one category as they seek the leading vector of $\boldsymbol{M}$ as the estimate. We categorize these methods as *correlation-focused* methods, since their spectral matrices endow more weight to the measuring vectors more correlated to the original signal. Conversely, the null vector and orthogonality promoting method can be regarded as *orthogonality-focused* methods. For the vector a_i with less correlation to the original signal, the corresponding weight w_i is larger in the spectral matrix.

These two types of initialization methods each have their own merits and demerits. The correlation-focused methods perform better when the oversalting rate m/n is low. While the situation is contrary as m/n is large enough, the orthogonality-focused method provides a more accurate estimate.

1.2. This Work

In this paper, we propose a method combining the advantages of the above two types of initialization methods. Our intuition is to construct a composite spectral matrix, which is a weighted sum of spectral matrices of correlation-focused methods and orthogonality-focused methods. Hence, the new method is termed as the composite initialization method.

1.3. Article Organization and Notations

This paper adopts the following notations. The bold-font lower-case letter denotes a column vector, e.g., $\boldsymbol{b}$, $\boldsymbol{z}$. Bold capital letters such as $\boldsymbol{A}$ denote matrices. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ stand for the transpose and the conjugate transpose, respectively. Calligraphic letters such as $\mathcal{I}$ denote a index set, and $|\mathcal{I}|$ denotes the number of elements of $\mathcal{I}$. $\{a_i\}_{i=1}^{m}$ is a simple notation of a set of elements in the form of $\{a_1, \ldots, a_m\}$. $\|\boldsymbol{x}\|_2$ denotes the ℓ_2-norm of $\boldsymbol{x}$. $\|\boldsymbol{x}\|$ means $\|\boldsymbol{x}\|_2$ for the sake of simplicity, if not specially specified.

The reminder of this article is organized as follows. The algorithm is given in Section 2. Section 3 provides the theoretical error analysis about the proposed initialization method with some relevant technique error bounds being placed in the Appendix. Section 4 illustrates the numerical performance and makes a comparison of the proposed method with other initialization methods.

2. The Formulation of the Composite Initialization Method

This section provides the formulation of the proposed method. Without loss of generality, we assume that $b_1 \leq b_2 \leq \cdots \leq b_m$. Measuring vectors $\{a_i\}_{i=1}^{m}$ are assumed to be independently sampled centered Gaussian variables with identity covariance for the convenience of theoretical analysis. Let $\boldsymbol{x}_0$ be the true solution of (1). For concreteness, we focus on the real-valued Gaussian model. We further assume that $\|\boldsymbol{x}_0\| = 1$ since this investigation focuses on the estimation of the original signal. In practice, $\|\boldsymbol{x}_0\|$ can be estimated by $\|\boldsymbol{x}_0\|^2 \approx \frac{1}{m} \sum_{1}^{m} b_i^2$. To begin with, we describe the null vector and the maximal correlation method which form the basis of our algorithm.

The null vector method first picks out a subset of vectors $\{a_i\}_{i \in \mathcal{I}_1}$, where $\mathcal{I}_1 \subset \{1, 2, \cdots, m\}$. Then, the null vector method approximates the original signal $\boldsymbol{x}_0$ by a vector that is most orthogonal to this subset. Since $\{b_i\}_{i=1}^{m}$ is in the ascending order, $\mathcal{I}_1$ can be directly written as $\{1, 2, \cdots, T_1\}$, where T_1 is the cardinality of $\mathcal{I}_1$. The intuition of the null vector method is as simple as follows. Since $\sum_{i \in \mathcal{I}_1} \boldsymbol{x}^{\mathrm{H}} a_i a_i^{\mathrm{H}} \boldsymbol{x}$ takes a very small value

$\sum_{i \in \mathcal{I}_1} b_i^2$ when $x = x_0$, one can construct the following minimizing problem to estimate x_0 to coincide this property.

$$\hat{x} = \arg \min_{\|z\|=1} z^{\mathrm{H}} \left(\frac{1}{T_1} \sum_{i=1}^{T_1} a_i a_i^{\mathrm{H}} \right) z. \qquad (2)$$

Solving (2) is equivalent to finding the smallest eigenvalue and the corresponding eigenvector of $M_1 := \frac{1}{T} \sum_{i=1}^{T_1} a_i a_i^{\mathrm{H}}$. The so-called orthogonality promoting method proposed by Wang is an approximately equivalent method that transforms the minimization problem (2) into a maximization problem.

The maximal correlation method is based on the opposite intuition, which picks out a subset of vectors $\{a_i\}_{i \in \mathcal{I}_2}$ corresponding to the largest T_2 elements in $\{b_i\}_{i=1}^m$. As $\{b_i\}_{i=1}^m$ is in the ascending order, $\mathcal{I}_2$ can also be written directly as $\{m - T_2 + 1, \cdots, m\}$. The core idea of maximal correlation method is searching for the vector that is most correlated with $\{a_i\}_{i \in \mathcal{I}_2}$. This is achieved by solving the following maximizing problem.

$$\hat{x} = \arg \max_{\|z\|=1} \frac{1}{T_2} \sum_{i=m-T_2+1}^{m} |\langle a_i, z \rangle|^2 = \arg \max_{\|z\|=1} z^{\mathrm{H}} \left(\frac{1}{T_2} \sum_{i=m-T_2+1}^{m} a_i a_i^{\mathrm{H}} \right) z. \qquad (3)$$

From (2) and (3), one can observe that these two methods only utilize a subset of vectors that are either most correlated or most orthogonal to the original signal. To make use of as much as possible information, we try to solve a composite problem combining (2) and (3) together.

$$\hat{x} = \arg \min_{\|z\|=1} \left\{ z^{\mathrm{H}} \left(\frac{1}{T_1} \sum_{i=1}^{T_1} a_i a_i^{\mathrm{H}} \right) z - \alpha z^{\mathrm{H}} \left(\frac{1}{T_2} \sum_{i=m-T_2+1}^{m} a_i a_i^{\mathrm{H}} \right) z \right\}. \qquad (4)$$

We can observe that (4) has a symmetrical structure and can be interpreted as a modified version of (2), which adds the objection of (3) as a penalization term. The solution of (4) will be a more accurate estimate than that of (3) since it utilize the information from (2).

Let $H_1 = \sum_{i=1}^{T_1} b_i^2 / T_1$, $H_2 = \sum_{i=m-T_2+1}^{T_2} b_i^2 / T_2$. We set the parameter as $\alpha = H_1 / H_2$ in (4). The true signal x is roughly the solution of (2) and (3); thus, it is also the approximated solution of (4). In next section, we will analyze the error of our proposed method.

The algorithm is presented in Algorithm 1. $2\alpha I$ is added to ensure the positivity of M. The x'_{k+1} in Step 4 is obtained by solving the following system of linear equations:

$$Mx = x_k. \qquad (5)$$

which can be solved by the conjugate descent method since M is a positive and Hermitian by using x_{k-1} as initializer.

Since x_0 and $e^{-i\phi} x_0$ are indistinguishable for the phase retrieval, we evaluate the estimate using the following metric.

$$\mathrm{dist}(\tilde{x}, x_0) = \min_{\phi \in [0, 2\pi)} \left\| e^{-i\phi} \tilde{x} - x_0 \right\|. \qquad (6)$$

The other initialization methods, e.g., the spectral method, the truncated spectral method, and the (weighted) maximal correlation method, are all based on the power iteration. The spectral method is for finding the leading eigenvector of matrix $\sum_{i=1}^m b_i^2 a_i a_i^T$, which is also realized by the power approach. Chen proposed the truncated spectral method to improve the performance of the spectral method, which constructed another matrix via the threshold parameter τ [17]. The corresponding matrix for iteration is $\sum_{i=1}^m \mathbf{1}_\tau(b_i) b_i^2 a_i a_i^{\mathrm{H}}$, where $\mathbf{1}_\tau$ is defined by the following.

$$\mathbf{1}_\tau(b_i) = \begin{cases} 1, & \text{if } b_i \le \tau \|b\|_2 / \sqrt{m}; \\ 0, & \text{else.} \end{cases} \tag{7}$$

The numerical performance of these initialization methods will be compared later.

Algorithm 1: The composite initialization method.

Input: Increasingly arranged $\{b_i\}_{i=1}^m$ and corresponding $\{a_i\}_{i=1}^m$, truncated value T_1, T_2, threshold value ϵ, a random initialization x_1.

1: $M_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} a_i a_i^{\mathrm{H}}$, $M_2 = \frac{1}{T_2}\sum_{i=T_2+1}^m a_i a_i^{\mathrm{H}}$, $\alpha = (\frac{1}{T_1}\sum_{i=1}^{T_1} b_i^2)/(\frac{1}{T_2}\sum_{i=m-T_2+1}^m b_i^2)$
2: $M = M_1 - \alpha M_2 + 2\alpha I$
3: **for** $k = 1, 2, \cdots$ **do**
4: $\quad x'_{k+1} \leftarrow M^{-1} x_k$
5: $\quad x_{k+1} \leftarrow x'_k / \|x'_{k+1}\|$
6: $\quad$ **if** $\|x_{k+1} - x_k\| \le \epsilon$, **break**
7: **end for**

Output: $x_{\mathrm{OPMP}} = x_{k+1}/\|x_{k+1}\|$.

---

3. Theoretical Analysis

In this section, we will present the error estimate of the proposed method under Gaussian assumptions in the real case.

Since the measuring vectors are assumed to be i.i.d. Gaussians, we can assume that $x_0 = e_1$ without loss of generality, where $e_1 = [1, 0, \cdots, 0]^{\mathrm{T}}$. Otherwise if $x_0 \ne e_1$, there exists an orthogonal matrix, P, satisfying $Px_0 = e_1$. Denote $A = [a_1, a_2, \cdots, a_m]$, then $A^{\mathrm{T}}x = A^{\mathrm{T}}P^T Px$. Let $z = Px$, then the original problem (1) is identical to $(PA)^T z = b$ because PA is composed by Gaussian vectors due to the invariance of Gaussian under the orthogonal transform.

The spectral matrix for orthogonality promoting method is the following.

$$M_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} a_i a_i^{\mathrm{T}}. \tag{8}$$

By denoting $d_i = (a_{i2}, a_{i3}, \cdots, a_{in})^{\mathrm{T}}$, then M can be written as follows:

$$M_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} \begin{bmatrix} a_{i1}^2 & a_{i1}d_i^{\mathrm{T}} \\ a_{i1}d_i & d_i d_i^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} H_1 & E_1^{\mathrm{T}} \\ E_1 & G_1 \end{bmatrix} \tag{9}$$

where the following is the case.

$$H_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} a_{i1}^2, \quad E_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} a_{i1}d_i, \quad G_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} d_i d_i^{\mathrm{T}}. \tag{10}$$

The matrix for maximal correlation method is the following:

$$M_2 = \frac{1}{T_2}\sum_{i=m-T_2+1}^m \begin{bmatrix} a_{i1}^2 & a_{i1}d_i^{\mathrm{T}} \\ a_{i1}d_i & d_i d_i^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} H_2 & E_2^{\mathrm{T}} \\ E_2 & G_2 \end{bmatrix} \tag{11}$$

where the following is the case.

$$H_2 = \frac{1}{T_2}\sum_{i=m-T_2+1}^m a_{i1}^2, \quad E_2 = \frac{1}{T_2}\sum_{i=m-T_2+1}^m a_{i1}d_i, \quad G_2 = \frac{1}{T_2}\sum_{i=m-T_2+1}^m d_i d_i^{\mathrm{T}}. \tag{12}$$

The constructed matrix for estimation is the following.

$$M = M_1 - \alpha M_2, \quad \alpha = H_1 / H_2. \tag{13}$$

To proceed, we notice the following basic facts:

1. The orthogonality promoting method is for finding the eigenvector of minimum eigenvalue of M.

$$M = \begin{bmatrix} 0 & E_1^{\mathrm{T}} - \alpha E_2^{\mathrm{T}} \\ E_1 - \alpha E_2 & G_1 - \alpha G_2 \end{bmatrix} := \begin{bmatrix} 0 & E^{\mathrm{T}} \\ E & G \end{bmatrix}. \tag{14}$$

2. In the ideal case that E is a zero matrix, M degenerates to the following.

$$\widetilde{M} = \begin{bmatrix} 0 & 0 \\ 0 & G \end{bmatrix}.$$

If we can also ensure that $G = G_1 - \alpha G_2$ is positively definite, the eigenvector corresponding to the smallest eigenvalue is exactly the true signal.

These two facts inspire us to estimate the error by computing the eigenvector of matrix $\widetilde{M}$ after adding perturbation $E = E_1 - \alpha E_2$. Thus, the outline of theoretical analysis is concluded as follows:

1. Estimate $\Delta\lambda = \lambda - \tilde{\lambda}$, where λ and $\tilde{\lambda}$ are the minimum eigenvalue of M and $\widetilde{M}$, respectively;
2. With $\Delta\lambda$, we can then compute the perturbation of corresponding eigenvector, $\|v - \tilde{v}\|$, which is the exact error of our algorithm.

Specifically, we have the following roadmap of theoretical analysis. Section 3.1 presents bound results for each component of the spectral matrix M. Using the results in Section 3.1, the bounds of $\Delta\lambda$ can then be easily obtained in Section 3.2. The relationship between perturbation of eigenvalues and eigenvectors is presented in Section 3.3, which finally induces the error estimation of our algorithm formally in Section 3.4.

3.1. Analysis of Each Component of the Spectral Matrix

In this part, we will provide the bounds for the variables involved of matrix M, which are basic ingredients for estimating perturbation of eigenvalue and eigenvector, namely $\Delta\lambda$ and $\|v - \tilde{v}\|$. In particular, the upper bound for H_1, lower bound for H_2, and the norm of E will be analysed.

3.1.1. Upper Bound of $\alpha = H_1 / H_2$

Finding the upper bound of α actually consists in finding the upper bound of H_1 and the lower bound of H_2. First, we have the upper bound of H_1 under statistical meaning in the follow lemma.

Lemma 1. *We have the following:*

$$\Pr(H_1 \leq \frac{p_1^2}{4\rho_0^2}) \geq 1 - \exp\left(-\frac{mp_1^2}{50}\right) \tag{15}$$

where $p_1 = T_1 / m$.

The proof of Lemma 1 is placed in the Appendixes A and B. As for the lower bound of H_2, we borrow a result from Wang [13], Lemma 3.

Lemma 2. *The following holds with probability exceeding* $1 - e^{-c_0 m}$:

$$H_2 \geq 0.99(1 + \log(m/T_2)), \tag{16}$$

provided that $T_2 \geq c_1 n$, $m \geq c_2 T_2$ *for some absolute constants* c_0, c_1, *and* c_2.

Then, we obtained the required upper bound of H_1/H_2.

Lemma 3. *Under the set-up of Lemmas 1 and 2, we have the following:*

$$\alpha = H_1/H_2 \leq \frac{p_1^2}{3.96\rho_0^2(1 - \log p_2)} \tag{17}$$

with probability of at least

$$1 - \exp\left(-\frac{mp_1^2}{50}\right) - \exp(-c_0 m), \tag{18}$$

where $\rho_0 = 1/\sqrt{2\pi}$, p_1, *and* p_2 *denote* T_1/m *and* T_2/m, *respectively.*

3.1.2. Lower Bound of the Smallest Eigenvalue of **G**

Since **G** is a linear combination of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$, the lower bound of the smallest eigenvalue of **G** can be estimated from the bounds of eigenvalues of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$.

According to (10) and (12), we rewrite $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ as the following:

$$\boldsymbol{G}_1 = \frac{1}{T_1}\boldsymbol{D}_1^{\mathrm{T}}\boldsymbol{D}_1, \quad \boldsymbol{G}_2 = \frac{1}{T_2}\boldsymbol{D}_2^{\mathrm{T}}\boldsymbol{D}_2, \tag{19}$$

where $\boldsymbol{D}_1 = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{T_1}]^{\mathrm{T}}$ and $\boldsymbol{D}_2 = [\boldsymbol{d}_{m-T_2+1}, \ldots, \boldsymbol{d}_m]^{\mathrm{T}}$. $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are termed Gaussian matrices here since their entries are all sampled from the Gaussian distribution. We notice that $\boldsymbol{D}_1^{\mathrm{T}}\boldsymbol{D}_1$ and $\boldsymbol{D}_2^{\mathrm{T}}\boldsymbol{D}_2$ are Wishart matrices, which can be written as the product of a Gaussian matrix and its adjoint [18]. Moreover, the eigenvalue perturbation of Wishart matrix obeys the following classical result.

Theorem 1 ([19], Corollary 5.35). *Let* **A** *be an* $N_1 \times N_2$ *matrix for which its entries are independent standard normal random variables. Then, for every* $t \geq 0$, *with a probability of at least* $1 - 2\exp\left(-t^2/2\right)$, *the following events hold simultaneously:*

$$\sqrt{N_1} - \sqrt{N_2} - t \leq s_{\min}(\boldsymbol{A}), \tag{20}$$

$$s_{\max}(\boldsymbol{A}) \leq \sqrt{N_1} + \sqrt{N_2} + t, \tag{21}$$

where $s_{\min}(\boldsymbol{A})$ *and* $s_{\max}(\boldsymbol{A})$ *stand for the smallest and the largest singular value of* **A**, *respectively.*

By pplying Theorem 1 to $\boldsymbol{D}_1 \in \mathbb{R}^{T_1 \times (n-1)}$ with the following replacements:

1. $\boldsymbol{D}_1 \to \boldsymbol{A}$,
2. $T_1 \to N_1$, $n - 1 \to N_2$,
3. $\sqrt{n-1}/10 \to t$,

one can see that the following is the case.

$$\Pr\left(s_{\min}(\boldsymbol{D}_1) \geq \sqrt{T_1} - 1.1\sqrt{n-1}\right) \geq 1 - 2\exp\left(-\frac{n-1}{200}\right). \tag{22}$$

In theoretical analysis, we assume that T_1 is large enough such that $\sqrt{T_1} - 1.1\sqrt{n-1} \geq 0$. However, it should be observed that setting $T_1 \geq n$ suffices to output reasonable estimates as shown in the numerical test. Based on this assumption, one then has the following.

$$s_{\min}(G_1) = s_{\min}\left(\frac{1}{T_1}D_1^{\mathsf{T}}D_1\right) = \frac{1}{T_1}s_{\min}^2(D_1)$$

$$\geq \frac{1}{T_1}(\sqrt{T_1} - 1.1\sqrt{n-1})^2 = (1 - 1.1\sqrt{s_1})^2 \tag{23}$$

The above holds with probability exceeding $1 - 2\exp(-\frac{n-1}{200})$, where $s_1 = (n-1)/T_1$. Similarly, we have the following:

$$\Pr\left(s_{\max}(G_2) \leq (1 + 1.1\sqrt{s_2})^2\right) \geq 1 - 2\exp\left(-\frac{n-1}{200}\right) \tag{24}$$

where $s_2 = (n-1)/T_2$.

Hence, we obtain the lower bound for the smallest eigenvalue of G.

$$s_{\min}(G) = s_{\min}(G_1 - \alpha G_2) \geq s_{\min}(G_1) - \alpha s_{\max}(G_2)$$

$$\geq (1 - 1.1\sqrt{s_1})^2 - \alpha(1 + 1.1\sqrt{s_2})^2. \tag{25}$$

A necessary condition for the validness of our method is $s_{\min}(G) \geq 0$, which can be satisfied by choosing proper s_1 and s_2.

3.1.3. The Upper Bound of $\|E\|^2$

Now, let us turn to the estimation of the norm of E. We denote each item of E as ξ_k.

$$\xi_k = \frac{1}{T_1}\sum_{i=1}^{T_1} a_{i1}a_{ik} - \frac{\alpha}{T_2}\sum_{i=m-T_2+1}^{m} a_{i1}a_{ik}. \tag{26}$$

Since a_{ik} obeys normal distribution, $H_1 = \frac{1}{T_1}\sum_{i=1}^{T_1} a_{i1}^2$ and $H_2 = \frac{1}{T_2}\sum_{i=m-T_2+1}^{m} a_{i1}^2$, we can derive the following.

$$\xi_k \sim \mathcal{N}(0, \frac{H_1}{T_1} + \frac{\alpha^2 H_2}{T_2}).$$

Let $\psi_k = \xi_k / \sqrt{\frac{H_1}{T_1} + \frac{\alpha^2 H_2}{T_2}}$, then ψ_k^2 obeys chi-squared distribution, which is a sub-exponential distribution. With ψ_k, we can reform $\|E\|_2^2$ as the following.

$$\|E\|^2 = \left(\frac{H_1(n-1)}{T_1} + \frac{\alpha^2 H_2(n-1)}{T_2}\right)\frac{1}{n-1}\sum_{k=2}^{n}\xi_k^2 = (H_1 s_1 + H_1\alpha s_2)\sum_{k=2}^{n}\frac{\xi_k^2}{n-1}. \tag{27}$$

The expectation of $\|E\|_2^2$ is obviously $H_1(s_1 + s_2\alpha)$.

The variance of $\|E\|_2^2$ is also needed in order obtain the upper bound of $\|E\|_2^2$. To this end, we need to recall the Bernstein-type inequality for sub-exponential random variable.

Theorem 2. *Let* $X_1, \cdots, X_N$ *be i.i.d. centered sub-exponential random variables with sub-exponential norm denoted as K. Then, for every* $t \geq 0$, *we have the following:*

$$\Pr\left(\frac{1}{N}\left|\sum_{i=1}^{N} X_i\right| \geq t\right) \leq 2\exp\left[-cN\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right] \tag{28}$$

where $c > 0$ is an absolute constant. K is the so-called sub-exponential norm defined as the following.

$$K = \sup_{p \geq 1} \frac{1}{p} (\mathbb{E}|X_1|^p)^{1/p}, \tag{29}$$

Define a centered sub-exponential random variable $Z_i = \xi_i - 1$. The sub-exponential norm of Z_i is computed in Appendix C. Appendix C tells that the sub-exponential norm $K = 1$ in our case. By replacing $n - 1 \to N, 0.1 \to t$ and $K \to 1$ in Theorem 2, one can then conclude the following result.

Theorem 3. *Since $K \leq 2$, let $t = 0.1$, then we have the following:*

$$\Pr\left(\frac{1}{n-1}\sum_{i=2}^{n} \xi_i^2 - 1 \geq 0.1\right) \leq \exp\left(-c\frac{n-1}{100}\right) \tag{30}$$

where $c > 0$ is an absolute constant.

Gathering these results, we have the following theorem.

Theorem 4.

$$\|E\|^2 \leq 1.1 H_1(s_1 + s_2\alpha) \tag{31}$$

The above holds with the probability stated below:

$$1 - \exp\left(-c\frac{n-1}{100}\right), \tag{32}$$

with some absolute constant $c > 0$.

3.2. Estimate of $\Delta\lambda$

To make an estimate for $\Delta\lambda$, we first recall a classical result from matrix perturbation theory in [20].

Theorem 5. *Let the following be the case:*

$$M = \begin{bmatrix} H & E^{\mathrm{T}} \\ E & G \end{bmatrix}, \quad \tilde{M} = \begin{bmatrix} H & 0 \\ 0 & G \end{bmatrix} \tag{33}$$

Let the above be Hermitian and set $\eta = \min|\mu - \nu|$ over all $\mu \in \sigma(H)$ and $\nu \in \sigma(G)$, where $\sigma(H)$ stands for the set of eigenvalue of H. Then, the following is the case:

$$\max_{1 \leq j \leq n} \left|\lambda_j^{\uparrow} - \tilde{\lambda}_j^{\uparrow}\right| \leq \frac{2\|E\|_2^2}{\eta + \sqrt{\eta^2 + 4\|E\|_2^2}}. \tag{34}$$

where $\lambda_j^{\uparrow}$ and $\tilde{\lambda}_j^{\uparrow}$ are the j-th smallest one among the eigenvalues of M and $\tilde{M}$, respectively.

To ensure the effectiveness of our algorithm, H_1 should still be the minimum eigenvalue after adding perturbation in our case. Then, η is simply $s_{\min}(G)$ that is bounded by (25). Hence, $\Delta\lambda$ can be estimated by the following.

$$|\Delta\lambda| \leq \max_{1 \leq j \leq n}\left|\lambda_j^{\uparrow} - \tilde{\lambda}_j^{\uparrow}\right| \leq \frac{2\|E\|_2^2}{\eta + \sqrt{\eta^2 + 4\|E\|_2^2}} \leq \frac{2\|E\|_2^2}{2\eta}$$

$$\leq \frac{1.1 H_1(s_1 + \alpha s_2)}{(1 - 1.1\sqrt{s_1})^2 - \alpha(1 - 1.1\sqrt{s_2})^2}. \tag{35}$$

3.3. Computing $\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|$

With the upper bound of $\Delta\lambda$, we can now estimate the perturbation of the eigenvector. The minimum eigenvector of $\boldsymbol{M}$ is computed by solving $(\boldsymbol{M} - (\lambda + \Delta\lambda)\boldsymbol{I}_n)\boldsymbol{v} = 0$:

$$(\boldsymbol{M} - (\lambda + \Delta\lambda)\boldsymbol{I}_m)\boldsymbol{v} = \begin{bmatrix} -\Delta\lambda & \boldsymbol{E}^{\mathrm{T}} \\ \boldsymbol{E} & \boldsymbol{G} - (\lambda + \Delta\lambda)\boldsymbol{I}_{n-1} \end{bmatrix}\boldsymbol{v} = 0, \tag{36}$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix here. Without loss of generality, the solution of (36) can be written in the form of $\boldsymbol{v} = (1, \delta\boldsymbol{v})^{\mathrm{T}}$. Then, we have the following.

$$\delta\boldsymbol{v} = -(\boldsymbol{G} - (\lambda + \Delta\lambda)\boldsymbol{I}_{n-1})^{-1}\boldsymbol{E}. \tag{37}$$

Therefore, $\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|$ can be bounded by the following:

$$\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|^2 \leq \|\delta\boldsymbol{v}\|^2$$

$$\leq \left\|(\boldsymbol{G} - (\lambda + \Delta\lambda)\boldsymbol{I}_{n-1})^{-1}\right\|^2 \|\boldsymbol{E}\|^2 \tag{38}$$

$$\leq \|\boldsymbol{E}\|^2 / \left(\eta - \|\boldsymbol{E}\|^2/\eta\right), \tag{39}$$

where (38) is derived based on (37). The form of (39) would be a bit complicated, as stated as follows.

$$\frac{\|\boldsymbol{E}\|^2}{\eta - \|\boldsymbol{E}\|^2/\eta} \leq \frac{1.1H_1(s_1 + s_2\alpha)\left((1 - 1.1\sqrt{s_1})^2 - \alpha(1 + 1.1\sqrt{s_2})^2\right)}{\left((1 - 1.1\sqrt{s_1})^2 - \alpha(1 + 1.1\sqrt{s_2})^2\right)^2 - 1.1H_1(s_1 + s_2\alpha)}. \tag{40}$$

3.4. Main Result

Notice that $\|\boldsymbol{v} - \tilde{\boldsymbol{v}}\|$ is exactly $\|\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}\|$. Then, it is straightforward to induce our main theorem by combining the above results.

Theorem 6 (Main result). *Consider the problem of estimating arbitrary $\boldsymbol{x} \in \mathbb{R}^n$ from m phaseless measurements* (1) *under the Gaussian assumption. Suppose the output of Algorithm 1 is $\tilde{\boldsymbol{x}}_0$. If $m \geq c_1T_1$, $m \geq c_2T_2$, $T_1 > n$, and $m \geq T_1 + T_2$, then we have the following error estimate for the composite initialization method:*

$$\|\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}\|^2 \leq \frac{RQ}{R - Q}, \tag{41}$$

with a probability of at least the following:

$$P = 1 - \exp\left(-\frac{mp_1^2}{144}e^{\frac{-p_1^2}{2\rho_0^2}}\right) - e^{-\frac{mp_1^2}{50}} - 2e^{-\frac{n-1}{200}} - e^{-c\frac{n-1}{100}} - e^{-c_0 m}, \tag{42}$$

where the following is the case,

$$Q = 1.1p_1^2\left(4\rho_0^2(1 - \log p_2)s_1 + 0.99p_1^2 s_2\right)(1 - \log p_2), \tag{43}$$

$$R = \left[4\rho_0^2(1 - \log p_2)(1 - 1.1\sqrt{s_1})^2 - 0.99p_1^2(1 + 1.1\sqrt{s_2})^2\right]^2, \tag{44}$$

$p_i = T_i/m$, $s_i = (n-1)/T_i$ *for* $i = 1, 2$, *and some absolute constants* $c_0, c_1, c_2, \rho_0 \geq 0$.

Probability P in (42) can be negative when m and n are too small, which shows the limitation of our analysis since m and n need to be large enough so as to render the

probability reasonable. In the extreme case that p_1, p_2, s_1 and s_2 are close to 0, our error estimation expression (41) can be approximated by a simpler form:

$$\|x_0 - \tilde{x}\|^2 \lessapprox \frac{1.1 p_1^2 s_1}{10 \rho_0^2} \tag{45}$$

which is verified by numerical experiments later.

4. Numerical Results

In this section, we test the accuracy of the proposed method and compare it with other methods, including the spectral method, the truncated spectral method, the reweighted maximal correlation method, the null vector method, and the orthogonality method. The sampling vectors a_i and the original signal $x_0 \in \mathbb{R}^{100}$ are independently randomly generated. To eliminate the influence of error brought by estimation of $\|x_0\|$, the original signal x_0 is normalized. We chose $T_1 = n, T_2 = 0.3n$ for numerical experiments for the proposed method. All the following simulation results are averaged over 80 independent Monte Carlo realizations.

Figure 1 provides the RMSE calculated by (6) versus the oversampling rate m/n of the mentioned initialization methods. Obviously, all methods exhibit better performance as m/n increases. In particular, the proposed initialization method outweighs the other methods. When $m/n \geq 10$ roughly, the composite initialization method performs closely as the null vector does, and the convergency behavior coincides with (45). When $m/n \leq 10$, the proposed method does not lose its accuracy as dramatically as the null vector method does. The proposed algorithm provides the most accurate estimate when oversampling rate is below the information limit, i.e., $m/n \approx 2$.
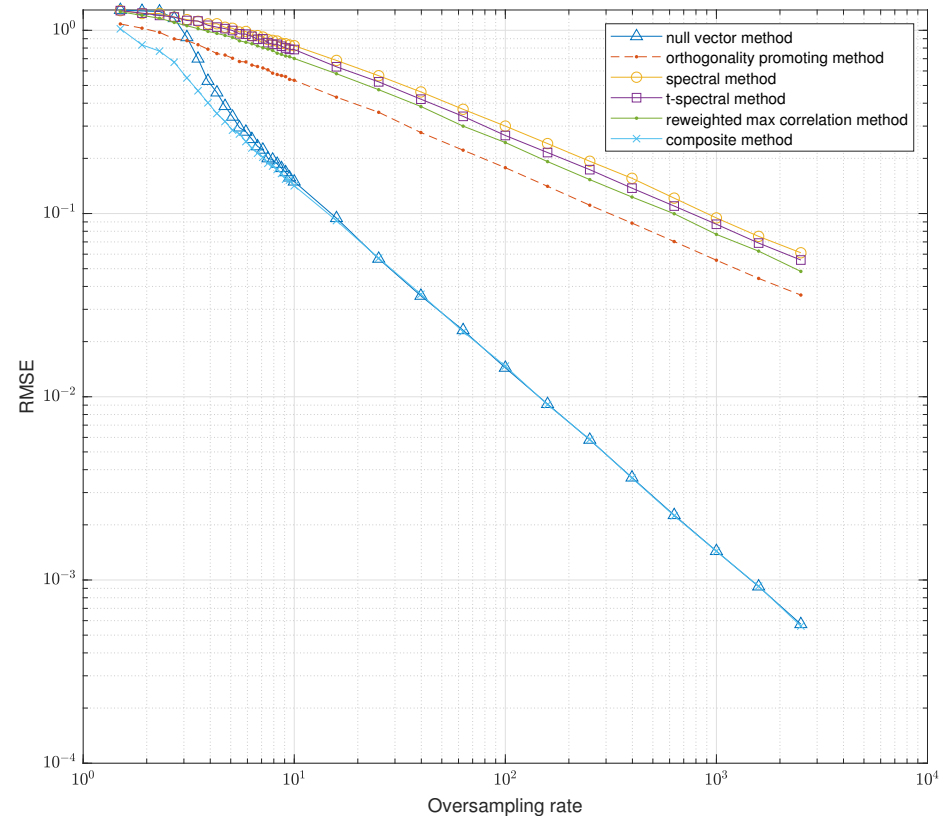

Figure 1. RMSE vs. oversampling rate of several initialization method. $n = 100, T_1 = n, T_2 = 0.3n$ for the composite initialization method, while for the other algorithms, the involved parameters are selected according to related articles.

Figure 2 illustrates the importance of initialization method on refining the success rate of TAF algorithm [12]. In each simulation, each initialization generates an estimate as the initializer for the TAF algorithm. A trial is considered to be successful if TAF algorithm can return a result with RMSE less than 10^{-5}. When m/n is large, all the presented initialization methods ensure almost 100% success rate. However, when m/n approaches 0, the differences in success rates of various methods appear gradually. Therefore, the composite initialization method can help TAF to achieve better success rate compared with the other two methods.

Table 1 illustrates the CPU time needed for the TAF using our method as an initializer compared with other two typical initialization methods. To distinguish the performance more clearly, the length of signal is set as a large number $n = 1000$. The proposed method and null vector method need to solve a linear system of equations at each inverse power computation, which makes it more time-consuming than the power computation of the maximal correlation method. However, the proposed method provides a more accurate initializer, which can help TAF converges faster. Hence, the overall efficiency of our algorithm is not far behind power-type methods as shown in Table 1.
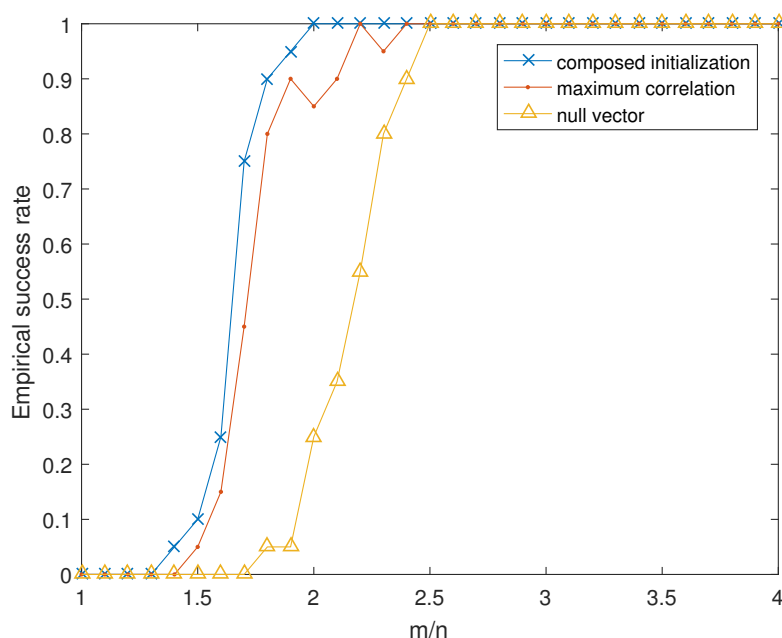

Figure 2. Empirical success rate of different initialization methods versus number of measurements with $n = 100$, with m/n varying form 1 to 4 for TAF.

Table 1. CPU time (s) for TAF using the proposed initialization method compared with other two typical initializer.

	Method	Initialization Stage	TAF Stage	Overall
	Null vector	1.24	1.34	2.58
$m/n = 2$	Maximal correlation	0.47	0.76	1.23
	Proposed	0.72	0.70	1.42
	Null vector	1.12	1.02	2.14
$m/n = 3$	Maximal correlation	0.42	0.49	0.91
	Proposed	0.73	0.47	1.2
	Null vector	0.95	1.02	1.97
$m/n = 5$	Maximal correlation	0.44	0.45	0.89
	Proposed	0.67	0.46	1.13

5. Conclusions

This paper proposes a new initialization method that combines the advantages of two methods constructed from two totally contrary intuitions. Both theoretical analysis and numerical experiments indicate the validness of our method and higher accuracy compared with other methods. Future work will focus on extending our initialization algorithms to the more generalized problem of PR, e.g., the quadratic sensing and the matrix reconstruction problems [21].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Upper Bound of H_1

H_1 is the average of the smallest T_1 elements of the set $\left\{ b_i^2 \right\}_{i=1}^m$. As b_i^2 has been sorted in ascending order, we have the following.

$$H_1 = \frac{1}{T_1} \sum_{i=1}^{T_1} a_{i1}^2 = \frac{1}{T_1} \sum_{i=1}^{T_1} b_i^2.$$

By the Gaussian assumption, $\left| a_i^T x_0 \right|^2 = a_{i1}^2$, obeys the following chi-squared distribution with probability function:

$$\rho(z) = \rho_0 \frac{e^{-z/2}}{\sqrt{z}}, \quad z \geq 0, \tag{A1}$$

where $\rho_0 = \frac{1}{\sqrt{2\pi}}$ is the normalization constant. The cumulative distribution function is the following.

$$F(\tau) := \int_0^\tau \rho_0 \frac{e^{-z/2}}{\sqrt{z}} dz. \tag{A2}$$

Let $p = \frac{T_1}{m}$ and $\tau_* = F^{-1}(p)$. The estimation of H_1 hinges on the value of τ_*. However, $\rho(z)$ is not explicitly integrable; hence we cannot derive an explicit expression of τ_* about p. To this end, we first calculate following bounds of τ_* using some basic inequalities, as the following two lemmas shows. The detailed proof is placed in Appendix B.

Lemma A1 (Upper bound of τ_*)**.** *Let $\tau_* = F^{-1}(p)$, then for $p \in (0, 0.75]$, we have the following.*

$$\tau_* < \frac{4p^2}{9\rho_0^2}. \tag{A3}$$

Lemma A2 (Lower bound of τ_*)**.** *Let $\tau_* = F^{-1}(p)$, then for $p \in (0, 1)$, we have the following.*

$$\tau_* > \frac{p^2}{4\rho_0^2}. \tag{A4}$$

Then, we turn back to the estimation of H_1. Notice that $\{b_i^2\}_{i=1}^{T_1}$ can be approximately regarded as random variables sampled from a bounded chi-squared distribution. Therefore, we first provide an upper bound of the largest element of $\{b_i^2\}_{i=1}^{T_1}$, namely, $b_{T_1}^2$. We prove the following result.

Theorem A1. *Assume that $F(\tau_*) = p_1 = T_1/m$. We have the following.*

$$\Pr\left(b_{T_1}^2 \geq \frac{p_1^2}{2\rho_0^2}\right) \leq \Pr\left(b_{T_1}^2 \geq \frac{9}{8}\tau_*\right) \leq \exp\left(-\frac{mp_1^2}{144}\exp\left(\frac{-p_1^2}{2\rho_0^2}\right)\right). \tag{A5}$$

Proof. Let $\psi := F(\tau_* + \frac{1}{8}\tau_*) - F(\tau_*)$, which satisfies the following:

$$\psi \geq \frac{1}{8}\tau_* F'(\tau_* + \frac{1}{8}\tau_*), \tag{A6}$$

as the probability function $\rho(z)$ is monotonically decreasing. Since $F'(\tau) = \rho(\tau)$ and (A3), we have the following.

$$\psi^2 \geq \frac{\tau_*^2}{64} \cdot \frac{\rho_0^2 \exp\left(-9\tau_*/8\right)}{9\tau_*/8} = \frac{1}{72}\rho_0^2 \tau_* \exp\left(\frac{-9\tau_*}{8}\right)$$
$$\geq \frac{p_1^2}{288}\exp\left(-\frac{9}{8}\tau_*\right) \geq \frac{p_1^2}{288}\exp\left(-\frac{p_1^2}{2\rho_0^2}\right). \tag{A7}$$

Define the following indicator random variables:

$$\left\{r_i \,\middle|\, r_i = \mathbf{1}_{[9\tau_*/8,+\infty)}(\tau_i), i = 1, \dots, m\right\}, \tag{A8}$$

where τ_i i.i.d. obeys the chi-squared distribution with the probability function (A1) and the characteristic function of the following.

$$\mathbf{1}_{[9\tau_*/8,+\infty)}(\tau_i) = \begin{cases} 1, & \tau_i \geq 9\tau_*/8; \\ 0, & \text{otherwise.} \end{cases} \tag{A9}$$

The event $\left\{b_{T_1}^2 \geq 9\tau_*/8\right\}$ means that at least $m - T_1 + 1$ measurements are larger than $9\tau_*/8$. Therefore, we have the following.

$$\Pr\left(b_{T_1}^2 \geq 9\tau_*/8\right) = \Pr\left(\sum_{i=1}^{m} r_i \geq m - T_1 + 1\right) = \Pr\left(\sum_{i=1}^{m} r_i > m - T_1\right). \tag{A10}$$

The expectation of r_i is then

$$\mathbb{E}r_i = 1 - F(9\tau_*/8).$$

r_i is the bounded distribution and the upper bound of the sum of r_i can be provided by the well-known Hoeffding's inequality.

Lemma A3 (Hoeffding's inequality). *Let $X_1, \cdots, X_n$ be i.i.d. random variables bounded by the interval $[a, b]$. Then, the following is the case.*

$$\Pr(\overline{X} - EX \geq t) \leq \exp\left(-\frac{2nt^2}{(b-a)^2}\right) \tag{A11}$$

Using (A10), Hoeffding's inequality yields the following:

$$\Pr\left(b_{T_1}^2 \geq \tau_* + \frac{1}{8}\tau_*\right) = \Pr\left(\sum_{i=1}^{m} r_i > m - T_1\right) \tag{A12}$$

$$= \Pr\left(\frac{1}{m}\sum_{i=1}^{m} r_i - \mathbb{E}r_i > 1 - T_1/m - \mathbb{E}r_i\right)$$

$$= \Pr\left(\frac{1}{m}\sum_{i=1}^{m} r_i - \mathbb{E}r_i > 1 - F(\tau_*) - (1 - F(9\tau_*/8))\right) \tag{A13}$$

$$\leq \exp\left(-2m\psi^2\right), \tag{A14}$$

with replacements $\tau_*/8 \to t$ and $[a, b] \to [0, 1]$ in Lemma A3. Combining (A7) and (A14), we can see that (A5) holds naturally, which completes the proof. $\quad\square$

Let us consider the i.i.d. bounded random variables of the following.

$$Z_i := b_i^2 \mathbf{1}_{[0,9\tau_*/8]}(b_i^2). \tag{A15}$$

Since $\frac{1}{T_1}\sum_{i=1}^{m} Z_i$ is larger than H_1, we can provide an upper bound of H_1 by finding the upper bound of $\frac{1}{T_1}\sum_{i=1}^{m} Z_i$.

The expectation of Z_i is the following:

$$\mathbb{E}\left(b_i^2 \mathbf{1}_{9\tau_*/8}(b_i^2)\right) = \int_0^{9\tau_*/8} x\rho(x)dx \leq \left(\frac{9}{8}\right)^{3/2} \int_0^{\tau_*} x\rho(x)dx, \tag{A16}$$

with the following.

$$\int_0^{\tau_*} x\rho(x)dx = \int_0^{\tau_*} \rho_0\sqrt{x}\exp(-x/2)dx$$

$$\leq \int_0^{\tau_*} \rho_0\sqrt{x}dx = \frac{2}{3}\rho_0\tau_*^{3/2} \leq \frac{48p_1^3}{243\rho_0^2} \leq \frac{p_1^3}{5\rho_0^2}. \tag{A17}$$

After computing the expectation of Z_i, we can bound $\sum_{i=1}^{m} Z_i$ by Hoeffding's inequality for bounded random variables similarly.

Since $0 \leq Z_i \leq 9\tau_*/8$, we can obtain the following result by replacing $p_1^3/(20\rho_0^2) \to t$ in Lemma A3.

$$\Pr(\frac{1}{m}\sum_{i=1}^{m} Z_i - \mathbb{E}Z_i \geq \frac{p_1^3}{20\rho_0^2}) \leq \exp\left(-\frac{1}{50}mp_1^2\right). \tag{A18}$$

Hence, we have the following.

$$\Pr\left(\frac{1}{T_1}\sum_{i=1}^{T} Z_i \geq \frac{p_1^2}{4\rho_0^2}\right) \leq \exp\left(-\frac{mp_1^2}{50}\right). \tag{A19}$$

Therefore, we obtain the upper bound of H_1 under statistical meaning.

$$\Pr(H_1 \leq \frac{p_1^2}{4\rho_0^2}) \geq 1 - \exp\left(-\frac{mp_1^2}{50}\right). \tag{A20}$$

Appendix B. The Bound Estimation of τ_*

Appendix B.1. The Upper Bound for τ_*

Construct the following function.

$$\hat{\rho}(x) = \rho_0\frac{1 - \sqrt{x}/2}{\sqrt{x}}, \tag{A21}$$

Then, it can be proved that the following is the case.

$$e^{-x^2/2} > 1 - x/2, \qquad \forall x \in (0, 2]. \tag{A22}$$

Therefore, we have the following.

$$\rho(x) > \hat{\rho}(x), \qquad \forall x \in (0, 4]. \tag{A23}$$

Define $\hat{F}(\tau) = \int_0^\tau \hat{\rho}(x) dx$. Then, it is obvious that the following is the case.

$$F(x) > \hat{F}(x), \qquad \forall x \in (0, 4]. \tag{A24}$$

Since $F(\tau)$ and $\hat{F}(\tau)$ are both strictly monotone decreasing over $(0, 4]$, their inverse functions satisfy the following:

$$\tau_* = F^{-1}(p) < \hat{F}^{-1}(p) := \hat{\tau}, \qquad \forall p \in (0, \hat{F}(4)]. \tag{A25}$$

where $\hat{F}(4)$ is about 0.79.

In the orthogonality promoting apporach using inverse power, $|\mathcal{I}|/m$ is always less than $1/2$ to achieve a tolerant performance. Therefore, in our concerned range ($0 < p < 1/2$), it is accessible to estimate τ_* by function $\hat{F}^{-1}$.

Now, we compute $\hat{F}$.

$$\hat{F}(\tau) = \int_0^\tau \rho_0 \frac{1 - \sqrt{x}/2}{\sqrt{x}} dx = \rho_0 \left(4\sqrt{\tau} - \tau \right)/2. \tag{A26}$$

Then, we obtain $\hat{\tau}$ by solving $\hat{F}(\tau) = p$.

$$\tau_\pm = \left(2 \pm \sqrt{4 - 2p/\rho_0} \right)^2. \tag{A27}$$

Picking out the unreasonable solution with excess $(0, 4]$, we have the following.

$$\hat{\tau} = \left(2 - \sqrt{4 - 2p/\rho_0} \right)^2. \tag{A28}$$

Since $2p/\rho_0 < 1$ and $\forall s \in [0, 1]$, $2 - \sqrt{4 - s} < s/3$, we can determine the following upper bound of $\hat{\tau}$.

$$\hat{\tau} < \left(\frac{2p}{3\rho_0} \right)^2. \tag{A29}$$

Appendix B.2. Lower Bound for τ_*

Similarly, consider the following function.

$$\check{\rho}(x) = \rho_0 \frac{1}{\sqrt{x}} \geq \rho(x). \tag{A30}$$

Define $\check{F}(\tau) = \int_0^\tau \check{\rho}(x) dx$. Then, it is obvious that the following is the case.

$$F(\tau) \leq \check{F}(\tau). \tag{A31}$$

Therefore, we have the following.

$$\check{\tau} = \check{F}^{-1}(p) = \left(\frac{p}{2\rho_0} \right)^2 < F^{-1}(p) = \tau_*. \tag{A32}$$

Appendix C. The Bound of K

The sub-exponential norm of ξ is defined by the following.

$$
\begin{aligned}
K &= \sup_{p \geq 1} p^{-1} \big(\mathbb{E} |Z_k|^p \big)^{1/p} \\
&= \sup_{p \geq 1} p^{-1} \left(\int_0^{+\infty} x^p \rho_0 \frac{\exp(-x)}{\sqrt{x}} dx \right)^{1/p}.
\end{aligned}
\tag{A33}
$$

Denote $I_p = \int_0^{+\infty} x^p \rho_0 \frac{\exp(-x)}{\sqrt{x}} dx$, then through integrating by part, we have the following.

$$
\begin{aligned}
I_p &= - \int_0^{\infty} \rho_0 x^{p-1/2} d \exp(-x) \\
&= -\left(\rho_0 e^{-x} x^{p-1/2} \right)\Big|_0^{+\infty} + (p - 1/2) \int_0^{+\infty} \rho_0 x^{p-1-1/2} \exp(-x) dx \\
&= (p - 1/2) I_{p-1}.
\end{aligned}
$$

Noticing that $I_0 = 1$, we have the following.

$$
I_p = \prod_{i=1}^{p} (i - 1/2) \leq p!.
\tag{A34}
$$

It is also obvious that the mean value of ξ is 1. Therefore, $Z_i = \xi_i - 1$ obeys a centered sub-exponential distribution since its mean value is 0 and sub-exponential norm is limited.

$$
\begin{aligned}
K &= \sup_{p \geq 1} p^{-1} \big(\mathbb{E} |Z_k|^p \big)^{1/p} \\
&= \sup_{p \geq 1} p^{-1} \big(\mathbb{E} |\xi - 1|^p \big)^{1/p} \\
&\leq 1 + \sup_{p \geq 1} p^{-1} \big(\mathbb{E} |\xi|^p \big)^{1/p} \\
&\leq 1 + \sup_{p \geq 1} p^{-1} (p!)^{1/p} \leq 2.
\end{aligned}
$$

References

1. Miao, J.; Charalambous, P.; Kirz, J.; Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **1999**, *400*, 342. [CrossRef]
2. Shechtman, Y.; Eldar, Y.C.; Cohen, O.; Chapman, H.N.; Miao, J.; Segev, M. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Process. Mag.* **2015**, *32*, 87–109. [CrossRef]
3. Stefik, M. Inferring DNA structures from segmentation data. *Artif. Intell.* **1978**, *11*, 85–114. [CrossRef]
4. Fienup, C.; Dainty, J. Phase retrieval and image reconstruction for astronomy. *Image Recover. Theory Appl.* **1987**, *231*, 275.
5. Luo, Q.; Wang, H. The Matrix Completion Method for Phase Retrieval from Fractional Fourier Transform Magnitudes. *Math. Probl. Eng.* **2016**, *2016*, 4617327. [CrossRef]
6. Candes, E.J.; Strohmer, T.; Voroninski, V. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pure Appl. Math.* **2013**, *66*, 1241–1274. [CrossRef]
7. Waldspurger, I.; d'Aspremont, A.; Mallat, S. Phase recovery, maxcut and complex semidefinite programming. *Math. Program.* **2015**, *149*, 47–81. [CrossRef]
8. Netrapalli, P.; Jain, P.; Sanghavi, S. Phase retrieval using alternating minimization. *IEEE Trans. Signal Process.* **2015**, *63*, 4814–4826. [CrossRef]
9. Elser, V. Phase retrieval by iterated projections. *JOSA A* **2003**, *20*, 40–55. [CrossRef] [PubMed]
10. Candes, E.J.; Li, X.; Soltanolkotabi, M. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory* **2015**, *61*, 1985–2007. [CrossRef]
11. Zhang, H.; Liang, Y. Reshaped wirtinger flow for solving quadratic system of equations. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2622–2630.

12. Wang, G.; Giannakis, G.; Saad, Y.; Chen, J. Solving most systems of random quadratic equations. *arXiv* **2017**, arXiv:1705.10407.
13. Wang, G.; Giannakis, G.B.; Saad, Y.; Chen, J. Phase retrieval via reweighted amplitude flow. *IEEE Trans. Signal Process.* **2018**, *66*, 2818–2833. [CrossRef]
14. Luo, Q.; Wang, H.; Lin, S. Phase retrieval via smoothed amplitude flow. *Signal Process.* **2020**, *177*, 107719. [CrossRef]
15. Luo, Q.; Lin, S.; Wang, H. Robust phase retrieval via median-truncated smoothed amplitude flow. In *Inverse Problems in Science and Engineering*; Taylor & Francis: Abingdon-on-Thames, UK, 2021; pp. 1–17.
16. Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*; Springer Science: New York, NY, USA, 2013; Volume 87.
17. Chen, P.; Fannjiang, A.; Liu, G.R. Phase retrieval by linear algebra. *SIAM J. Matrix Anal. Appl.* **2017**, *38*, 854–868. [CrossRef]
18. Wishart, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **1928**, *20*, 32–52. [CrossRef]
19. Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *arXiv* **2010**, arXiv:1011.3027
20. Li, C.K.; Li, R.C. A note on eigenvalues of perturbed Hermitian matrices. *Linear Algebra Its Appl.* **2005**, *395*, 183–190. [CrossRef]
21. Chi, Y.; Lu, Y.M.; Chen, Y. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.* **2019**, *67*, 5239–5269. [CrossRef]