Uniform Cramér–Rao Lower Bound for Phase Retrieval: An Empirical Study

Zongyu Li
Dept. of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, USA
zonyul@umich.edu

Jeffrey A. Fessler
Dept. of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, USA
fessler@umich.edu

2022 56th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5906-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEEECONF56349.2022.10052041

***Abstract*—This paper derives and analyzes the uniform Cramér–Rao Lower Bound (UCRLB) for the phase retrieval problem, and then compares the bias-variance trade-offs of several phase retrieval algorithms (e.g., Wirtinger flow (WF), Gerchberg-Saxton (GS), phaselift, majorize-minimize (MM), alternating direction method of multipliers (ADMM)) that were derived from maximum likelihood (ML) estimates where the measurements follow independent Gaussian distribution. We also consider regularizers that exploit assumed properties of the latent signal, e.g., ℓ_2 norm and ℓ_1 norm (approximated by the Huber function) that corresponds to the sparsity of finite differences (anisotropic total variation (TV)) or of the detail coefficients of a discrete wavelet transform. Simulation results show that the classical CRLB fails to lower-bound the variances of phase retrieval algorithms, whereas the UCRLB provides a variance lower-bound. Simulation results also show that the regularized algorithms that better approximate the properties of the true signal have better bias-variance trade-offs (when compared to UCRLB).**

***Index Terms*—Uniform Cramér–Rao Lower Bound, phase retrieval, inverse problem.**

I. Introduction

It is well known that the variance of any unbiased estimator is bounded by the Cramér–Rao Lower Bound (CRLB). However, many estimators, e.g., derived from regularized maximum likelihood (maximum a posteriori in Bayesian setting) are typically biased. Hence their variance cannot be bounded by the classical CRLB. Hero *et al.* [1] proposed the uniform CRLB (UCRLB), which is a bound on the smallest attainable variance that can be achieved using any estimator with bias gradient of which norm is bounded by a constant. This paper analyzes the UCRLB for the challenging phase retrieval problem.

Phase retrieval refers to the problem of recovering a signal or image from intensity (or magnitude)-only measurements [2, 3]. It is inherently ill-posed since many signals share the same magnitude spectrum. Mathematically, consider a length-N signal $\boldsymbol{x}$ reconstructed from M squared-magnitude measurements $\boldsymbol{y}$ based on the maximum likelihood (ML) estimate:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{R}^N}{\arg\min} \, f(\boldsymbol{x}), \quad f(\boldsymbol{x}) \triangleq \frac{1}{2\sigma^2} \sum_{i=1}^{M} \left(y_i - \left| \boldsymbol{a}_i' \boldsymbol{x} \right|^2 \right)^2, \quad (1)$$

where $\boldsymbol{a}_i' \in \mathbb{C}^N$ denotes the ith row of the system matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, $i = 1, \ldots, M$. For simplicity, this paper assumes the noise samples are i.i.d. Gaussian distributed with zero mean and known standard deviation σ.

To solve (1), many algorithms have been proposed, such as Wirtinger Flow (WF) [4] and its variants [5–7] that descend the cost function with a (projected/thresholded/truncated) Wirtinger gradient using an appropriate step size [8], where the Wirtinger gradient is defined as

$$\nabla f(\boldsymbol{x}) = \frac{2}{\sigma^2} \mathrm{real}\left\{ \boldsymbol{A}' \, \mathrm{diag}\{ |\boldsymbol{A}\boldsymbol{x}|^2 - \boldsymbol{y} \} \boldsymbol{A}\boldsymbol{x} \right\}, \quad (2)$$

where $\mathrm{diag}\{\cdot\}$ denotes a diagonal matrix constructed by a vector. Another famous method is "matrix lifting" [3, 4, 9] that constructs a rank-one matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}'$ so that it converts the original problem to a semi-definite programming (SDP) after relaxing the rank-one constraint. To monotonically descend the cost function (1), one method is majorizer-minimize (MM) that iteratively construct a majorizer and minimize the surrogate function [10, 11]. An alternative to (1) (aka intensity model) is the magnitude model that works with the square root of $\boldsymbol{y}$. For example, [12] proposes an algorithm known as Gerchberg Saxton (GS) that introduces a new variable to represent the phase and alternatively update the two variables; [13] uses alternating direction methods of multipliers (ADMM) with variables representing the magnitude and the phase of the signal. Other methods such as Gauss-Newton methods [14] and iterative soft-thresholding with exact line search algorithm (STELA) have also been proposed to solve (1).

There have been several studies involving the classical CRLB for phase retrieval in the literature [13, 15, 16]. Balan *et al.* [16] derived and analyzed the CRLB for two different types of phase retrieval problems. Qian *et al.* [15] proposed a novel method known as feasible point pursuit (FPP) that is based on quadratically constrained quadratic programming (QCQP) and is measured against the classical CRLB. However, we argue that the classical CRLB may have limited applicability because it is unknown whether an estimator determined by an iterative phase retrieval algorithm is unbiased or not, especially for those with tuning parameters or regularizers. Instead, the UCRLB should be used to evaluate these algorithms, as will be presented next.

II. Methods

A. Uniform CRLB

As discussed in Section I, a limitation of the classical CRLB analysis is that unbiased estimation is often impractical. The uniform CRLB is applicable for biased estimation [1]. For simplicity, this paper considers the scalar UCRLB, which is the smallest attainable variance of a single element of the true signal. Following Theorem 1 in [1], the scalar UCRLB can be written as

$$\begin{aligned} &\mathrm{Var}(\hat{\boldsymbol{x}}_j) \geq B_\gamma \triangleq (\boldsymbol{e}_j + \boldsymbol{v}_\gamma)' \boldsymbol{F}^+ (\boldsymbol{e}_j + \boldsymbol{v}_\gamma), \\ &\boldsymbol{v}_\gamma \triangleq -(\gamma \boldsymbol{C} + \boldsymbol{F}^+)^{-1} \boldsymbol{F}^+ \boldsymbol{e}_j, \\ &\boldsymbol{F} \triangleq \frac{4}{\sigma^2} \boldsymbol{G}\boldsymbol{G}', \quad \boldsymbol{G} \triangleq \mathrm{real}\{\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}|\}\}, \end{aligned} \tag{3}$$

where $\boldsymbol{F}^+$ denotes the Moore–Penrose inverse of $\boldsymbol{F}$, $\boldsymbol{C}$ is a positive definite matrix, and $\boldsymbol{e}_j$ is a unit vector whose the jth element is 1. The scalar UCRLB can be represented by a set of points $\left(\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}}, \sqrt{B_\gamma}\right)$ with varying γ and an appropriate choice of $\boldsymbol{C}$. This paper used $\boldsymbol{C} = \boldsymbol{I}$ for simplicity (so $\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}} = \|\boldsymbol{v}_\gamma\|_2$).

With the UCRLB, the limiting variance of an estimator becomes a function of the norm of bias gradient, where empirically we approximate the bias gradient by [1]:

$$\begin{aligned} \nabla \boldsymbol{b}(\hat{\boldsymbol{x}}) &\triangleq \nabla_{\boldsymbol{x}} \left(\mathbb{E}[\hat{\boldsymbol{x}}] - \boldsymbol{x}\right) \\ &\approx \frac{1}{L-1} \sum_{l=1}^{L} \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right) \left(-\nabla f(\boldsymbol{x}; \boldsymbol{y}_l)\right)' - \boldsymbol{I}, \end{aligned} \tag{4}$$

where L denotes the number of realizations of $\boldsymbol{y}$, $\boldsymbol{I}$ denotes an identity matrix, and $\hat{\boldsymbol{x}}(\boldsymbol{y}_l)$ denotes the estimator based on $\boldsymbol{y}_l$. $\overline{\hat{\boldsymbol{x}}}$ denotes the sample mean of $\hat{\boldsymbol{x}}(\boldsymbol{y}_l)$ shown as follows

$$\overline{\hat{\boldsymbol{x}}} \triangleq \frac{1}{L} \sum_{l=1}^{L} \hat{\boldsymbol{x}}(\boldsymbol{y}_l). \tag{5}$$

Next, we define the norm of bias gradient for the jth element in $\nabla \boldsymbol{b}(\hat{\boldsymbol{x}})$ by $\|\delta(\hat{\boldsymbol{x}}_j)\|_{\boldsymbol{C}}$ with

$$\delta(\hat{\boldsymbol{x}}_j) \triangleq \nabla \boldsymbol{b}(\hat{\boldsymbol{x}})' \boldsymbol{e}_j. \tag{6}$$

Then, we estimate the variance of $\hat{\boldsymbol{x}}_j$ (the jth element in $\hat{\boldsymbol{x}}$) by the sample variance

$$\mathrm{Var}(\hat{\boldsymbol{x}}_j) \approx \boldsymbol{e}_j' \left(\frac{1}{L-1} \sum_{l=1}^{L} \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right) \left(\hat{\boldsymbol{x}}(\boldsymbol{y}_l) - \overline{\hat{\boldsymbol{x}}}\right)' \right) \boldsymbol{e}_j. \tag{7}$$

Finally we compare the point $\left(\|\delta(\hat{\boldsymbol{x}}_j)\|_{\boldsymbol{C}}, \sqrt{\mathrm{Var}(\hat{\boldsymbol{x}}_j)}\right)$ with the UCRLB, i.e., the set of points $\left(\|\boldsymbol{v}_\gamma\|_{\boldsymbol{C}}, \sqrt{B_\gamma}\right)$ as we vary γ, to illustrate the bias-variance trade-off of the corresponding estimator.

B. Wirtinger Flow

To descend the cost function (1), one method is Wirtinger flow (WF) with Fisher information for step size [8]. Let subscript k denote the kth iteration. The WF update is:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mu_k \nabla f(\boldsymbol{x}_k), \tag{8}$$

where

$$\mu_k \triangleq \frac{\sigma^2 \|\nabla f(\boldsymbol{x}_k)\|_2^2}{4\|\operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}_k|\} \boldsymbol{d}_k\|_2^2}, \quad \boldsymbol{d}_k \triangleq \boldsymbol{A} \nabla f(\boldsymbol{x}_k). \tag{9}$$

To further potentially improve the reconstruction quality, one can impose a regularization term in (1) so that the cost function becomes

$$\Phi(\boldsymbol{x}) \triangleq f(\boldsymbol{x}) + \beta R(\boldsymbol{x}), \tag{10}$$

where β denotes the regularization strength.

A simple choice of $R(\boldsymbol{x})$ is $\|\boldsymbol{x}\|_2^2$, which leads to the following iteration update:

$$\begin{aligned} &\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \tilde{\mu}_k \nabla \Phi(\boldsymbol{x}_k), \quad \nabla \Phi(\boldsymbol{x}_k) \triangleq \nabla f(\boldsymbol{x}_k) + 2\beta \boldsymbol{x}_k, \\ &\tilde{\mu}_k \triangleq \frac{\sigma^2 \|\nabla \Phi(\boldsymbol{x}_k)\|_2^2}{4 \nabla \Phi(\boldsymbol{x}_k)' \left(\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}_k|^2\} \boldsymbol{A} + 2\beta \boldsymbol{I}\right) \nabla \Phi(\boldsymbol{x}_k)}. \end{aligned} \tag{11}$$

Another possible choice of regularizer is based on the assumption that $\boldsymbol{T}\boldsymbol{x}$ is approximately sparse, e.g., $\|\boldsymbol{T}\boldsymbol{x}\|_1$, for a $K \times N$ matrix $\boldsymbol{T}$. However, the ℓ_1 norm is not differentiable everywhere, so here we use the Huber function to approximate the ℓ_1 norm, so that $R(\boldsymbol{x})$ has the form

$$\begin{aligned} &R(\boldsymbol{x}) = \boldsymbol{1}' h.(\boldsymbol{T}\boldsymbol{x}; \alpha) = \min_{\boldsymbol{z}} \frac{1}{2} \|\boldsymbol{T}\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \alpha \|\boldsymbol{z}\|_1, \\ &h(t; \alpha) = \begin{cases} \frac{1}{2}|t|^2, & |t| < \alpha, \\ \alpha|t| - \frac{1}{2}\alpha^2, & \text{otherwise}, \end{cases} \end{aligned} \tag{12}$$

where $f.$ means element-wise application as in the Julia language. Then we majorize the Huber function $h(t)$ using quadratic polynomials with the optimal curvature using the ratio $\dot{h}(z)/z$ [17, p. 184], leading to iteration of the form

$$\begin{aligned} &\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \tilde{\mu}_k \nabla \Phi(\boldsymbol{x}_k), \\ &\tilde{\mu}_k \triangleq \frac{\sigma^2 \|\nabla \Phi(\boldsymbol{x}_k)\|_2^2}{4 \nabla \Phi(\boldsymbol{x}_k)' \left(\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}_k|^2\} \boldsymbol{A} + \beta \boldsymbol{T}' \boldsymbol{D} \boldsymbol{T}\right) \nabla \Phi(\boldsymbol{x}_k)}, \\ &\nabla \Phi(\boldsymbol{x}_k) \triangleq \nabla f(\boldsymbol{x}_k) + \beta \boldsymbol{T}' \dot{h}.(\boldsymbol{T}\boldsymbol{x}; \alpha), \\ &\boldsymbol{D} \triangleq \operatorname{diag}\{\min.(\alpha \oslash |\boldsymbol{T}\boldsymbol{x}_k|, 1)\}. \end{aligned} \tag{13}$$

Here we overload the notation $\tilde{\mu}_k$ and $\nabla \Phi(\boldsymbol{x}_k)$. $\oslash$ denotes element-wise division and $\dot{h}(\cdot)$ denotes the derivative of $h(\cdot)$.

C. Gerchberg Saxton (GS)

The GS algorithm introduces a variable $\boldsymbol{\theta}$ to represent the phase, leading to the following optimization problem:

$$\begin{aligned} \hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{x} \in \mathbb{C}^N,\, \boldsymbol{\theta} \in \mathbb{C}^N}{\arg\min} \|\boldsymbol{A}\boldsymbol{x} - \operatorname{diag}\{\sqrt{\boldsymbol{y}}\}\, \boldsymbol{\theta}\|_2^2, \\ &\text{subject to} \quad |\theta_i| = 1, \ i = 1, ..., N. \end{aligned} \tag{14}$$

The resulting GS iteration update is

$$\begin{aligned} &\boldsymbol{\theta}_{k+1} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}_k), \\ &\boldsymbol{x}_{k+1} = \left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1} \boldsymbol{A}' \operatorname{diag}\{\sqrt{\boldsymbol{y}}\} \boldsymbol{\theta}_{k+1}. \end{aligned} \tag{15}$$

If calculating the inverse of $(\boldsymbol{A}'\boldsymbol{A})$ is expensive, one can run several iterations of a gradient descent method like conjugate gradient.

D. PhaseLift

The PhaseLift algorithm reformulates (1) as

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\succeq 0} \frac{1}{2}\|\mathcal{A}(\boldsymbol{X}) - \boldsymbol{y}\|_2^2 + \beta\|\boldsymbol{X}\|_*, \tag{16}$$

where $\mathcal{A}$ is a linear operator that maps from $N \times N$ Hermitian matrices to $\mathbb{R}^M$:

$$\mathcal{A}(\boldsymbol{X}) = \left[|\boldsymbol{a}_1'\boldsymbol{x}|^2, \quad ..., \quad |\boldsymbol{a}_i'\boldsymbol{x}|^2, \quad ..., \quad |\boldsymbol{a}_M'\boldsymbol{x}|^2\right]'. \tag{17}$$

We used a fast adaptive shrinkage/thresholding algorithm (FASTA) [18] to solve (16).

E. Majorize-minimize (MM)

An MM method known as "PRIME" [10] has the following iteration update:

$$\begin{aligned} \boldsymbol{W}_k &= \boldsymbol{x}_k\boldsymbol{x}_k' + \frac{1}{D}\boldsymbol{A}' \operatorname{diag}\{\boldsymbol{y} - |\boldsymbol{A}\boldsymbol{x}_k|^2\}\boldsymbol{A}, \\ \boldsymbol{x}_{k+1} &= \sqrt{\lambda_{\max}(\boldsymbol{W}_k)}\boldsymbol{u}_{\max}(\boldsymbol{W}_k), \end{aligned} \tag{18}$$

where $\lambda_{\max}$ and $\boldsymbol{u}_{\max}$ denote the largest eigenvalue (in magnitude) and the corresponding eigenvector. D is a constant that needs to be no smaller than $\lambda_{\max}(\boldsymbol{B}\boldsymbol{B}')$, where $\boldsymbol{B}$ is defined as

$$\boldsymbol{B} \triangleq \left[\operatorname{vec}(\boldsymbol{a}_1\boldsymbol{a}_1'), \quad ..., \quad \operatorname{vec}(\boldsymbol{a}_M\boldsymbol{a}_M')\right] \in \mathbb{C}^{N^2\times M}. \tag{19}$$

The most straightforward way would be to first construct $\boldsymbol{B}\boldsymbol{B}'$ or $\boldsymbol{B}'\boldsymbol{B}$ and then perform power iteration. But constructing $\boldsymbol{B}\boldsymbol{B}'$ is computationally expensive and very memory hungry. Hence, we developed an upper bound on $\lambda_{\max}(\boldsymbol{B}'\boldsymbol{B})$ that can be computed efficiently. Rewrite $\boldsymbol{B}$ as

$$\boldsymbol{B} = \boldsymbol{C}\boldsymbol{P}, \quad \boldsymbol{C} \triangleq \boldsymbol{A}' \otimes \boldsymbol{A}^T, \tag{20}$$

where $\otimes$ denotes Kronecker product. $\boldsymbol{P} \in \mathbb{R}^{M^2\times M}$ is a concatenation of unit vectors $\boldsymbol{e}_j \in \mathbb{R}^{M^2}$, where $j = 1, (M+1)+1, 2(M+1)+1, ..., (M-1)(M+1)+1$, so $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_M$. Let $\boldsymbol{v}$ denote a unit-norm eigenvector corresponding to the largest eigenvalue of $\boldsymbol{B}'\boldsymbol{B}$, so that

$$\lambda_{\max}(\boldsymbol{B}'\boldsymbol{B}) = \boldsymbol{v}'\boldsymbol{B}'\boldsymbol{B}\boldsymbol{v} = \boldsymbol{v}'\boldsymbol{P}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{P}\boldsymbol{v} = \tilde{\boldsymbol{v}}'\boldsymbol{C}'\boldsymbol{C}\tilde{\boldsymbol{v}}, \tag{21}$$

where $\tilde{\boldsymbol{v}} = \boldsymbol{P}\boldsymbol{v}$. Because $\tilde{\boldsymbol{v}}'\tilde{\boldsymbol{v}} = \boldsymbol{v}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{v} = 1$, then by the spectral theorem,

$$\tilde{\boldsymbol{v}}'\boldsymbol{C}'\boldsymbol{C}\tilde{\boldsymbol{v}} = \|\boldsymbol{C}\tilde{\boldsymbol{v}}\|_2^2 \le \|\boldsymbol{C}\|_2^2 = \lambda_{\max}(\boldsymbol{C}'\boldsymbol{C}). \tag{22}$$

Using the property of Kronecker product, it follows that

$$\lambda_{\max}(\boldsymbol{B}'\boldsymbol{B}) \le \lambda_{\max}(\boldsymbol{C}'\boldsymbol{C}) = \lambda_{\max}^2(\boldsymbol{A}'\boldsymbol{A}). \tag{23}$$

To compensate for outliers, reference [11] used an ℓ_1 norm for the data-fit term also with ℓ_1 norm regularizer that assumes the true signal is sparse, leading to the following cost function:

$$\sum_{i=1}^{M} \left|y_i - \left|\boldsymbol{a}_i'\boldsymbol{x}\right|^2\right| + \beta\|\boldsymbol{x}\|_1. \tag{24}$$

They then construct a convex but non-smooth majorizer and minimize the surrogate iteratively.

F. Alternating direction method of multipliers (ADMM)

Reference [19] proposed to use ADMM with variable splitting $\boldsymbol{a}_i'\boldsymbol{x} = u_i\boldsymbol{e}^{\imath\omega_i}$ to represent the magnitude and phase separately, which leads to easy least-squares variable update.

III. Experiment

A. Compared algorithms

For unregularized algorithms, we compared WF, GS, phaselift with variant β for the nuclear norm, PRIME and ADMM with variant augmented Lagrangian penalty parameter. For regularized algorithms, we compared WF with ℓ_2 norm regularizer and ℓ_1 norm approximated by the Huber function $h.(\boldsymbol{T}\boldsymbol{x};\alpha)$. We set $\boldsymbol{T}$ to be the total variation (TV) matrix or the detailed coefficients of orthogonal discrete wavelet transform (ODWT).

B. NRMSE Comparison

We first compared the NRMSE of the estimation results of different phase retrieval algorithms. We used an empirical transmission system matrix of size 4096×256 [20, 21], where the true image size is 16×16 as shown in Fig. 1 (a). We used spectral initialization [4], i.e., the (real part of) leading eigenvector of $\boldsymbol{A}'\operatorname{diag}\{\boldsymbol{y}\}\boldsymbol{A}$ as the initial estimate of $\hat{\boldsymbol{x}}$. To handle the phase ambiguity, before computing the bias gradient, variance and the normalized root mean square error (NRMSE), we corrected the phase of $\hat{\boldsymbol{x}}$ by

$$\hat{\boldsymbol{x}}_{\text{corrected}} \triangleq \operatorname{sign}(\langle\hat{\boldsymbol{x}}, \boldsymbol{x}\rangle)\,\hat{\boldsymbol{x}}. \tag{25}$$

The NRMSE is defined as

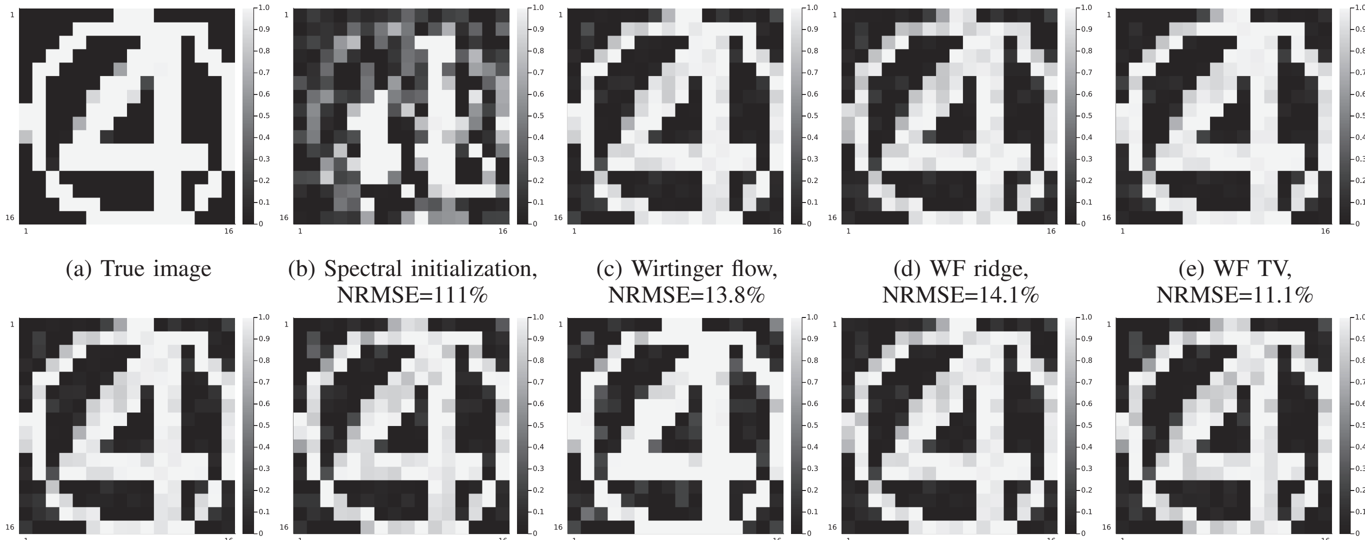
$$\text{NRMSE} \triangleq \frac{\|\hat{\boldsymbol{x}}_{\text{corrected}} - \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}. \tag{26}$$

Fig. 1 compares the reconstruction NRMSE among different phase retrieval algorithms; WF with TV and ODWT regularizer showed improved accuracy compared to unregularized algorithms, demonstrating the value of regularization. As regularized algorithms can yield biased estimates, the next section examines the bias-variance trade-off.

C. Bias-variance Comparison

We compared the bias-variance trade-off of phase retrieval algorithms. We used an oversampled random complex Gaussian matrix of size 400×16. The true signal was a 1D real signal shown in Fig. 2. The standard deviation of noise σ was set to be 1. To reduce the estimation uncertainty for bias-gradient and variance, we drew $L = 10000$ realizations of $\boldsymbol{y}$. We used ℓ_2 norm regularizer and ℓ_1 norm approximated by the Huber function $h.(\boldsymbol{T}\boldsymbol{x};\alpha)$. We set $\boldsymbol{T}$ to be the total variation (TV) matrix or the detailed coefficients of orthogonal discrete wavelet transform (ODWT), and α was set to be 0.1. All algorithms were implemented in Julia 1.6.3 (MacOS) and were ran for 100 iterations.

Fig. 3 shows that although the classical CRLB lower-bounds the variance of algorithms like PhaseLift and ADMM, it fails to bound the variance of WF and PRIME. In contrast, variances of all algorithms were lower-bounded by the UCRLB.



(a) True image
(b) Spectral initialization, NRMSE=111%
(c) Wirtinger flow, NRMSE=13.8%
(d) WF ridge, NRMSE=14.1%
(e) WF TV, NRMSE=11.1%
(f) WF ODWT, NRMSE=12.5%
(g) Gerchberg Saxton, NRMSE=16.7%
(h) PhaseLift, NRMSE=30.7%
(i) PRIME, NRMSE=13.5%
(j) ADMM, NRMSE=16.3%

Fig. 1: NRMSE comparison of images estimated by different phase retrieval algorithms. For visualization purpose, the color scales of all images were set to $[0, 1]$.

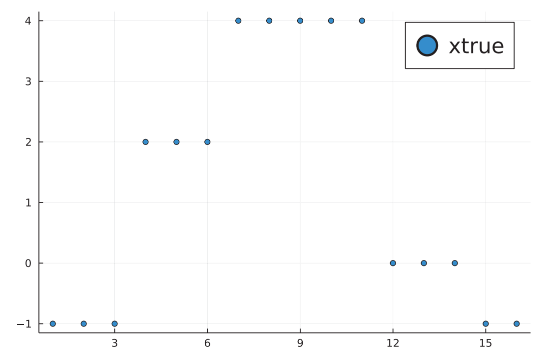


Fig. 2: The true signal used in the bias-variance trade-off experiment.

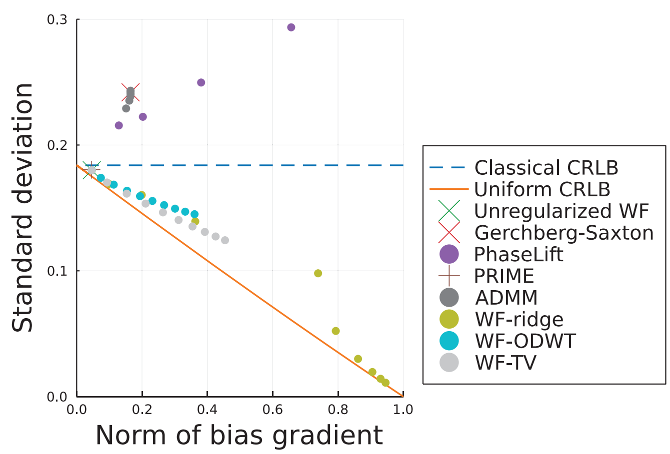


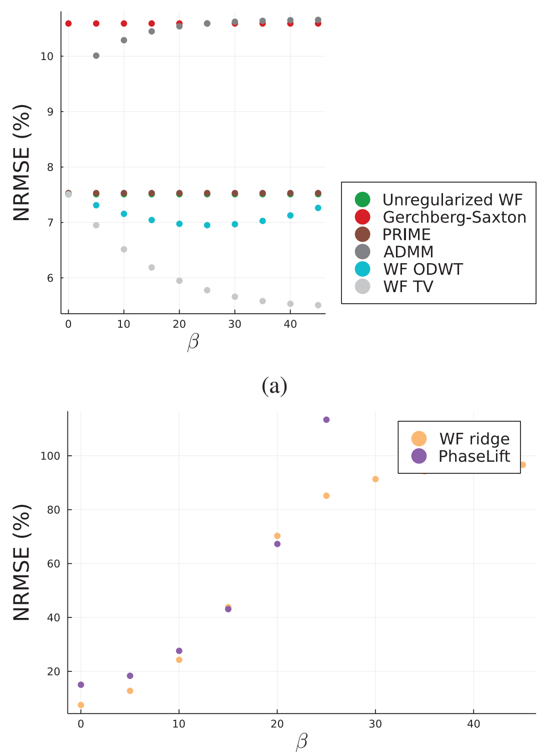
Fig. 3: The bias-variance trade-off of the uniform CRLB and variants of phase retrieval algorithms regarding the first element in the true signal. For regularized algorithms, we varied β from 1 to 100 with interval 5.

For regularized algorithms, we found that WF with ℓ_2 norm regularization (WF-ridge) achieved the lowest variance for some β [22], but at the cost of the largest bias; WF-TV is closer to the UCRLB compared to WF-ODWT, presumably due to the true signal is piece-wise uniform, which better matches the assumption of TV regularization. This was also evident in Fig. 4 (a), where we found WF-TV consistently showed lower NRMSE than WF-ODWT with varied β. In Fig. 4 (b), we found that for WF-ridge and PhaseLift, the NRMSE was sensitive to the change of β and a small β might be more preferable.

IV. Discussion and Conclusion

In this paper, we first derived the uniform CRLB for the phase retrieval problem, and then compared the bias-variance trade-off for some commonly used phase retrieval algorithms. Simulation results showed that the phase retrieval algorithms can be biased so that the classical CRLB can fail to bound their variances. Simulation results showed that regularizers that better match the assumed property of the true signal has better bias-variance trade-offs (when compared to UCRLB).

Future work includes investigating the bias-resolution-variance trade-off [23], e.g., estimators with the same bias-gradient norm but with different resolution properties, which could lead to some counter-intuitive results when applying the UCRLB in imaging systems like SPECT [24]. Future work also includes experimenting on more system matrices such as discrete Fourier transform (DFT), comparing with more



(a)

(b)

Fig. 4: NRMSE of phase retrieval algorithms with varying regularization parameter β.

phase retrieval algorithms such as machine learning methods [25–27], and testing on a wider variety of signal and image dataset.

References

[1] A. Hero, J. Fessler, and M. Usman. "Exploring estimator bias-variance tradeoffs using the uniform CR bound". In: *IEEE Trans. Sig. Proc.* 44.8 (1996), pp. 2026–2041. DOI: 10.1109/78.533723.

[2] M. V. Klibanov, P. E. Sacks, and A. V. Tikhonravov. "The phase retrieval problem". In: *Inverse Problems* 11.1 (1995), pp. 1–28. DOI: 10.1088/0266-5611/11/1/001.

[3] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. "Phase Retrieval with Application to Optical Imaging: A contemporary overview". In: *IEEE Signal Processing Magazine* 32.3 (2015), pp. 87–109. DOI: 10.1109/MSP.2014.2352673.

[4] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski. "Phase retrieval via matrix completion". In: *SIAM J. Imaging Sci.* 6.1 (2013), 199–225. DOI: 10.1137/110848074.

[5] X. Jiang, S. Rajan, and X. Liu. "Wirtinger Flow Method With Optimal Stepsize for Phase Retrieval". In: *IEEE Sig. Proc. Letters* 23.11 (2016), pp. 1627–1631. DOI: 10.1109/LSP.2016.2611940.

[6] T. T. Cai, X. Li, and Z. Ma. "Optimal Rates of Convergence for Noisy Sparse Phase Retrieval via Thresholded Wirtinger Flow". In: *Annals Stat.* 44.5 (2016), pp. 2221–2251. DOI: 10.1214/16-AOS1443.

[7] M. Soltanolkotabi. "Structured Signal Recovery From Quadratic Measurements: Breaking Sample Complexity Barriers via Nonconvex Optimization". In: *IEEE Trans. Info. Theory* 65.4 (2019), pp. 2374–2400. DOI: 10.1109/TIT.2019.2891653.

[8] Z. Li, K. Lange, and J. A. Fessler. "Poisson Phase Retrieval in Very Low-Count Regimes". In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 838–850. DOI: 10.1109/TCI.2022.3209936.

[9] E. J. Candès, T. Strohmer, and V. Voroninski. "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming". In: *Comm. Pure Appl. Math.* 66.8 (), pp. 1241–1274. DOI: 10.1002/cpa.21432.

[10] T. Qiu, P. Babu, and D. P. Palomar. "PRIME: Phase Retrieval via Majorization-Minimization". In: *IEEE Trans. Sig. Proc.* 64.19 (2016), pp. 5174–5186. DOI: 10.1109/TSP.2016.2585084.

[11] D. S. Weller, A. Pnueli, G. Divon, O. Radzyner, Y. C. Eldar, and J. A. Fessler. "Undersampled Phase Retrieval With Outliers". In: *IEEE Transactions on Computational Imaging* 1.4 (2015), pp. 247–258. DOI: 10.1109/TCI.2015.2498402.

[12] R. W. Gerchberg and W. O. Saxton. "Practical Algorithm for Determination of Phase from Image and Diffraction Plane Pictures". In: *OPTIK* 35.2 (1972), 237–246.

[13] J. Liang, P. Stoica, Y. Jing, and J. Li. "Phase Retrieval via the Alternating Direction Method of Multipliers". In: *IEEE Signal Processing Letters* 25.1 (2018), pp. 5–9. DOI: 10.1109/LSP.2017.2767826.

[14] B. Gao and Z. Xu. "Phaseless Recovery Using the Gauss-Newton Method". In: *IEEE Trans. Sig. Proc.* 65.22 (2017), pp. 5885–5896. DOI: 10.1109/TSP.2017.2742981.

[15] C. Qian, N. D. Sidiropoulos, K. Huang, L. Huang, and H. C. So. "Phase Retrieval Using Feasible Point Pursuit: Algorithms and Cramér–Rao Bound". In: *IEEE Trans. Sig. Proc.* 64.20 (2016), pp. 5282–5296. DOI: 10.1109/TSP.2016.2593688.

[16] R. Balan and D. Bekkerman. "The Cramer-Rao Lower Bound in the Phase Retrieval Problem". In: *SampTA*. 2019, pp. 1–5. DOI: 10.1109/SampTA45681.2019.9030920.

[17] P. J. Huber. *Robust statistics*. New York: Wiley, 1981.

[18] T. Goldstein, C. Studer, and R. Baraniuk. *A Field Guide to Forward-Backward Splitting with a FASTA Implementation*. 2014. DOI: 10.48550/ARXIV.1411.3406.

[19] J. Liang, P. Stoica, Y. Jing, and J. Li. "Phase Retrieval via the Alternating Direction Method of Multipliers". In: *IEEE Signal Processing Letters* 25.1 (2018), pp. 5–9. DOI: 10.1109/LSP.2017.2767826.

[20] R. Chandra, Z. Zhong, J. Hontz, V. McCulloch, C. Studer, and T. Goldstein. "PhasePack: A Phase Retrieval Library". In: *Asil. Conf. Sig. Sys. Comp.* (2017), pp. 1617–1621.

[21] C. A. Metzler, M. K. Sharma, S. Nagesh, R. G. Baraniuk, O. Cossairt, and A. Veeraraghavan. "Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset". In: *Proc. Intl. Conf. Comp. Photography*. 2017, 1–16. DOI: 10.1109/ICCPHOT.2017.7951483.

[22] Y. Eldar. "Minimum variance in biased estimation: bounds and asymptotically optimal estimators". In: *IEEE Trans. Sig. Proc.* 52.7 (2004), pp. 1915–1930. DOI: 10.1109/TSP.2004.828929.

[23] T. Kragh and A. Hero. "Bias-resolution-variance tradeoffs for single pixel estimation tasks using the uniform Cramer-Rao bound". In: *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*. Vol. 2. 2000, 15/296–15/298 vol.2. DOI: 10.1109/NSSMIC.2000.950124.

[24] L. Meng and N. Clinthorne. "A modified uniform Cramer-Rao bound for multiple pinhole aperture design". In: *IEEE Trans. Med. Imag.* 23.7 (2004), pp. 896–902. DOI: 10.1109/TMI.2004.828356.

[25] P. Hand, O. Leong, and V. Voroninski. "Phase Retrieval Under a Generative Prior". In: *NeurIPS*. Vol. 31. Curran Associates, Inc., 2018.

[26] Y. Zhang, M. A. Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, and P. Villanueva-Perez. "PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets". In: *Opt. Express* 29.13 (2021), pp. 19593–19604. DOI: 10.1364/OE.423222.

[27] C. Metzler, P. Schniter, A. Veeraraghavan, and R. Baraniuk. "prDeep: Robust Phase Retrieval with a Flexible Deep Network". In: *ICML*. Vol. 80. PMLR, 2018, pp. 3501–3510.