Linear Spectral Estimators and an Application to Phase Retrieval

Ramina Ghods [1] **Andrew S. Lan** [2] **Tom Goldstein** [3] **Christoph Studer** [1]

Abstract

Phase retrieval refers to the problem of recovering real- or complex-valued vectors from magnitude measurements. The best-known algorithms for this problem are iterative in nature and rely on so-called spectral initializers that provide accurate initialization vectors. We propose a novel class of estimators suitable for general nonlinear measurement systems, called linear spectral estimators (LSPEs), which can be used to compute accurate initialization vectors for phase retrieval problems. The proposed LSPEs not only provide accurate initialization vectors for noisy phase retrieval systems with structured or random measurement matrices, but also enable the derivation of sharp and nonasymptotic mean-squared error bounds. We demonstrate the efficacy of LSPEs on synthetic and real-world phase retrieval problems, and show that our estimators significantly outperform existing methods for structured measurement systems that arise in practice.

1. Introduction

Phase retrieval refers to the problem of recovering an unknown N-dimensional signal vector $\mathbf{x} \in \mathcal{H}^N$, with $\mathcal{H}$ being the set of either real ($\mathbb{R}$) or complex ($\mathbb{C}$) numbers, from the following nonlinear measurement process:

$$\mathbf{y} = f(\mathbf{A}\mathbf{x} + \mathbf{e}^z) + \mathbf{e}^y. \tag{1}$$

Here, the measurement vector $\mathbf{y} \in \mathbb{R}^M$ contains M real-valued observations, for example measured through the nonlinear function $f(z) = |z|^2$ that operates element-wise on vectors, $\mathbf{A} \in \mathcal{H}^{M \times N}$ is a given measurement matrix, and the vectors $\mathbf{e}^z \in \mathcal{H}^M$ and $\mathbf{e}^y \in \mathbb{R}^N$ model signal and measurement noises, respectively. In contrast to the majority of existing results on phase retrieval that assume randomness in the measurement matrix $\mathbf{A}$, we focus on the practical scenario in which the measurement matrix $\mathbf{A}$ is deterministic, but the signal vector $\mathbf{x}$ to be recovered as well as the two noise sources $\mathbf{e}^z$ and $\mathbf{e}^y$ are random.

[1]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY [2]Department of EE, Princeton University [3]University of Maryland. Correspondence to: Ramina Ghods <rg548@cornell.edu>, Christoph Studer <studer@cornell.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018. Copyright 2018 by the author(s).

1.1. Phase Retrieval

Phase retrieval has been studied extensively over the last decades (Gerchberg & Saxton, 1972; Fienup, 1982) and finds use in a range of applications, including imaging (Fogel et al., 2016; Yeh et al., 2015; Holloway et al., 2016), microscopy (Kou et al., 2010; Faulkner & Rodenburg, 2004), and X-ray crystallography (Harrison, 1993; Miao et al., 2008; Pfeiffer et al., 2006). Phase retrieval problems were solved traditionally using alternating projection methods, such as the Gerchberg-Saxton (Gerchberg & Saxton, 1972) and Fienup (Fienup, 1982) algorithms. More recent results have shown that semidefinite programming enables the design of algorithms with performance guarantees (Candès et al., 2013; Candès & Li, 2014; Candès et al., 2015a; Waldspurger et al., 2015). These methods lift the problem to a higher dimension, resulting in excessive complexity and memory requirements. To perform phase retrieval for high-dimensional problems with performance guarantees, a range of convex (Bahmani & Romberg, 2017; Goldstein & Studer, 2017; Hand & Voroninski, 2016; Dhifallah et al., 2017; Dhifallah & Lu, 2017; Yuan & Wang, 2017; Salehi et al., 2018) and nonconvex methods (Netrapalli et al., 2013; Schniter & Rangan, 2015; Candès et al., 2015b; Chen & Candès, 2015; Zhang & Liang, 2016; Wang et al., 2017a; Zhang et al., 2016; Wei, 2015; Sun et al., 2016; Zeng & So, 2017; Lu & Li, 2017; Ma et al., 2018) have been proposed recently.

1.2. Spectral Initializers

All of the above non-lifting-based phase retrieval methods rely on accurate initial estimates of the signal vector to be recovered. Such estimates are typically obtained by means of so-called *spectral initializers* put forward in (Netrapalli et al., 2013). Spectral initializers first compute a Hermitian matrix of the following form:

$$\mathbf{D}_\beta = \beta \sum_{m=1}^{M} \mathcal{T}(y_m)\mathbf{a}_m\mathbf{a}_m^H, \tag{2}$$

where $\beta > 0$ is a suitably-chosen scaling factor, y_m denotes the mth measurement, $\mathbf{a}_m^H$ corresponds to the mth row of the measurement matrix $\mathbf{A}$ and $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ is a (possibly nonlinear) *preprocessing function*. While the identity $\mathcal{T}(y) = y$ was used originally in (Netrapalli et al., 2013), recent results revealed that carefully crafted preprocessing functions yield more accurate estimates (Chen & Candès, 2015; Chen et al., 2015; Wang et al., 2017a;b; Lu & Li, 2017; Mondelli & Montanari, 2017). From the matrix $\mathbf{D}_\beta$ in (2), one then extracts the (scaled) eigenvector $\hat{\mathbf{x}}$ associated with the largest eigenvalue, which serves as an initial estimate of the solution to the phase retrieval problem.

As shown in (Netrapalli et al., 2013; Chen & Candès, 2015; Chen et al., 2015; Wang et al., 2017a;b; Lu & Li, 2017; Mondelli & Montanari, 2017), for i.i.d. Gaussian measurement matrices $\mathbf{A}$, sufficiently large measurement ratios $\delta = M/N$, and carefully crafted preprocessing functions $\mathcal{T}$, spectral initializers provide accurate initialization vectors. In fact, the results in (Mondelli & Montanari, 2017) for the large-system limit with δ fixed and $M \to \infty$ show that spectral initializers in combination with an optimal preprocessing function $\mathcal{T}$ achieve the fundamental information-theoretic limits of phase retrieval. However, the assumption of having i.i.d. Gaussian measurement matrices $\mathbf{A}$ is impractical—it is more natural to assume that the signal vector $\mathbf{x}$ is random and the measurement matrix $\mathbf{A}$ is deterministic and structured (Bendory & Eldar, 2017).

1.3. Contributions

We propose a novel class of estimators, called *linear spectral estimators (LSPEs)*, that provide accurate estimates for general nonlinear measurement systems of the form (1) and enable a nonasymptotic mean-squared error (MSE) analysis. We showcase the efficacy of LSPEs by applying them to phase retrieval problems, where we compute initialization vectors for real- and complex-valued systems with deterministic and finite-dimensional measurement matrices. For the proposed LSPEs, we derive nonasymptotic and sharp bounds on the MSE for signal estimation from phaseless measurements. We use synthetic and real-world phase retrieval problems to demonstrate that LSPEs are able to significantly outperform existing spectral initializers on systems that acquire structured measurements. We furthermore show that preprocessing the phaseless measurements enables LSPEs to generate improved initialization vectors for an even broader class of measurement systems.

1.4. Notation

Lowercase and uppercase boldface letters represent column vectors and matrices, respectively. For a matrix $\mathbf{A}$, its transpose and Hermitian conjugate is $\mathbf{A}^T$ and $\mathbf{A}^H$, respectively, and the kth row and ℓth column entry is $[\mathbf{A}]_{k,\ell} = A_{k,\ell}$. For a vector $\mathbf{a}$, the kth entry is $[\mathbf{a}]_k = a_k$. The ℓ_2-norm of $\mathbf{a}$ is denoted by $\|\mathbf{a}\|_2$ and the Frobenius norm of $\mathbf{A}$ by $\|\mathbf{A}\|_F$. The Kronecker product is $\otimes$, the Hadamard product is $\odot$, the Hadamard division is $\oslash$, and the trace operator is $\mathrm{tr}(\cdot)$. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$; the $M \times N$ all-zeros and all-ones matrices are denoted by $\mathbf{0}_{M\times N}$ and $\mathbf{1}_{M\times N}$, respectively. For a vector $\mathbf{a}$, $\mathrm{diag}(\mathbf{a})$ is a square matrix with $\mathbf{a}$ on the main diagonal; for a matrix $\mathbf{A}$, $\mathrm{diag}(\mathbf{A})$ is a column vector containing the diagonal elements of $\mathbf{A}$.

2. Linear Spectral Estimators

We start by reviewing the essentials of spectral initializers and then, introduce linear spectral estimators (LSPEs) for measurement systems of the form (1) with general nonlinearities f. We furthermore provide nonasymptotic expressions for the associated estimation error, and we compare our analytical results to that of conventional spectral initializers in (2). In Section 3, we will apply LSPEs to phase retrieval.

2.1. Spectral Estimation and Initializers

One of the key issues of the phase retrieval problem is the fact that if $\mathbf{x}$ is a solution to (1), then $e^{j\phi}\mathbf{x}$ for any $\phi \in [0, 2\pi)$ is also a valid solution (assuming $\mathcal{H} = \mathbb{C}$). Put simply, the solution is nonunique up to a global phase shift. One way of combating this issue is to directly recover the outer product $\mathbf{x}\mathbf{x}^H$ instead of $\mathbf{x}$, which is unaffected by phase shifts; this insight is the key underlying lifting-based phase retrieval methods (Candès et al., 2013; Candès & Li, 2014; Candès et al., 2015a; Waldspurger et al., 2015). With this in mind, one could envision the design of an estimator that directly minimizes the conditional MSE:

$$\dot{\mathbf{x}} = \underset{\tilde{\mathbf{x}}\in\mathcal{H}^N}{\arg\min}\, \mathbb{E}\big[\|\mathbf{x}\mathbf{x}^H - \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\|_F^2 \mid \mathbf{y}\big]. \tag{3}$$

Here, expectation is with respect to the signal vector $\mathbf{x}$ and the two noise sources $\mathbf{e}^z$ and $\mathbf{e}^y$. This optimization problem resembles that of a posterior mean estimator (PME) which is, in general, difficult to derive, even for simple observation models—for phase retrieval, we have two additional challenges: (i) nonlinear phaseless measurements as in (1) and (ii) the quantity $\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H$ has rank-1.

Spectral initializers avoid the issues of the estimator in (3) by first replacing the true outer product $\mathbf{x}\mathbf{x}^H$ with a so-called spectral estimator matrix $\mathbf{D}_\beta$ as in (2) that depends on the measurement vector $\mathbf{y}$. In a second step, one then computes the best rank-1 approximation as follows:

$$\hat{\mathbf{x}} = \underset{\tilde{\mathbf{x}}\in\mathcal{H}^N}{\arg\min}\, \|\mathbf{D}_\beta - \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\|_F^2 \tag{4}$$

from which the estimate $\hat{\mathbf{x}}$ can be extracted. By performing an eigenvalue decomposition $\mathbf{D}_\beta = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ with $\mathbf{U}^H\mathbf{U} = \mathbf{I}_M$ and the eigenvalues in the diagonal matrix

$\mathbf{\Lambda} = \operatorname{diag}([\lambda_1, \dots, \lambda_M]^T)$ are sorted in descending order of their magnitudes, a spectral initializer is given by the *scaled leading eigenvector* $\hat{\mathbf{x}} = \sqrt{\lambda_1}\mathbf{u}_1$. In practice, one can use power iterations to efficiently compute $\hat{\mathbf{x}}$.

2.2. Linear Spectral Estimators

We now propose a novel class of estimators, which we call *linear spectral estimators (LSPEs)*, that provide accurate estimates for general nonlinear measurement systems of the form (1). To this end, we borrow ideas from the spectral initializer, the PME in (3), and the linear phase retrieval algorithm put forward in (Ghods et al., 2018). In the first step, LSPEs apply a linear estimator to the nonlinear observations in $\mathcal{T}(\mathbf{y})$ to construct a *spectral estimator matrix* $\mathbf{D}_\mathbf{y}$ for which the *spectral MSE* (or matrix MSE) defined as

$$\textit{S-MSE} = \mathbb{E}\left[\left\|\mathbf{D}_\mathbf{y} - \mathbf{x}\mathbf{x}^H\right\|_F^2\right] \tag{5}$$

is minimal. We restrict ourselves to spectral estimator matrices $\mathbf{D}_\mathbf{y}$ that are *affine* in $\mathcal{T}(\mathbf{y})$, i.e., are of the form

$$\mathbf{D}_\mathbf{y} = \mathbf{W}_0 + \sum_{m=1}^{M} \mathcal{T}(y_m)\mathbf{W}_m \tag{6}$$

with $\mathbf{W}_m \in \mathcal{H}^{N\times N}$, $m = 0, \dots, M$. In the second step, we use the spectral estimator matrix $\mathbf{D}_\mathbf{y}$ to extract a (scaled) leading eigenvector as in (3), which is the *linear spectral estimate* of the signal vector $\mathbf{x}$. Intuitively, if we can construct a matrix $\mathbf{D}_\mathbf{y}$ from the preprocessed measurements in $\mathcal{T}(\mathbf{y})$ for which the S-MSE in (5) is minimal, then we expect that computing its best rank-1 approximation would yield an accurate estimate of the signal vector $\mathbf{x}$ up to a global phase shift. We will justify this claim in Section 2.3.

Mathematically, we wish to compute a matrix $\mathbf{D}_\mathbf{y}$ of the form (6) that is the solution to the following problem:

$$\underset{\substack{\widetilde{\mathbf{W}}_m \in \mathcal{H}^{N\times N} \\ m=0,\dots,M}}{\text{minimize}} \; \mathbb{E}\left[\left\|\widetilde{\mathbf{W}}_0 + \sum_{m=1}^{M} \mathcal{T}(y_m)\widetilde{\mathbf{W}}_m - \mathbf{x}\mathbf{x}^H\right\|_F^2\right]. \tag{7}$$

Clearly, the spectral estimator matrix $\mathbf{D}_\mathbf{y}$ will depend on the measurement matrix $\mathbf{A}$, the statistics of the signal to be estimated $\mathbf{x}$ and the two noise sources $\mathbf{e}^z$ and $\mathbf{e}^y$, the nonlinearity f, as well as the preprocessing function $\mathcal{T}$. For this setting, we have the following general result which summarizes the LSPE; the proof is given in Appendix A.

Theorem 1 (Linear Spectral Estimator)**.** *Let the measurement vector $\mathbf{y}$ be a result of the general measurement model in (1) and select a preprocessing function $\mathcal{T}$. Define the vector $\overline{\mathcal{T}}(\mathbf{y}) = \mathbb{E}[\mathcal{T}(\mathbf{y})]$ and assume the matrix*

$$\mathbf{T} = \mathbb{E}\left[(\mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y}))(\mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y}))^T\right]$$

is full rank. Let $\mathbf{t} \in \mathbb{R}^M$ satisfy $\mathbf{T}\mathbf{t} = \mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y})$ and

$$\mathbf{V}_m = \mathbb{E}\left[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))(\mathbf{x}\mathbf{x}^H - \mathbf{K}_\mathbf{x})\right]$$

for $m = 1, \dots, M$ with $\mathbf{K}_\mathbf{x} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$. Then, the LSPE matrix that minimizes the S-MSE in (5) is given by

$$\mathbf{D}_\mathbf{y} = \mathbf{K}_\mathbf{x} + \sum_{m=1}^{M} t_m \mathbf{V}_m. \tag{8}$$

The linear spectral estimate $\hat{\mathbf{x}}$ is then given by the scaled leading eigenvector of the matrix $\mathbf{D}_\mathbf{y}$ in (8).

The vector $\mathbf{t}$ is the only quantity in Theorem 1 that depends on the *actual* (nonlinear) observations contained in the measurement vector $\mathbf{y}$. All other quantities depend only on the first two moments of $\mathbf{x}\mathbf{x}^H$ as well as the considered signal, noise, and measurement models. The key features of the LSPE are as follows: (i) the involved quantities can often be computed in closed form (see Section 3 for two applications to phase retrieval) and (ii) LSPEs enable a nonasymptotic and sharp analysis of the associated estimation error.

Remark 1. *Theorem 1 requires the matrix $\mathbf{T}$ to be invertible. This condition is satisfied in most practical situations with nondegenerate measurement matrices $\mathbf{A}$ or in situations with nonzero measurement noise.*

2.3. Estimation Error Analysis of LSPEs

The remaining piece of the proposed LSPE is to show that the result of this two-step estimation procedure indeed yields a vector that is close to the signal vector $\mathbf{x}$. We start with the following result; the proof is given in Appendix B.

Theorem 2 (S-MSE of the LSPE)**.** *Let the assumptions of Theorem 1 hold. Then, the S-MSE in (5) for the LSPE matrix in (8) is given by*

$$\textit{S-MSE}_{\textit{LSPE}} = C_{\mathbf{x}\mathbf{x}^H} - \sum_{m=1}^{M}\sum_{m'=1}^{M} [\mathbf{T}^{-1}]_{m,m'} \operatorname{tr}\left(\mathbf{V}_m^H \mathbf{V}_{m'}\right) \tag{9}$$

with $C_{\mathbf{x}\mathbf{x}^H} = \mathbb{E}\left[\left\|\mathbf{x}\mathbf{x}^H - \mathbf{K}_\mathbf{x}\right\|_F^2\right]$.

With this result, we are ready to establish a bound on the estimation error of the LSPE. The proof of the following result follows from Theorem 2 and is given in Appendix C.

Corollary 1 (LSPE Estimation Error)**.** *Let the assumptions of Theorem 1 hold. Then, the estimation error (EER) of the LSPE satisfies the following inequality:*

$$\textit{EER}_{\textit{LSPE}} = \mathbb{E}\left[\|\hat{\mathbf{x}}\hat{\mathbf{x}}^H - \mathbf{x}\mathbf{x}^H\|_F^2\right] \leq 4\,\textit{S-MSE}_{\textit{LSPE}}. \tag{10}$$

This result implies that by minimizing the S-MSE in (5) via (7), we are also reducing the EER of the LSPE. In other words, if the spectral error $\mathbf{E} = \mathbf{D}_\mathbf{y} - \hat{\mathbf{x}}\hat{\mathbf{x}}^H$ is small, then the EER of the LSPE (10) will be small.

Remark 2. *Corollary 1 is nonasymptotic and depends on the instance of measurement matrix* $\mathbf{A}$*. This result is in stark contrast to existing performance bounds for spectral initializers (Netrapalli et al., 2013; Chen & Candès, 2015; Chen et al., 2015; Wang et al., 2017a;b) that strongly rely on randomness in the measurement matrix. In addition to randomness, the sharp performance guarantees in (Lu & Li, 2017; Mondelli & Montanari, 2017) focus on the asymptotic regime for which* $\delta = M/N$ *is fixed and* $M \to \infty$*.*

2.4. S-MSE of Spectral Initializers

We can also derive an exact expression for the S-MSE of the conventional spectral initializer in (2). We assume optimal scaling, i.e., the parameter β is set to minimize the S-MSE. The following result characterizes the S-MSE of such a scaled spectral initializer; the proof is given in Appendix D.

Proposition 1 (S-MSE of the Spectral Initializer)**.** *Let* $\mathbf{D}_\beta$ *be the conventional spectral initializer matrix in (2). Then, the optimally-scaled S-MSE defined as*

$$\textit{S-MSE}_{SI} = \min_{\beta \in \mathcal{H}} \mathbb{E}\left[\|\mathbf{D}_\beta - \mathbf{x}\mathbf{x}^H\|_F^2\right] \tag{11}$$

is given by

$$\textit{S-MSE}_{SI} = R_{\mathbf{x}\mathbf{x}^H} - \frac{\left|\sum_{m=1}^M \mathbf{a}_m^H \widetilde{\mathbf{V}}_m \mathbf{a}_m\right|^2}{\sum_{m=1}^M \sum_{m'=1}^M \widetilde{T}_{m,m'} |\mathbf{a}_m^H \mathbf{a}_{m'}|^2}, \tag{12}$$

where $R_{\mathbf{x}\mathbf{x}^H} = \mathbb{E}\left[\|\mathbf{x}\mathbf{x}^H\|_F^2\right]$, $\widetilde{\mathbf{V}}_m = \mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right]$, $m = 1, \ldots, M$, *and* $\widetilde{\mathbf{T}} = \mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right]$.

Since the matrix in (2) is a special case of the LSPE matrix in (6), we have the following simple yet important property:

$$\textit{S-MSE}_{\text{LSPE}} \leq \textit{S-MSE}_{\text{SI}}.$$

In words, the spectral MSE of the LSPE cannot be worse than that of a spectral initializer. As we will show in Section 4, LSPEs are able to outperform spectral initializers on both synthetic and real-world phase retrieval problems given that the same preprocessing function $\mathcal{T}$ is used.

3. LSPEs for Phase Retrieval Problems

The LSPE provides a framework for estimating signal vectors from the general observation model in (1). To make the concept of LSPEs explicit and to demonstrate their efficacy in practice, we now show two application examples to phase retrieval in complex-valued systems. The LSPE for real-valued phase retrieval can be found in Appendix E.

3.1. Phase Retrieval without Preprocessing

We first focus on the case where the signal vector $\mathbf{x}$ to be estimated and the measurement matrix $\mathbf{A}$ are both complex-valued. The phaseless measurements $\mathbf{y}$, however, remain real-valued. We need the following assumptions.

Assumptions 1. *Let* $\mathcal{H} = \mathbb{C}$*. Assume square absolute measurements* $f(z) = |z|^2$ *and the identity preprocessing function* $\mathcal{T}(y) = y$*. Assume that the signal vector* $\mathbf{x} \in \mathbb{C}^N$ *is i.i.d. circularly-symmetric complex Gaussian with covariance matrix* $\mathbf{C}_\mathbf{x} = \sigma_x^2 \mathbf{I}_N$*, i.e.,* $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \sigma_x^2 \mathbf{I}_N)$*. Assume that the signal noise vector* $\mathbf{e}^z$ *is circularly-symmetric complex Gaussian with covariance matrix* $\mathbf{C}_{\mathbf{e}^z}$*, i.e.,* $\mathbf{e}^z \sim \mathcal{CN}(\mathbf{0}_{M\times 1}, \mathbf{C}_{\mathbf{e}^z})$*, and the measurement noise vector* $\mathbf{e}^y$ *is a real-valued Gaussian vector with mean* $\bar{\mathbf{e}}^y$ *and covariance matrix* $\mathbf{C}_{\mathbf{e}^y}$*, i.e.,* $\mathbf{e}^y \sim \mathcal{N}(\bar{\mathbf{e}}^y, \mathbf{C}_{\mathbf{e}^y})$*. Furthermore assume that* $\mathbf{x}$*,* $\mathbf{e}^z$*, and* $\mathbf{e}^y$ *are independent.*

Under these assumptions, we can derive the following LSPE which we call LSPE-$\mathbb{C}$; the detailed derivations of this spectral estimator are given in Appendix G.

Estimator 1 (LSPE-$\mathbb{C}$)**.** *Let Assumptions 1 hold. Then, the spectral estimation matrix is given by*

$$\mathbf{D}_\mathbf{y}^\mathbb{C} = \mathbf{K}_\mathbf{x} + \sum_{m=1}^M t_m \mathbf{V}_m, \tag{13}$$

where $\mathbf{K}_\mathbf{x} = \sigma_x^2 \mathbf{I}_N$*, the vector* $\mathbf{t} \in \mathbb{R}^M$ *is given by the solution to the linear system* $\mathbf{T}\mathbf{t} = \mathbf{y} - \overline{\mathbf{y}}$ *with*

$$\begin{aligned} \overline{\mathbf{y}} &= \operatorname{diag}(\mathbf{C}_\mathbf{z}) + \bar{\mathbf{e}}^y \\ \mathbf{C}_\mathbf{z} &= \sigma_x^2 \mathbf{A}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z} \\ \mathbf{T} &= \mathbf{C}_\mathbf{z} \odot \mathbf{C}_\mathbf{z}^* + \mathbf{C}_{\mathbf{e}^y} \end{aligned}$$

and $\mathbf{V}_m = \sigma_x^4 \mathbf{a}_m \mathbf{a}_m^H$*,* $m = 1, \ldots, M$*. The spectral estimate* $\hat{\mathbf{x}}$ *is given by the (scaled) leading eigenvector of* $\mathbf{D}_\mathbf{y}^\mathbb{C}$ *in (13). Furthermore, the S-MSE is given by Theorem 2.*

We emphasize that the spectral estimator matrix in (13) resembles that of the conventional spectral initializer matrix (2) with the following key differences. First and foremost, each outer product contained in $\mathbf{V}_m = \sigma_x^4 \mathbf{a}_m \mathbf{a}_m^H$ in Estimator 1 is weighted by t_m, which is a function of *all* phaseless measurements in $\mathbf{y}$ and of the covariance matrix $\mathbf{C}_\mathbf{x}$. In contrast, each outer product in the conventional spectral initializer matrix in (2) is only weighted by the associated measurement y_m. This difference enables the LSPE to weight each outer product depending on correlations in the phaseless measurements caused by structure in the matrix $\mathbf{A}$. Second, the spectral estimator matrix includes a mean term $\mathbf{K}_\mathbf{x}$, which is absent in the spectral initializer matrix. As we will show in Section 4, for the same preprocessing function $\mathcal{T}$, Estimator 1 is able to outperform spectral initializers for systems with structured measurement matrices $\mathbf{A}$. For large i.i.d. Gaussian measurement matrices, there is no particular correlation structure to exploit and LSPEs perform on par with spectral initializers.

3.2. Phase Retrieval with Exponential Preprocessing

To demonstrate the flexibility and generality of our framework, we now design an LSPE with an exponential preprocessing function for complex-valued phase retrieval. We derive the LSPE under the following assumptions.

Assumptions 2. *Let $\mathcal{H} = \mathbb{C}$. Assume square absolute measurements $f(z) = |z|^2$ and the exponential preprocessing function $\mathcal{T}(y) = \exp(-\gamma y)$ with $\gamma > 0$, i.e., we consider*

$$\mathcal{T}(\mathbf{y}) = \exp\big(-\gamma(|\mathbf{z}|^2 + \mathbf{e}^y)\big) \quad \text{and} \quad \mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}^z,$$

where the exponential function is applied element-wise to vectors. The remaining assumptions are the same as in Assumptions 1.

We now derive the following LSPE called LSPE-Exp; the derivation of this spectral estimator is given in Appendix H.

Estimator 2 (LSPE-Exp)**.** *Let Assumptions 2 hold. Then, the spectral estimation matrix is given by*

$$\mathbf{D}_{\mathbf{y}}^{\text{Exp}} = \mathbf{K}_{\mathbf{x}} + \sum_{m=1}^{M} t_m \mathbf{V}_m, \tag{14}$$

where $\mathbf{K}_{\mathbf{x}} = \sigma_x^2 \mathbf{I}_N$, the vector $\mathbf{t} \in \mathbb{R}^M$ is given by the solution to the linear system $\mathbf{T}\mathbf{t} = \mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y})$ with

$$\begin{aligned}
\overline{\mathcal{T}}(\mathbf{y}) &= \mathbf{p}_\gamma \oslash \mathbf{q}_\gamma \\
\mathbf{T} &= (\mathbf{p}_\gamma \mathbf{p}_\gamma^T) \odot \exp(\gamma^2 \mathbf{C}_{\mathbf{e}^y}) \oslash (\mathbf{q}_\gamma \mathbf{q}_\gamma^T - \gamma^2 \mathbf{C}_{\mathbf{z}} \odot \mathbf{C}_{\mathbf{z}}^*) \\
&\quad - (\mathbf{p}_\gamma \mathbf{p}_\gamma^T) \oslash (\mathbf{q}_\gamma \mathbf{q}_\gamma^T) \\
\mathbf{V}_m &= -\frac{\gamma \sigma_x^4 [\mathbf{p}_\gamma]_m}{(\gamma [\mathbf{C}_{\mathbf{z}}]_{m,m} + 1)^2} \mathbf{a}_m \mathbf{a}_m^H, \quad m = 1, \dots, M,
\end{aligned}$$

where we use the following definitions:

$$\begin{aligned}
\mathbf{q}_\gamma &= \gamma \operatorname{diag}(\mathbf{C}_{\mathbf{z}}) + \mathbf{1}_{M \times 1} \\
\mathbf{p}_\gamma &= \exp\big(-\gamma \bar{\mathbf{e}}^y + \gamma^2 \tfrac{1}{2} \operatorname{diag}(\mathbf{C}_{\mathbf{e}^y})\big) \\
\mathbf{C}_{\mathbf{z}} &= \sigma_x^2 \mathbf{A}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z}.
\end{aligned}$$

The spectral estimate $\hat{\mathbf{x}}$ is given by the (scaled) leading eigenvector of $\mathbf{D}_{\mathbf{y}}^{\text{Exp}}$ in (14). Furthermore, the S-MSE of this estimator is given by Theorem 2.

At first sight, the choice of the exponential preprocessing function used in Estimator 2 seems to be arbitrary. We emphasize, however, that this particular function is inspired by the asymptotically-optimal preprocessing function for properly-normalized Gaussian measurement ensembles proposed in (Mondelli & Montanari, 2017) which is given by

$$\mathcal{T}_{\text{opt}}(y) = \frac{y - 1}{y + \sqrt{\delta} - 1}. \tag{15}$$

As it turns out, we can scale, negate, and shift the exponential preprocessing function $\mathcal{T}(y) = \exp(-\gamma y)$ to make it take a similar shape as the function in (15). More concretely, exponential preprocessing as well as $\mathcal{T}_{\text{opt}}(y)$ enables one to attenuate the effect of measurements with large magnitude, which is also the idea underlying the class of orthogonal spectral initializers, as proposed in (Chen et al., 2015; Wang et al., 2017a;b), that perform well in practice.

4. Numerical Results

We now compare the performance of our LSPEs against existing spectral initializers proposed for phase retrieval on synthetic and real image data. All our results use the spectral initializers and experimental setups provided by PhasePack (Chandra et al., 2017).

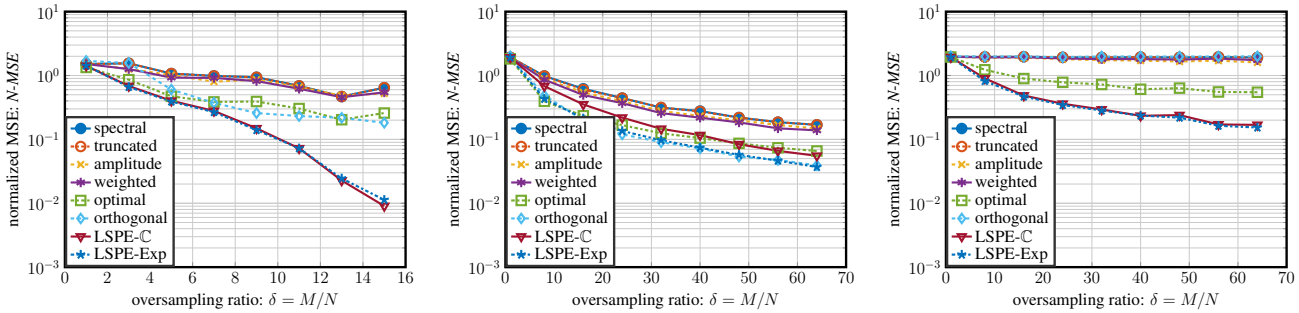
4.1. Impact of Measurement Ensemble

We start by comparing the normalized MSE (N-MSE) defined as (Chandra et al., 2017)

$$\textit{N-MSE} = \frac{\min_{\alpha \in \mathcal{H}} \|\mathbf{x} - \alpha \hat{\mathbf{x}}\|^2}{\|\mathbf{x}\|^2}$$

for a range of spectral initializers on different measurement ensembles. Specifically, we focus on the complex-valued case and consider (i) an i.i.d. Gaussian measurement matrix with signal dimension $N = 16$, (ii) an i.i.d. Gaussian measurement matrix with $N = 256$, and (iii) the structured "transmission matrix" used for image recovery through multiple scattering media as detailed in (Metzler et al., 2017). We vary the oversampling ratio $\delta = M/N$ and compare the N-MSE of the proposed complex-valued LSPEs, LSPE-$\mathbb{C}$ (Estimator 1) and LSPE-Exp (Estimator 2 with $\gamma = 0.001$), to the following spectral initializers: the original spectral initializer (Netrapalli et al., 2013; Candès et al., 2015a) called "spectral," truncated spectral initializer (Chen & Candès, 2015) called "truncated," weighted spectral initializer (Wang et al., 2017b) called "weighted," amplitude spectral initializer (Wang et al., 2017a) called "amplitude," orthogonal spectral initializer (Chen et al., 2015) called "orthogonal," and the asymptotically-optimal spectral initializer (Mondelli & Montanari, 2017) called "optimal." For the following synthetic experiments, we generate the signals to be recovered according to Assumptions 1 and Assumptions 2 for LSPE-$\mathbb{C}$ and LSPE-Exp, respectively.

Figure 1a shows that the proposed LSPEs significantly outperform all existing spectral initializers for small problem dimensions with Gaussian measurements; this improvement is even more pronounced for large oversampling ratios. The reason is that since we randomly generate a low-dimensional sensing matrix, the system will exhibit strong correlations among the measurements that can be exploited by LSPEs. For larger dimensions with Gaussian measurements, we see in Figure 1b that the proposed LSPEs do not provide an advantage over other methods. In fact, only LSPE-Exp is



(a) *Gaussian measurements,* $N = 16$. (b) *Gaussian measurements,* $N = 256$. (c) *Transmission measurements,* $N = 256$.

Figure 1: *Comparison of normalized MSE (N-MSE) as a function of the oversampling ratio* $\delta = M/N$ *for complex-valued phase retrieval with different spectral initializers and with different measurement matrices. The proposed LSPEs perform well on low-dimensional problems, for structured measurement ensembles, or at high oversampling ratios* δ.

able to perform as well as the orthogonal spectral initializer, which achieves the best performance in this scenario. This behavior can be attributed to the facts that (i) for large random matrices there is no particular correlation structure among the measurements to exploit and (ii) ignoring measurements associated to large values in y_m is increasingly important. For structured measurements, as it is the case for the transmission matrix from (Metzler et al., 2017), we see in Figure 1c that LSPEs significantly outperform existing methods that are designed for random measurement ensembles. In this scenario, exponential preprocessing does not improve performance since correlations in the transmission matrix are dominating the performance.

4.2. S-MSE Expressions and Approximation Error

We now validate our theoretical S-MSE expressions in Theorem 2 and Proposition 1, and confirm the accuracy of the EER bound given in Corollary 1. In the following experiment, we set $M = 8N$ and vary the dimension N from 8 to 64. For each pair (M, N), we randomly generate one instance of an i.i.d. circularly symmetric complex Gaussian measurement matrix and average the different errors (S-MSE and EER) over $10,000$ Monte-Carlo trials. We consider a noiseless setting and assume identity preprocessing, i.e., $\mathcal{T}(y) = y$. The signal vectors are generated according to an i.i.d. circularly complex Gaussian random vector. From Figure 2, we see that our analytical S-MSE expressions for the LSPE-$\mathbb{C}$ and spectral initializers match their empirical values. We furthermore see that the empirical EER is only about 6 dB to 10 dB lower than our non-asymptotic upper bound given in Corollary 1.

4.3. Real-World Image Recovery

We finally illustrate the efficacy of LSPEs in a more realistic scenario. In particular, we show results for a real image reconstruction task by using LSPEs and spectral initializers

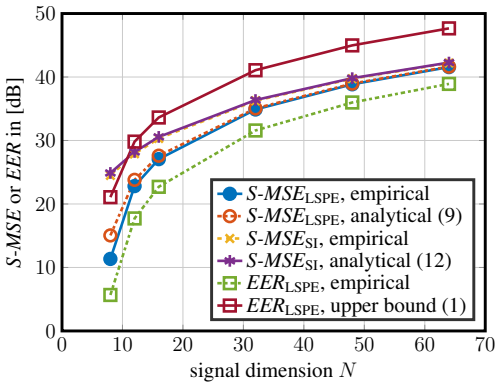
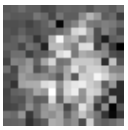


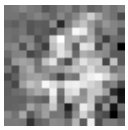
Figure 2: *Comparison of the analytical and empirical spectral MSE (S-MSE) and estimation error (EER) for LSPEs and spectral initializers (SI) at oversampling ratio* $\delta = 8$. *Our analytical expressions in Theorem 2 and Proposition 1 match the empirical S-MSE; the upper bound in Corollary 1 accurately characterizes the empirical EER.*

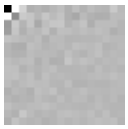
only, i.e., we are *not* using any additional phase retrieval algorithm. Our goal is to recover a 16×16-pixel and a 40×40-pixel image that was captured through a multiple scattering media using the deterministic and highly-structured transmission matrix as detailed in (Metzler et al., 2017). We compare the proposed LSPEs to the same set of spectral initializers as in Section 4.1. The signal priors are as in Assumptions 1 (LSPE-$\mathbb{C}$) and Assumptions 2 (LSPE-Exp).

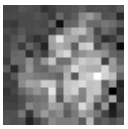
Figures 3 and 4 show the recovered images along with the N-MSE values. The proposed LSPEs (often significantly) outperform all spectral initializers in terms of visual quality as well as the N-MSE. This result confirms the observations made in Figure 1c that LSPEs outperform existing spectral initializers for structured measurement matrices. We note that exponential preprocessing for LSPEs does not noticeably improve the N-MSE (over LSPE-$\mathbb{C}$) in this setting since correlations in the transmission measurement matrix are dominating the recovery performance.

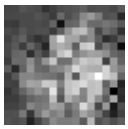

(a) *original*

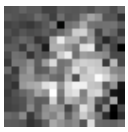

(b) *amplitude* *N-MSE* $= 0.4927$

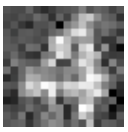

(c) *optimal* *N-MSE* $= 0.4833$


(d) *orthogonal* *N-MSE* $= 0.6850$


(e) *spectral* *N-MSE* $= 0.4764$


(f) *truncated* *N-MSE* $= 0.4764$


(g) *weighted* *N-MSE* $= 0.4797$


(h) *LSPE-*$\mathbb{C}$ *N-MSE* $= 0.3377$

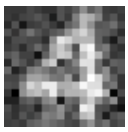
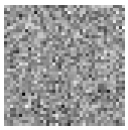
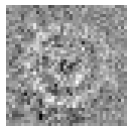
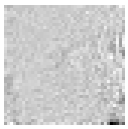

(i) *LSPE-Exp* *N-MSE* $= 0.2928$

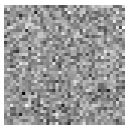
Figure 3: *Recovery of a* 16×16 *image from with* $M = 5N$ *measurements captured through a scattering medium without the use of a phase retrieval algorithm. LSPEs outperform all initializers for structured measurements.*

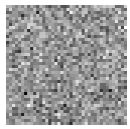

(a) *original*

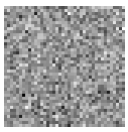

(b) *amplitude* *N-MSE* $= 0.7010$

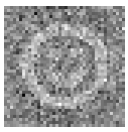

(c) *optimal* *N-MSE* $= 0.5849$


(d) *orthogonal* *N-MSE* $= 0.7028$


(e) *spectral* *N-MSE* $= 0.7016$


(f) *truncated* *N-MSE* $= 0.7020$


(g) *weighted* *N-MSE* $= 0.7013$


(h) *LSPE-*$\mathbb{C}$ *N-MSE* $= 0.4920$

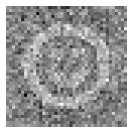

(i) *LSPE-Exp* *N-MSE* $= 0.4896$

Figure 4: *Recovery of a* 40×40 *image from with* $M = 10N$ *measurements captured through a scattering medium without the use of a phase retrieval algorithm. LSPEs outperform all initializers for structured measurements.*

5. Conclusions

We have proposed a novel class of estimators, called *linear spectral estimators* (LSPEs), which are suitable for the recovery of signals from general nonlinear measurement systems. We have developed nonasymptotic and deterministic performance guarantees for LSPEs that provide accurate bounds on the estimation error, especially for structured or low-dimensional measurement systems. To demonstrate the efficacy of LSPEs in practice, we have applied them to complex-valued phase retrieval problems, in which LSPEs can be used to compute accurate signal estimates or initialization vectors for other convex or nonconvex phase retrieval algorithms. We have shown that properly preprocessing the nonlinear measurements can further improve the performance of LSPEs in practical scenarios. Our simulations with synthetic and real data have shown that LSPEs are able to significantly outperform existing spectral initializers, especially for low-dimensional problems, for structured measurement matrices, or for large oversampling ratios.

There are many avenues for future work. First, one could derive LSPEs for the asymptotically-optimal preprocessing function in (15) or for other commonly used functions, which may lead to further performance improvements. Second, the proposed error analysis could be used to generate improved measurement matrices. Third, an exploration of LSPEs for other nonlinearities that arise in machine learning and signal processing applications is left for future work.

Acknowledgments

R. Ghods and C. Studer were supported in part by Xilinx, Inc. and by the US National Science Foundation (NSF) under grants ECCS-1408006, EECS-1740286, CCF-1535897, CCF-1652065, and CNS-1717559. T. Goldstein was supported by the US NSF under grant CCF-1535902, the US ONR under grant N00014-15-1-2676, the DARPA Lifelong Learning Machines program, and the Sloan Foundation.

A. Proof of Theorem 1

The proof proceeds in two steps detailed as follows.

Mean Matrix We first compute the mean matrix $\mathbf{W}_0$. Since (7) is a quadratic form, we can take the derivative in $\widetilde{\mathbf{W}}_0^H$ and set it to zero, i.e.,

$$\frac{d}{d\widetilde{\mathbf{W}}_0^H} \mathbb{E}\left[\left\|\widetilde{\mathbf{W}}_0 + \sum_{m=1}^{M} \mathcal{T}(y_m)\widetilde{\mathbf{W}}_m - \mathbf{x}\mathbf{x}^H\right\|_F^2\right] = 0.$$

Basic matrix calculus yields

$$\widetilde{\mathbf{W}}_0 = \mathbf{K}_\mathbf{x} - \textstyle\sum_{m=1}^{M} \overline{\mathcal{T}}(y_m)\widetilde{\mathbf{W}}_m \tag{16}$$

with $\overline{\mathcal{T}}(y_m) = \mathbb{E}[\mathcal{T}(y_m)]$ and $\mathbf{K}_\mathbf{x} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$.

Linear Estimation Matrix With (16) and the fact that (7) is a quadratic form in the matrices $\mathbf{W}_m$, $m = 1, \ldots, M$,

we take the derivatives in $\mathbf{W}_m^H$ and setting them to zero:

$$\frac{d}{d\widetilde{\mathbf{W}}_m^H}\mathbb{E}\left[\left\|\sum_{m=1}^{M}(\mathcal{T}(y_m)-\overline{\mathcal{T}}(y_m))\widetilde{\mathbf{W}}_m-(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right\|_F^2\right]=0.$$

By interchanging the derivative with expectation and with basic manipulations, we obtain the following set of optimality conditions for $\mathbf{W}_m$ for $m=1,\dots,M$:

$$\begin{aligned}\textstyle\sum_{m'=1}^{M}\widetilde{\mathbf{W}}_{m'}\,\mathbb{E}\big[(\mathcal{T}(y_m)-\overline{\mathcal{T}}(y_m))(\mathcal{T}(y_{m'})-\overline{\mathcal{T}}(y_{m'}))\big]\\ =\mathbb{E}\big[(\mathcal{T}(y_m)-\overline{\mathcal{T}}(y_m))(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\big].\end{aligned}\tag{17}$$

In compact matrix form, the above condition reads

$$(\mathbf{T}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{W}}=\underline{\mathbf{V}},\tag{18}$$

where we used the following shortcuts:

$$\begin{aligned}\mathbf{T}&=\mathbb{E}\big[(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))^T\big]\\ \underline{\mathbf{W}}&=[\widetilde{\mathbf{W}}_1^T,\dots,\widetilde{\mathbf{W}}_m^T,\dots,\widetilde{\mathbf{W}}_M^T]^T\\ \mathbf{V}_m&=\mathbb{E}\big[(\mathcal{T}(y_m)-\overline{\mathcal{T}}(y_m))(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\big],m=1,\dots,M\\ \underline{\mathbf{V}}&=[\mathbf{V}_1^T,\dots,\mathbf{V}_m^T,\dots,\mathbf{V}_M^T]^T.\end{aligned}$$

The condition in (18) can be solved for the estimation matrices in $\underline{\mathbf{W}}$ leading to $\underline{\mathbf{W}}=(\mathbf{T}^{-1}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}$, where we require the matrix $\mathbf{T}$ to be full rank. To obtain the linear spectral estimator matrix, we simplify as

$$\begin{aligned}\mathbf{D}_\mathbf{y}&=\mathbf{K}_\mathbf{x}+((\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))^T\otimes\mathbf{I}_{N\times N})\underline{\mathbf{W}}\\ &=\mathbf{K}_\mathbf{x}+\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m,\end{aligned}$$

where we define the vector $\mathbf{t}=\mathbf{T}^{-1}(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))$.

B. Proof of Theorem 2

To compute the spectral MSE in (5), we simplify

$$\textit{S-MSE}=\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m-(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right\|_F^2\right].$$

We expand this expression into four terms

$$\begin{aligned}&\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m-(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right\|_F^2\right]\\ &=\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right\|_F^2\right] &(19)\\ &\quad+\mathbb{E}\left[\left\|\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x}\right\|_F^2\right]\\ &\quad-\mathbb{E}\left[\mathrm{tr}\left((\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})^H\left(\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right)\right)\right] &(20)\\ &\quad-\mathbb{E}\left[\mathrm{tr}\left(\left(\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right)^H(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right)\right] &(21)\end{aligned}$$

and simplify each expression individually. We start with (19) and use the fact that

$$\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m=((\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))^T\mathbf{T}^{-1}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}$$

and rewrite the quantity within expectation as follows:

$$\begin{aligned}&\underline{\mathbf{V}}^H(\mathbf{T}^{-1}(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))\otimes\mathbf{I}_{N\times N})\\ &\quad\times((\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))^T\mathbf{T}^{-1}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}\\ &=\underline{\mathbf{V}}^H((\mathbf{T}^{-1}(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))\\ &\quad\times(\mathcal{T}(\mathbf{y})-\overline{\mathcal{T}}(\mathbf{y}))^T\mathbf{T}^{-1})\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}.\end{aligned}$$

We now evaluate the expectation which leads to

$$\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right\|_F^2\right]=\mathrm{tr}\Big(\underline{\mathbf{V}}^H(\mathbf{T}^{-1}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}\Big)$$

or, equivalently, to

$$\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right\|_F^2\right]=\displaystyle\sum_{m=1}^{M}\sum_{m'=1}^{M}[\mathbf{T}^{-1}]_{m,m'}\,\mathrm{tr}\big(\mathbf{V}_m^H\mathbf{V}_{m'}\big).$$

We next will simplify (20). Recall that

$$t_m=\textstyle\sum_{m'=1}^{M}[\mathbf{T}^{-1}]_{m,m'}(\mathcal{T}(y_{m'})-\overline{\mathcal{T}}(y_{m'})),$$

which enables us to write (20) as

$$\begin{aligned}&\mathbb{E}\Big[\mathrm{tr}\Big((\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})^H\left(\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right)\Big)\Big]\\ &=\textstyle\sum_{m=1}^{M}\sum_{m'=1}^{M}[\mathbf{T}^{-1}]_{m,m'}\\ &\quad\times\mathrm{tr}\big(\mathbb{E}\big[(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})^H(\mathcal{T}(y_{m'})-\overline{\mathcal{T}}(y_{m'}))\big]\,\mathbf{V}_m\big)\\ &=\textstyle\sum_{m=1}^{M}\sum_{m'=1}^{M}[\mathbf{T}^{-1}]_{m,m'}\,\mathrm{tr}\big(\mathbf{V}_{m'}^H\mathbf{V}_m\big).\end{aligned}\tag{22}$$

Seeing as (21) is the Hermitian conjugate of (20), we have

$$\begin{aligned}&\mathbb{E}\left[\mathrm{tr}\left(\left(\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m\right)^H(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right)\right]\\ &=\textstyle\sum_{m=1}^{M}\sum_{m'=1}^{M}[\mathbf{T}^{-1}]_{m,m'}^{*}\,\mathrm{tr}\big(\mathbf{V}_m^H\mathbf{V}_{m'}\big).\end{aligned}\tag{23}$$

Combining all these terms yield the spectral MSE

$$\begin{aligned}\textit{S-MSE}&=\mathbb{E}\left[\left\|\textstyle\sum_{m=1}^{M}t_m\mathbf{V}_m-(\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x})\right\|_F^2\right]\\ &=\mathbf{C}_{\mathbf{x}\mathbf{x}^H}-\mathrm{tr}\Big(\underline{\mathbf{V}}^H(\mathbf{T}^{-1}\otimes\mathbf{I}_{N\times N})\underline{\mathbf{V}}\Big).\end{aligned}$$

with $\mathbf{C}_{\mathbf{x}\mathbf{x}^H}=\mathbb{E}\left[\left\|\mathbf{x}\mathbf{x}^H-\mathbf{K}_\mathbf{x}\right\|_F^2\right]$.

C. Proof of Corollary 1

We bound the estimation error with the spectral MSE of the LSPE as follows. For a given instance, we have

$$\begin{aligned}\|\hat{\mathbf{x}}\hat{\mathbf{x}}^H-\mathbf{x}\mathbf{x}^H\|_F^2&=\|\hat{\mathbf{x}}\hat{\mathbf{x}}^H-\mathbf{D}_\mathbf{y}+\mathbf{D}_\mathbf{y}-\mathbf{x}\mathbf{x}^H\|_F^2\\ &\overset{(a)}{\leq}2\|\hat{\mathbf{x}}\hat{\mathbf{x}}^H-\mathbf{D}_\mathbf{y}\|_F^2+2\|\mathbf{D}_\mathbf{y}-\mathbf{x}\mathbf{x}^H\|_F^2\\ &\overset{(b)}{\leq}4\|\mathbf{D}_\mathbf{y}-\mathbf{x}\mathbf{x}^H\|_F^2,\end{aligned}$$

where (a) follows from the squared triangle inequality and (b) because $\hat{\mathbf{x}}\hat{\mathbf{x}}^H$ is the best rank-1 approximation of $\mathbf{D}_\mathbf{y}$. Averaging over all instances finally yields

$$\mathbb{E}\big[\|\hat{\mathbf{x}}\hat{\mathbf{x}}^H-\mathbf{x}\mathbf{x}^H\|_F^2\big]\leq 4\,\textit{S-MSE}_{\text{LSPE}}.$$

D. Proof of Proposition 1

Our goal is to first evaluate the S-MSE of the unnormalized spectral initializer in (2)

$$\textit{US-MSE}_{\text{SI}} = \mathbb{E}\left[\left\|\beta \sum_{m=1}^{M} \mathcal{T}(y_m)\mathbf{a}_m\mathbf{a}_m^H - \mathbf{x}\mathbf{x}^H\right\|_F^2\right]$$

and then minimize the resulting expression over the parameter β. The unnormalized spectral MSE can be expanded into the following form:

$$\begin{aligned}
&|\beta|^2 \sum_{m=1}^{M}\sum_{m'=1}^{M} \mathbb{E}[\mathcal{T}(y_m)\mathcal{T}(y_{m'})]\,\mathrm{tr}(\mathbf{a}_m\mathbf{a}_m^H\mathbf{a}_{m'}\mathbf{a}_{m'}^H)\\
&- \beta^* \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\,\mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right]\right)\\
&- \beta \sum_{m=1}^{M} \mathrm{tr}\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\mathcal{T}(y_m)\right]\mathbf{a}_m\mathbf{a}_m^H\right)\\
&+ \mathbb{E}\left[\|\mathbf{x}\mathbf{x}^H\|_F^2\right].
\end{aligned}$$

By using the definitions

$$\begin{aligned}
\widetilde{\mathbf{V}}_m &= \mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right], m = 1,\dots,M,\\
\widetilde{\mathbf{T}} &= \mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right],
\end{aligned}$$

we can simplify the above expression into

$$\begin{aligned}
&|\beta|^2 \sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2 + \mathbb{E}\left[\|\mathbf{x}\mathbf{x}^H\|_F^2\right]\\
&- \beta^* \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right) - \beta \sum_{m=1}^{M} \mathrm{tr}\left(\widetilde{\mathbf{V}}_m^H\mathbf{a}_m\mathbf{a}_m^H\right). \quad (24)
\end{aligned}$$

We can now find the optimal parameter for β by taking the derivative with respect to β^* and setting the expression to zero. The resulting optimal scaling parameter is given by

$$\hat{\beta} = \frac{\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2}.$$

We now plug in $\hat{\beta}$ into the expression (24), which yields

$$\begin{aligned}
\textit{S-MSE}_{\text{SI}} = &\left|\frac{\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2}\right|^2 \quad (25)\\
&\times \sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2\\
&- \frac{\sum_{m=1}^{M} \mathrm{tr}\left(\widetilde{\mathbf{V}}_m^H\mathbf{a}_m\mathbf{a}_m^H\right)}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}^*_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2} \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)\\
&- \frac{\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2} \sum_{m=1}^{M} \mathrm{tr}\left(\widetilde{\mathbf{V}}_m^H\mathbf{a}_m\mathbf{a}_m^H\right)\\
&+ \mathbb{E}\left[\|\mathbf{x}\mathbf{x}^H\|_F^2\right].
\end{aligned}$$

This expression can be simplified further to obtain:

$$\begin{aligned}
\textit{S-MSE}_{\text{SI}} =&\frac{\left|\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)\right|^2}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}^*_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2}\\
&- \frac{\left|\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)\right|^2}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}^*_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2}\\
&- \frac{\left|\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{a}_m\mathbf{a}_m^H\widetilde{\mathbf{V}}_m\right)\right|^2}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2}\\
&+ \mathbb{E}\left[\|\mathbf{x}\mathbf{x}^H\|_F^2\right],\\
=&\, R_{\mathbf{x}\mathbf{x}^H} - \frac{\left|\sum_{m=1}^{M} \mathbf{a}_m^H\widetilde{\mathbf{V}}_m\mathbf{a}_m\right|^2}{\sum_{m=1}^{M}\sum_{m'=1}^{M} \widetilde{T}_{m,m'}|\mathbf{a}_m^H\mathbf{a}_{m'}|^2},
\end{aligned}$$

which is what we wanted to show in (12).

E. Real-Valued Phase Retrieval

We now focus on the case where the signal vector $\mathbf{x}$ to be recovered and the measurement matrix $\mathbf{A}$ are both real-valued. We derive the LSPE by using the following assumptions, which are reasonable for phase retrieval problems.

Assumptions 3. *Let $\mathcal{H} = \mathbb{R}$. Assume square measurements $f(z) = z^2$ and the identity preprocessing function $\mathcal{T}(y) = y$. Assume that the signal vector $\mathbf{x} \in \mathbb{R}^N$ is i.i.d. zero-mean Gaussian distributed with covariance matrix $\mathbf{C}_\mathbf{x} = \sigma_x^2\mathbf{I}_N$, i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_{N\times 1}, \sigma_x^2\mathbf{I}_N)$; the parameter σ_x^2 denotes the signal variance. Assume that the signal noise vector $\mathbf{e}^z$ is zero-mean Gaussian with covariance matrix $\mathbf{C}_{\mathbf{e}^z}$, i.e., $\mathbf{e}^z \sim \mathcal{N}(\mathbf{0}_{M\times 1}, \mathbf{C}_{\mathbf{e}^z})$, and the measurement noise vector $\mathbf{e}^y$ is Gaussian with mean $\bar{\mathbf{e}}^y$ and covariance matrix $\mathbf{C}_{\mathbf{e}^y}$, i.e., $\mathbf{e}^y \sim \mathcal{N}(\bar{\mathbf{e}}^y, \mathbf{C}_{\mathbf{e}^y})$. Furthermore assume that $\mathbf{x}$, $\mathbf{e}^z$, and $\mathbf{e}^y$ are independent.*

Under these assumptions, we can derive the following LSPE which we call LSPE-$\mathbb{R}$; the detailed derivations of this spectral estimator are given in Appendix F.

Estimator 3 (LSPE-$\mathbb{R}$)**.** *Let Assumptions 3 hold. Then, the spectral estimation matrix is given by*

$$\mathbf{D}_\mathbf{y}^{\mathbb{R}} = \mathbf{K}_\mathbf{x} + \sum_{m=1}^{M} t_m\mathbf{V}_m, \quad (26)$$

where $\mathbf{K}_\mathbf{x} = \sigma_x^2\mathbf{I}_N$, the vector $\mathbf{t} \in \mathbb{R}^M$ is given by the solution to the linear system $\mathbf{T}\mathbf{t} = \mathbf{y} - \overline{\mathbf{y}}$ with

$$\overline{\mathbf{y}} = \mathrm{diag}(\mathbf{C}_\mathbf{z}) + \bar{\mathbf{e}}^y$$

$$\mathbf{C_z} = \sigma_x^2 \mathbf{A}\mathbf{A}^T + \mathbf{C}_{\mathbf{e}^z}$$
$$\mathbf{T} = 2\mathbf{C_z} \odot \mathbf{C_z} + \mathbf{C}_{\mathbf{e}^y}$$

and $\mathbf{V}_m = 2\sigma_x^4 \mathbf{a}_m \mathbf{a}_m^T$, $m = 1, \dots, M$. *The spectral estimate* $\hat{\mathbf{x}}$ *is given by the (scaled) leading eigenvector of* $\mathbf{D}_{\mathbf{y}}^{\mathbb{R}}$ *in (26). Furthermore, the S-MSE is given by Theorem 2.*

F. Derivation of Estimator 3

We now use Theorem 1 to derive Estimator 3 under Assumptions 3. To this end, we require the three quantities: $\overline{\mathcal{T}}(\mathbf{y})$, $\mathbf{T}$, and $\mathbf{V}_m$, $m = 1, \dots, M$, which we derive separately.

Computing $\overline{\mathcal{T}}(\mathbf{y})$ To compute the real-valued vector

$$\overline{\mathcal{T}}(\mathbf{y}) = \mathbb{E}[\mathcal{T}(\mathbf{y})], \tag{27}$$

we need the following result on the bivariate folded normal distribution developed in (Kan & Robotti, 2017, Sec. 3.1).

Lemma 1. *Let* $[u_1, u_2] \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *be a pair of real-valued jointly Gaussian random variables with covariance matrix*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2}^2 \\ \sigma_{1,2}^2 & \sigma_2^2 \end{bmatrix}.$$

Then, for $m = 1, 2$, *the pair of random variables* (ν_1, ν_2) *with* $\nu_1 = u_1^2$ *and* $\nu_2 = u_2^2$ *follows the bivariate folded normal distribution with the following (centered) moments:*

$$\begin{aligned} \bar{\nu}_m &= \mathbb{E}\left[u_m^2\right] = \sigma_m^2 + \mu_m^2 \\ [\mathbf{C}_{\boldsymbol{\nu}}]_{1,2} &= \mathbb{E}[(\nu_1 - \bar{\nu_1})(\nu_2 - \bar{\nu_2})] \\ &= 4\mu_1\mu_2\sigma_{1,2}^2 + 2\sigma_{1,2}^4 \\ [\mathbf{C}_{\boldsymbol{\nu}}]_{1,1} &= \mathbb{E}\left[(\nu_1 - \bar{\nu_1})^2\right] = 2\sigma_1^4 + 4\mu_1^2\sigma_1^2. \end{aligned}$$

Let $\bar{\mathbf{z}} = \mathbb{E}[\mathbf{z}]$ denote the mean vector and $\mathbf{C_z} = \mathbf{A}\mathbf{C}_{\mathbf{x}}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z} = \sigma_x^2 \mathbf{A}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z}$ the covariance matrix of the "phased" measurements $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}^z$. Then, by defining $\sigma_m^2 = [\mathbf{C_z}]_{m,m}$, we can compute the mth entry $\overline{\mathcal{T}}(y_m)$ using Lemma 1 as follows:

$$\overline{\mathcal{T}}(y_m) = \bar{y}_m = \mathbb{E}\left[|z_m|^2 + n_m^y\right] = \sigma_m^2 + \bar{e}_m^y. \tag{28}$$

Hence, in compact vector notation we have

$$\overline{\mathcal{T}}(\mathbf{y}) = \bar{\mathbf{y}} = \operatorname{diag}(\mathbf{C_z}) + \bar{\mathbf{e}}^y. \tag{29}$$

Computing $\mathbf{T}$ To compute the real-valued matrix

$$\begin{aligned} \mathbf{T} &= \mathbb{E}\left[(\mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y}))(\mathcal{T}(\mathbf{y}) - \overline{\mathcal{T}}(\mathbf{y}))^T\right] \\ &= \mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right] - \overline{\mathcal{T}}(\mathbf{y})\overline{\mathcal{T}}(\mathbf{y})^T, \end{aligned} \tag{30}$$

we only need to compute the matrix $\mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right]$ as the vector $\overline{\mathcal{T}}(\mathbf{y})$ was computed in (29). We compute this matrix entry-wise as

$$\begin{aligned} T_{m,m'} &= \mathbb{E}\left[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))(\mathcal{T}(y_{m'}) - \overline{\mathcal{T}}(y_{m'}))\right] \\ &= \mathbb{E}[y_m y_{m'}^*] - \bar{y}_m \bar{y}_{m'}^* \\ &\overset{(a)}{=} \mathbb{E}\left[(|z_m|^2 + e_m^y)(|z_{m'}|^2 + e_{m'}^y)\right] \\ &\quad - (\sigma_m^2 + \bar{e}_m^y)(\sigma_{m'}^2 + \bar{e}_{m'}^y) \\ &= \mathbb{E}\left[|z_m|^2 |z_{m'}|^2\right] - \sigma_{m'}^2\sigma_m^2 + [\mathbf{C}_{\mathbf{e}^y}]_{m,m'}, \end{aligned}$$

where (a) follows from (28). The only unknown term in the above expression is $\mathbb{E}\left[|z_m|^2|z_{m'}|^2\right]$. This term is the second moment of the random vector $[|z_m|^2, |z_{m'}|^2]$, which follows a bivariate folded normal distribution. For $m \neq m'$, Lemma 1 yields

$$\mathbb{E}\left[|z_m|^2|z_{m'}|^2\right] = \sigma_m^2\sigma_{m'}^2 + 2\sigma_{m,m'}^4$$

with $\sigma_{m,m'}^2 = [\mathbf{C_z}]_{m,m'}$. For $m = m'$, Lemma 1 yields

$$\mathbb{E}\left[|y_m|^2\right] = \mathbb{E}\left[|z_m|^4\right] = 3\sigma_m^4.$$

Hence, we have

$$T_{m,m'} = [\mathbf{C}_{\mathbf{e}^y}]_{m,m'} + \begin{cases} 2\sigma_{m,m'}^4 & \text{if } m \neq m' \\ 2\sigma_m^4 & \text{if } m = m', \end{cases}$$

which can be written in compact matrix form as

$$\mathbf{T} = 2\mathbf{C_z} \odot \mathbf{C_z} + \mathbf{C}_{\mathbf{e}^y}.$$

Computing $\mathbf{V}_m$ To compute the matrices

$$\begin{aligned} \mathbf{V}_m &= \mathbb{E}\left[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))(\mathbf{x}\mathbf{x}^H - \mathbf{K}_{\mathbf{x}})\right] \\ &= \mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right] - \overline{\mathcal{T}}(y_m)\mathbf{K}_{\mathbf{x}} \end{aligned} \tag{31}$$

for $m = 1, \dots, M$, we only need to compute the complex-valued matrix $\mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right]$ as the two other quantities $\mathbf{K}_{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]$ and $\overline{\mathcal{T}}(y_m)$ are known. We compute this matrix entry-wise as

$$\begin{aligned} [\mathbf{V}_m]_{n,n'} &= \mathbb{E}\left[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))x_n x_{n'}^*\right] \\ &= \mathbb{E}[y_m x_n x_{n'}^*] - \bar{y}_m [\mathbf{C}_{\mathbf{x}}]_{n,n'}. \end{aligned}$$

Since $\bar{y}_m$ is known from (28), we focus on computing

$$\begin{aligned} &\mathbb{E}[y_m x_n x_{n'}^*] \\ &= \mathbb{E}\left[\left(\left(\sum_{j=1}^N A_{m,j}^* x_j^* + e_m^z\right)\right.\right. \\ &\quad \left.\left. \times \left(\sum_{j'=1}^N A_{m,j'} x_{j'} + e_m^z\right) + e_m^y\right) x_n x_{n'}^*\right] \\ &= \mathbb{E}\left[\left(\sum_{j=1}^N A_{m,j}^* x_j^* \sum_{j'=1}^N A_{m,j'} x_{j'}\right) x_n x_{n'}^*\right] \\ &\quad + \mathbb{E}\left[|e_m^z|^2 x_n x_{n'}^*\right] + \mathbb{E}[e_m^y x_n x_{n'}^*] \\ &= \sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\left[x_j^* x_{j'} x_n x_{n'}^*\right] \end{aligned} \tag{32}$$

$$+ ([\mathbf{C}_{\mathbf{e}^z}]_{m,m} + \bar{e}^y_m)[\mathbf{C}_\mathbf{x}]_{n,n'}.$$

The only unknown in the above expression is the double summation in (32). Since we assumed that the entries of the signal vector $\mathbf{x}$ are i.i.d., most of the terms in this summation are zero. For $n \neq n'$, there are only two nonzero terms, corresponding to the cases of $(j, j') = (n, n')$ and $(j, j') = (n', n)$. Thus, for $n \neq n'$ we have

$$\begin{aligned}\sum_{j=1}^{N} A^*_{m,j} \sum_{j'=1}^{N} A_{m,j'} \mathbb{E}\left[x^*_j x_{j'} x_n x^*_{n'}\right] &= 2A^*_{m,n} A_{m,n'} \mathbb{E}\left[|x_n|^2 |x_{n'}|^2\right] \\ &\overset{(b)}{=} 2A^*_{m,n} A_{m,n'} [\mathbf{C}_\mathbf{x}]_{n,n} [\mathbf{C}_\mathbf{x}]_{n',n'}, \end{aligned} \tag{33}$$

where (b) follows from Lemma 1. For $n = n'$, we have

$$\begin{aligned}&\sum_{j=1}^{N} A^*_{m,j} \sum_{j'=1}^{N} A_{m,j'} \mathbb{E}\left[x^*_j x_{j'} x_n x^*_n\right] \\ &= |A_{m,n}|^2 \mathbb{E}\left[|x_n|^4\right] + \sum_{j\neq n, j=1}^{N} |A_{m,j}|^2 \mathbb{E}\left[|x_j|^2 |x_n|^2\right] \\ &\overset{(c)}{=} 3|A_{m,n}|^2 [\mathbf{C}_\mathbf{x}]^2_{n,n} + \sum_{j\neq n, j=1}^{N} |A_{m,j}|^2 [\mathbf{C}_\mathbf{x}]_{j,j} [\mathbf{C}_\mathbf{x}]_{n,n} \\ &= 2|A_{m,n}|^2 [\mathbf{C}_\mathbf{x}]^2_{n,n} + \sum_{j=1}^{N} |A_{m,j}|^2 [\mathbf{C}_\mathbf{x}]_{j,j} [\mathbf{C}_\mathbf{x}]_{n,n}.\end{aligned}$$

As for (33), (c) follows from Lemma 1. By combining the above results, we have

$$\begin{aligned}\mathbf{V}_m =& 2\mathbf{C}^H_\mathbf{x} \mathbf{a}_m \mathbf{a}^H_m \mathbf{C}_\mathbf{x} + (\mathbf{a}^H_m \mathbf{C}_\mathbf{x} \mathbf{a}_m)(\mathbf{C}^H_\mathbf{x} \odot \mathbf{I}) \\ &+ ([\mathbf{C}_{\mathbf{e}^z}]_{m,m} - \sigma^2_m)\mathbf{C}_\mathbf{x} = 2\sigma^4_x \mathbf{a}_m \mathbf{a}^H_m,\end{aligned}$$

where $\mathbf{a}^H_m$ denotes the mth row of the matrix $\mathbf{A}$.

G. Derivation of Estimator 1

We now use Theorem 1 to derive Estimator 1 under Assumptions 1. To this end, we require the three quantities: $\overline{\mathcal{T}}(\mathbf{y})$, $\mathbf{T}$, and $\mathbf{V}_m$, $m = 1, \ldots, M$, which we derive separately.

Computing $\overline{\mathcal{T}}(\mathbf{y})$ To compute the real-valued vector $\overline{\mathcal{T}}(\mathbf{y}) = \bar{\mathbf{y}}$ in (27), we need the following definitions. Let $\bar{\mathbf{z}} = \mathbb{E}[\mathbf{z}]$ denote the mean vector and $\mathbf{C}_\mathbf{z} = \mathbf{A}\mathbf{C}_\mathbf{x}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z} = \sigma^2_x \mathbf{A}\mathbf{A}^H + \mathbf{C}_{\mathbf{e}^z}$ the covariance matrix of the "phased" measurements $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}^z$. Then, using Lemma 1 with the definitions $\bar{\mathbf{z}}$ and $\mathbf{C}_\mathbf{z}$, we have

$$\begin{aligned}\bar{y}_m &= \mathbb{E}\left[|z_m|^2 + \bar{e}^y_m\right] = \mathbb{E}\left[|z_{m,\mathcal{R}}|^2 + |z_{m,\mathcal{I}}|^2 + \bar{e}^y_m\right] \\ &= \sigma^2_m + \bar{e}^y_m,\end{aligned} \tag{34}$$

where we have used the definition $\sigma^2_m = [\mathbf{C}_\mathbf{z}]_{m,m}$. Hence, in compact vector notation we have

$$\overline{\mathcal{T}}(\mathbf{y}) = \bar{\mathbf{y}} = \operatorname{diag}(\mathbf{C}_\mathbf{z}) + \bar{\mathbf{e}}^y.$$

Computing $\mathbf{T}$ To compute the real-valued matrix $\mathbf{T}$ in (30), we will frequently use the following result. Since the vector $\mathbf{z}$ is a complex circularly-symmetric jointly Gaussian vector, we can extract the covariance matrices of the real and imaginary parts separately as:

$$\mathbb{E}\left[\mathbf{z}_\mathcal{I} \mathbf{z}^H_\mathcal{I}\right] \overset{(a)}{=} \mathbb{E}\left[\mathbf{z}_\mathcal{R} \mathbf{z}^H_\mathcal{R}\right] = \frac{1}{2}\Re\{\mathbf{C}_\mathbf{z}\} = \frac{1}{2}\mathbf{C}_{\mathbf{z},\mathcal{R}} \tag{35}$$

$$\mathbb{E}\left[\mathbf{z}_\mathcal{R} \mathbf{z}^H_\mathcal{I}\right] = -\mathbb{E}\left[\mathbf{z}_\mathcal{I} \mathbf{z}^H_\mathcal{R}\right] = \frac{1}{2}\Im\{\mathbf{C}_\mathbf{z}\} = \frac{1}{2}\mathbf{C}_{\mathbf{z},\mathcal{I}}, \tag{36}$$

where (a) follows from circular symmetry of the random vector $\mathbf{x}$. We are now ready to compute the individual entries of $\mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right]$ as

$$\begin{aligned}T_{m,m'} &= \mathbb{E}\left[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))(\mathcal{T}(y_{m'}) - \overline{\mathcal{T}}(y_{m'}))\right] \\ &= \mathbb{E}[(y_m - \bar{y}_m)(y_{m'} - \bar{y}_{m'})^*] \\ &= \mathbb{E}[y_m y^*_{m'}] - \bar{y}_m \bar{y}^*_{m'}.\end{aligned}$$

The quantity $\bar{y}_m$ is given by (34). Hence, we now compute

$$\begin{aligned}&\mathbb{E}[y_m y^*_{m'}] \\ &= \mathbb{E}\left[(|z_m|^2 + e^y_m)(|z'_m|^2 + e^y_{m'})^*\right] \\ &= \mathbb{E}\left[\left(|z_{m,\mathcal{R}}|^2 + |z_{m,\mathcal{I}}|^2\right)\left(|z_{m',\mathcal{R}}|^2 + |z_{m',\mathcal{I}}|^2\right)\right] \\ &\quad + [\mathbf{C}_{\mathbf{e}^y}]_{m,m} \\ &= 2\,\mathbb{E}\left[|z_{m,\mathcal{R}}|^2 |z_{m',\mathcal{R}}|^2\right] + 2\,\mathbb{E}\left[|z_{m,\mathcal{R}}|^2 |z_{m',\mathcal{I}}|^2\right] \\ &\quad + [\mathbf{C}_{\mathbf{e}^y}]_{m,m}.\end{aligned}$$

The first two terms above are a second moment of the variables $[|z_{m,\mathcal{R}}|^2, |z_{m',\mathcal{R}}|^2]$ and $[|z_{m,\mathcal{R}}|^2, |z_{m',\mathcal{I}}|^2]$, which follow a bivariate folded normal distributions. We first focus on $[|z_{m,\mathcal{R}}|^2, |z_{m',\mathcal{R}}|^2]$. With Lemma 1, we can calculate the moments using the covariance $\mathbb{E}\left[\mathbf{z}_\mathcal{R} \mathbf{z}^H_\mathcal{R}\right]$ given in (35). To this end, define $\sigma^2_{m,m',\mathcal{R}} = [\mathbf{C}_{\mathbf{z},\mathcal{R}}]_{m,m'}$ and $\sigma^2_{m,\mathcal{R}} = [\mathbf{C}_{\mathbf{z},\mathcal{R}}]_{m,m}$. Thus, we have

$$\mathbb{E}\left[|z_{m,\mathcal{R}}|^2 |z_{m',\mathcal{R}}|^2\right] = \begin{cases} \frac{\sigma^2_{m,\mathcal{R}}}{2}\frac{\sigma^2_{m',\mathcal{R}}}{2} + \frac{\sigma^4_{m,m',\mathcal{R}}}{2}, & m \neq m' \\ 3\frac{\sigma^4_{m,\mathcal{R}}}{4}, & m = m'. \end{cases}$$

Analogously, we can compute $\mathbb{E}\left[\mathbf{z}_\mathcal{R} \mathbf{z}^H_\mathcal{I}\right]$ in (36) from the covariance matrix of $[|z_{m,\mathcal{R}}|^2, |z_{m',\mathcal{I}}|^2]$, with $\sigma^2_{m,m',\mathcal{I}} = [\mathbf{C}_{\mathbf{z},\mathcal{I}}]_{m,m'}$ and noting that $\sigma^2_{m,\mathcal{I}} = [\mathbf{C}_{\mathbf{z},\mathcal{I}}]_{m,m} = 0$ as

$$\mathbb{E}\left[|z_{m,\mathcal{R}}|^2 |z_{m',\mathcal{I}}|^2\right] = \begin{cases} \frac{\sigma^2_{m,\mathcal{R}}}{2}\frac{\sigma^2_{m',\mathcal{R}}}{2} + 2\frac{\sigma^4_{m,m',\mathcal{I}}}{4}, & m \neq m' \\ 3\frac{\sigma^4_{m,\mathcal{R}}}{4}, & m = m'. \end{cases}$$

By combining the above results, we have

$$\begin{aligned}T_{m,m'} =& \begin{cases} \sigma^2_{m,\mathcal{R}}\sigma^2_{m',\mathcal{R}} + \sigma^4_{m,m',\mathcal{R}} + \sigma^4_{m,m',\mathcal{I}}, & m \neq m' \\ 2\,\sigma^4_{m,\mathcal{R}}, & m = m', \end{cases} \\ &+ [\mathbf{C}_{\mathbf{e}^y}]_{m,m'} - \bar{y}_m \bar{y}^*_{m'} \\ =& [\mathbf{C}_{\mathbf{e}^y}]_{m,m'} + \begin{cases} \sigma^4_{m,m',\mathcal{R}} + \sigma^4_{m,m',\mathcal{I}}, & m \neq m' \\ \sigma^4_{m,\mathcal{R}}, & m = m', \end{cases}\end{aligned}$$

which can be written in matrix form as

$$\mathbf{T} = \mathbf{C}_\mathbf{z} \odot \mathbf{C}^*_\mathbf{z} + \mathbf{C}_{\mathbf{e}^y}.$$

Computing $\mathbf{V}_m$ To compute the matrices $\mathbf{V}_m$, $m = 1, \dots, M$, in (31), we need the complex-valued matrix $\mathbb{E}\big[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\big]$. We compute this matrix entry-wise as

$$\begin{aligned}[\mathbf{V}_m]_{n,n'} &= \mathbb{E}\big[(\mathcal{T}(y_m) - \overline{\mathcal{T}}(y_m))x_n x_{n'}^*\big] \\ &= \mathbb{E}[y_m x_n x_{n'}^*] - \bar{y}_m[\mathbf{C}_\mathbf{x}]_{n,n'}.\end{aligned}$$

Since $\bar{y}_m$ is given by (34), we only need to compute

$$\begin{aligned}&\mathbb{E}[y_m x_n x_{n'}^*] \\ &= \mathbb{E}\Bigg[\Bigg(\Big(\sum_{j=1}^N A_{m,j}^* x_j^* + e_m^{z*}\Big) \\ &\qquad \times \Big(\sum_{j'=1}^N A_{m,j'} x_{j'} + e_m^z\Big) + e_m^y\Bigg) x_n x_{n'}^*\Bigg] \\ &= \sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\big[x_j^* x_{j'} x_n x_{n'}^*\big] \\ &\quad + \mathbb{E}\big[|e_m^z|^2 x_n x_{n'}^*\big] + \mathbb{E}[e_m^y x_n x_{n'}^*] \\ &= \sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\big[x_j^* x_{j'} x_n x_{n'}^*\big] \qquad (37) \\ &\quad + ([\mathbf{C}_{\mathbf{e}^z}]_{m,m} + \bar{e}_m^y)[\mathbf{C}_\mathbf{x}]_{n,n'}.\end{aligned}$$

We will first simplify the term

$$\sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\big[x_j^* x_{j'} x_n x_{n'}^*\big].$$

Since we assumed that the signal vector $\mathbf{x}$ has i.i.d. zero-mean entries, most of the terms in this summation are zero. For $n \neq n'$, there is only one non-zero term for $(j, j') = (n, n')$. Thus, for $n \neq n'$ we have

$$\begin{aligned}&\sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\big[x_j^* x_{j'} x_n x_{n'}^*\big] \\ &\qquad = A_{m,n}^* A_{m,n'} [\mathbf{C}_\mathbf{x}]_{n,n} [\mathbf{C}_\mathbf{x}]_{n',n'},\end{aligned}$$

since the term that corresponds to $(j, j') = (n', n)$, i.e. $A_{m,n'}^* A_{m,n}\, \mathbb{E}[x_{n'}^* x_{n'}^*]\, \mathbb{E}[x_n x_n]$, is zero.

Next, for $n = n'$, we have

$$\begin{aligned}&\sum_{j=1}^N A_{m,j}^* \sum_{j'=1}^N A_{m,j'}\, \mathbb{E}\big[x_j^* x_{j'} x_n x_{n'}^*\big] \\ &= |A_{m,n}|^2\, \mathbb{E}\big[|x_n|^4\big] + \sum_{j\neq k, j=1}^N |A_{m,j}|^2\, \mathbb{E}\big[|x_j|^2 |x_n|^2\big] \\ &= |A_{m,n}|^2\, \mathbb{E}\big[|x_{n,\mathcal{R}}|^4\big] + |A_{m,n}|^2\, \mathbb{E}\big[|x_{n,\mathcal{I}}|^4\big] \\ &\quad + 2|A_{m,n}|^2\, \mathbb{E}\big[|x_{n,\mathcal{R}}|^2 |x_{n,\mathcal{I}}|^2\big] \\ &\quad + \sum_{j\neq n, j=1}^N |A_{m,j}|^2 \\ &\quad \times \mathbb{E}\big[(|x_{j,\mathcal{R}}|^2 + |x_{j,\mathcal{I}}|^2)(|x_{n,\mathcal{R}}|^2 + |x_{n,\mathcal{I}}|^2)\big] \\ &\overset{(a)}{=} 2|A_{m,n}|^2\, \mathbb{E}\big[|x_{n,\mathcal{R}}|^4\big] \\ &\quad + 2\sum_{j=1}^N |A_{m,j}|^2\, \mathbb{E}\big[|x_{j,\mathcal{R}}|^2 |x_{n,\mathcal{I}}|^2\big] \\ &\quad + 2\sum_{j\neq n, j=1}^N |A_{m,j}|^2\, \mathbb{E}\big[|x_{j,\mathcal{R}}|^2 |x_{n,\mathcal{R}}|^2\big] \\ &\overset{(b)}{=} |A_{m,n}|^2 [\mathbf{C}_\mathbf{x}]_{n,n}^2 + \sum_{j=1}^N |A_{m,j}|^2 [\mathbf{C}_\mathbf{x}]_{j,j} [\mathbf{C}_\mathbf{x}]_{n,n},\end{aligned}$$

where (a) follows from circular symmetry of $\mathbf{x}$ and (b) from Lemma 1. By combining the above results, we have

$$\begin{aligned}\mathbf{V}_m &= \mathbf{C}_\mathbf{x}^H \mathbf{a}_m \mathbf{a}_m^H \mathbf{C}_\mathbf{x} + (\mathbf{a}_m^H \mathbf{C}_\mathbf{x} \mathbf{a}_m)(\mathbf{C}_\mathbf{x}^H \odot \mathbf{I}) \\ &\quad + ([\mathbf{C}_{\mathbf{e}^z}]_{m,m} - \sigma_m^2)\mathbf{C}_\mathbf{x} = \sigma_x^4 \mathbf{a}_m \mathbf{a}_m^H.\end{aligned}$$

H. Derivation of Estimator 2

We now use Theorem 1 to derive Estimator 2 under Assumptions 2. To this end, we require the three quantities: $\overline{\mathcal{T}}(\mathbf{y})$, $\mathbf{T}$, and $\mathbf{V}_m$, $m = 1, \dots, M$, which we derive separately.

Computing $\overline{\mathcal{T}}(\mathbf{y})$ To derive an expression for $\overline{\mathcal{T}}(\mathbf{y})$ in (27), we need the following two results.

Lemma 2. *Let $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}_{M\times 1}, \Sigma)$ be a complex-valued circularly-symmetric jointly Gaussian random vector with positive definite covariance matrix $\Sigma \in \mathbb{C}^{M\times M}$. Then, for the random variable $\nu = \exp(-\mathbf{u}^H \mathbf{G}\mathbf{u})$ with positive definite $\mathbf{G} \in \mathbb{C}^{M\times M}$ and $\mathbf{G} + \Sigma^{-1}$ positive definite, we have the following result:*

$$\mathbb{E}[\nu] = \frac{1}{|\mathbf{G}\Sigma + \mathbf{I}_M|}.$$

Proof. We first expand the expected value into

$$\begin{aligned}&\mathbb{E}[\nu] = \mathbb{E}\big[\exp(-\mathbf{u}^H \mathbf{G}\mathbf{u})\big] = \\ &\int_{\mathbb{C}^M} \exp(-\mathbf{u}^H \mathbf{G}\mathbf{u}) \frac{1}{\pi^M |\Sigma|} \exp(-\mathbf{u}^H \Sigma^{-1}\mathbf{u}) \mathrm{d}\mathbf{u},\end{aligned}$$

where $|\Sigma| > 0$ is the determinant of Σ. We can now simplify the above expression as follows:

$$\begin{aligned}&\int_{\mathbb{C}^M} \exp(-\mathbf{u}^H \mathbf{G}\mathbf{u}) \frac{1}{\pi^M |\Sigma|} \exp(-\mathbf{u}^H \Sigma^{-1}\mathbf{u}) \mathrm{d}\mathbf{u} \\ &= \int_{\mathbb{C}^M} \frac{1}{\pi^M |\Sigma|} \exp\big(-\mathbf{u}^H (\mathbf{G} + \Sigma^{-1})\mathbf{u}\big) \mathrm{d}\mathbf{u} \\ &= \frac{\pi^M |(\mathbf{G} + \Sigma^{-1})^{-1}|}{\pi^M |\Sigma|} \frac{1}{\pi^M |(\mathbf{G} + \Sigma^{-1})^{-1}|} \\ &\quad \times \int_{\mathbb{C}^M} \exp\big(-\mathbf{u}^H (\mathbf{G} + \Sigma^{-1})\mathbf{u}\big) \mathrm{d}\mathbf{u}\end{aligned}$$

$$= \frac{|(\mathbf{G}+\Sigma^{-1})^{-1}|}{|\Sigma|} = \frac{1}{|\mathbf{G}+\Sigma^{-1}||\Sigma|} = \frac{1}{|\mathbf{G}\Sigma+\mathbf{I}|},$$

where we also required that $\mathbf{G}+\Sigma^{-1}$ is positive definite. ■

Lemma 3. *Let $\mathbf{u} \sim \mathcal{N}(\bar{\mathbf{u}}, \Sigma)$ be a real-valued Gaussian random vector with mean $\bar{\mathbf{u}}$ and covariance Σ, and $\boldsymbol{\gamma} \in \mathbb{R}^N$ be a given vector. Then, we have*

$$\mathbb{E}\left[\exp(-\boldsymbol{\gamma}^T\mathbf{u})\right] = \exp\left(-\boldsymbol{\gamma}^T\bar{\mathbf{u}} + \tfrac{1}{2}\boldsymbol{\gamma}^T\Sigma\boldsymbol{\gamma}\right).$$

Proof. The proof is an immediate consequence of the moment generating function of a Gaussian random vector. ■

By considering Lemma 2 and Lemma 3 for scalar random variables, the mth entry of the preprocessed phaseless measurement is given by

$$\begin{aligned}\overline{\mathcal{T}}(y_m) &= \mathbb{E}[\mathcal{T}(y_m)] = \mathbb{E}\left[\exp(-\gamma|z_m|^2 - \gamma[\mathbf{e}^y]_m)\right] \\ &= \frac{1}{\gamma[\mathbf{C}_\mathbf{z}]_{m,m}+1}\exp\left(-\gamma[\bar{\mathbf{e}}^y]_m + \tfrac{1}{2}\gamma^2[\mathbf{C}_{\mathbf{e}^y}]_{m,m}\right).\end{aligned}$$

We define the following auxiliary vectors

$$\mathbf{q}_\gamma = \gamma\,\mathrm{diag}(\mathbf{C}_\mathbf{z}) + \mathbf{1}_{M\times 1} \tag{38}$$

$$\mathbf{p}_\gamma = \exp\left(-\gamma\bar{\mathbf{e}}^y + \tfrac{1}{2}\gamma^2\,\mathrm{diag}(\mathbf{C}_{\mathbf{e}^y})\right), \tag{39}$$

which enable us to rewrite the above expression in compact vector form as

$$\overline{\mathcal{T}}(\mathbf{y}) = \mathbf{p}_\gamma \oslash \mathbf{q}_\gamma.$$

Computing $\mathbf{T}$ To compute the matrix $\mathbf{T}$ in (30), we only need to compute $\mathbb{E}\left[\mathcal{T}(\mathbf{y})\mathcal{T}(\mathbf{y})^T\right]$, which we will compute entry-wise and in two separate steps. Concretely, we have

$$\begin{aligned}\mathbb{E}[\mathcal{T}(y_m)\mathcal{T}(y_{m'})] &= \mathbb{E}\left[\exp(-\gamma(|z_m|^2+|z_{m'}|^2))\right] \\ &\quad\times \mathbb{E}[\exp(-\gamma([\mathbf{e}^y]_m + [\mathbf{e}^y]_{m'}))],\end{aligned}$$

where we compute both expected values separately. In the first step, we compute

$$\mathbb{E}\left[\exp(-\gamma(|z_m|^2+|z_{m'}|^2))\right] = \mathbb{E}\left[\exp(-\mathbf{u}^H\mathbf{G}\mathbf{u}))\right],$$

with $\mathbf{u} = [z_m, z_{m'}]^T$ and $\mathbf{G} = \mathbf{I}_2\gamma$. By invoking Lemma 2 with $[\Sigma]_{m,m'} = [\mathbf{C}_\mathbf{z}]_{m,m'}$, we obtain

$$\begin{aligned}&\mathbb{E}\left[\exp(-\gamma(|z_m|^2+|z_{m'}|^2))\right] = \frac{1}{|\gamma\Sigma+\mathbf{I}_2|} \\ &= \frac{1}{(\gamma[\mathbf{C}_\mathbf{z}]_{m,m}+1)(\gamma[\mathbf{C}_\mathbf{z}]_{m',m'}+1) - \gamma^2|[\mathbf{C}_\mathbf{z}]_{m,m'}|^2}.\end{aligned}$$

With the definition of $\mathbf{q}_\gamma$ in (38), we can rewrite the above expression in vector form as

$$\begin{aligned}&\mathbb{E}\left[\exp(-\gamma|\mathbf{z}|^2)\exp(-\gamma|\mathbf{z}|^2)^T\right] \\ &\quad= \mathbf{1}_{M\times M} \oslash (\mathbf{q}_\gamma\mathbf{q}_\gamma^T - \gamma^2\mathbf{C}_\mathbf{z}\odot\mathbf{C}_\mathbf{z}^*).\end{aligned}$$

In the second step, we compute

$$\mathbb{E}[\exp(-\gamma([\mathbf{e}^y]_m + [\mathbf{e}^y]_{m'}))] = \mathbb{E}\left[\exp(-\boldsymbol{\gamma}^T\mathbf{u})\right]$$

with $\mathbf{u} = [[\mathbf{e}^y]_m, [\mathbf{e}^y]_{m'}]^T$ and $\boldsymbol{\gamma}^T = [\gamma,\gamma]$. By invoking Lemma 3 with mean $\bar{\mathbf{u}} = [[\bar{\mathbf{e}}^y]_m, [\bar{\mathbf{e}}^y]_{m'}]$ and covariance Σ given by the entries of the covariance matrix $\mathbf{C}_{\mathbf{e}^y}$ associated to the indices m and m', we obtain

$$\begin{aligned}\mathbb{E}[\exp(-\gamma([\mathbf{e}^y]_m + [\mathbf{e}^y]_{m'}))] &= \exp(-\gamma([\bar{\mathbf{e}}^y]_m + [\bar{\mathbf{e}}^y]_{m'})) \\ &\quad\times\exp(\tfrac{1}{2}\gamma^2([\mathbf{C}_{\mathbf{e}^y}]_{m,m} + [\mathbf{C}_{\mathbf{e}^y}]_{m',m'} + 2[\mathbf{C}_{\mathbf{e}^y}]_{m,m'})).\end{aligned}$$

With the definition of $\mathbf{p}_\gamma$ in (39), we can rewrite the above expression in vector form as

$$\mathbb{E}\left[\exp(-\gamma\mathbf{e}^y)\exp(-\gamma\mathbf{e}^y)^T\right] = (\mathbf{p}_\gamma\mathbf{p}_\gamma^T)\odot\exp(\gamma^2\mathbf{C}_{\mathbf{e}^y})$$

We furthermore have

$$\overline{\mathcal{T}}(\mathbf{y})\overline{\mathcal{T}}(\mathbf{y})^T = (\mathbf{p}_\gamma\mathbf{p}_\gamma^T)\oslash(\mathbf{q}_\gamma\mathbf{q}_\gamma^T).$$

By combining the two steps with the above results, we have

$$\begin{aligned}\mathbf{T} = (\mathbf{p}_\gamma\mathbf{p}_\gamma^T)\odot\big(&\exp(\gamma^2\mathbf{C}_{\mathbf{e}^y})\oslash(\mathbf{q}_\gamma\mathbf{q}_\gamma^T - \gamma^2\mathbf{C}_\mathbf{z}\odot\mathbf{C}_\mathbf{z}^*) \\ &- \mathbf{1}_{M\times M}\oslash(\mathbf{q}_\gamma\mathbf{q}_\gamma^T)\big).\end{aligned}$$

Computing $\mathbf{V}_m$ To compute the matrices $\mathbf{V}_m$, $m = 1,\dots,M$, in (31), we only need $\mathbb{E}\left[\mathcal{T}(y_m)\mathbf{x}\mathbf{x}^H\right]$ which we will compute entry-wise and in two steps. We have

$$\begin{aligned}\mathbb{E}[\mathcal{T}(y_m)x_n x_{n'}^*] &= \mathbb{E}\left[\exp(-\gamma|\mathbf{a}_m^H\mathbf{x} + [\mathbf{e}^z]_m|^2)x_n x_{n'}^*\right] \\ &\quad\times\mathbb{E}[\exp(-\gamma[\mathbf{e}^y]_m)]\,,\end{aligned}$$

where we next compute both expected values separately. As a first step, we use direct integration to compute the following expected value:

$$\begin{aligned}\mathbb{E}\left[\exp(-\gamma|\mathbf{a}_m^H\mathbf{x} + [\mathbf{e}^z]_m|^2)x_n x_{n'}^*\right] &= \int_{\mathbb{C}^{N+1}}\exp(-\gamma|\mathbf{a}_m^H\mathbf{x} + [\mathbf{e}^z]_m|^2) \\ &\times\frac{1}{(\pi\sigma_x^2)^N}\exp\left(-\frac{\|\mathbf{x}\|^2}{\sigma_x^2}\right) \\ &\times\frac{1}{\pi\sigma_n^2}\exp\left(-\frac{|[\mathbf{e}^z]_m|^2}{\sigma_n^2}\right)x_n x_{n'}^*\mathrm{d}\mathbf{x}\mathrm{d}[\mathbf{e}^z]_m.\end{aligned}$$

We define the following auxiliary quantities:

$$\begin{aligned}\tilde{\mathbf{a}}_m^H &= [\,\mathbf{a}_m^H, 1\,] \\ \tilde{\mathbf{x}}^T &= [\,\mathbf{x}^T, [\mathbf{e}^z]_m\,] \\ \mathbf{C}_{\tilde{\mathbf{x}}} &= \begin{bmatrix}\sigma_x^2\mathbf{I}_N & \mathbf{0}_{N\times 1}\\ \mathbf{0}_{1\times N} & \sigma_m^2\end{bmatrix} \\ \widetilde{\mathbf{K}}^{-1} &= \gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H + \mathbf{C}_{\tilde{\mathbf{x}}}^{-1},\end{aligned}$$

where $\sigma_m^2 = \mathbb{E}\big[|[\mathbf{e}^z]_m|^2\big] = [\mathbf{C}_{\mathbf{n}^z}]_{m,m}$. We now derive the above expectation in compact form as

$$\begin{aligned}
&\mathbb{E}\big[\exp(-\gamma|\tilde{\mathbf{a}}_m^H\tilde{\mathbf{x}}|^2)\tilde{x}_n\tilde{x}_{n'}^*\big] = \\
&= \frac{1}{(\pi\sigma_x^2)^N}\frac{1}{\pi\sigma_n^2}\int_{\mathbb{C}^{N+1}}\exp(-\gamma|\tilde{\mathbf{a}}^H\tilde{\mathbf{x}}|^2 - \tilde{\mathbf{x}}^H\mathbf{C}_{\tilde{\mathbf{x}}}^{-1}\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}} \\
&= \frac{1}{|\pi\mathbf{C}_{\tilde{\mathbf{x}}}|}\int_{\mathbb{C}^{N+1}}\exp(-\tilde{\mathbf{x}}^H(\gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H + \mathbf{C}_{\tilde{\mathbf{x}}}^{-1})\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}} \\
&= \frac{1}{|\pi\mathbf{C}_{\tilde{\mathbf{x}}}|}\int_{\mathbb{C}^{N+1}}\exp(-\tilde{\mathbf{x}}^H\widetilde{\mathbf{K}}^{-1}\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}},
\end{aligned}$$

where $n = 1,\ldots,N+1$, $n' = 1,\ldots,N+1$. We can further rewrite this expression as

$$\begin{aligned}
&\frac{1}{|\pi\mathbf{C}_{\tilde{\mathbf{x}}}|}\int_{\mathbb{C}^{N+1}}\exp(-\tilde{\mathbf{x}}^H\widetilde{\mathbf{K}}^{-1}\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}} \\
&\qquad = \frac{|\pi\widetilde{\mathbf{K}}|}{|\pi\widetilde{\mathbf{K}}||\pi\mathbf{C}_{\tilde{\mathbf{x}}}|}\int_{\mathbb{C}^{N+1}}\exp(-\tilde{\mathbf{x}}^H\widetilde{\mathbf{K}}^{-1}\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}}.
\end{aligned}$$

It is now key to realize that

$$\begin{aligned}
&\frac{1}{|\pi\widetilde{\mathbf{K}}|}\int_{\mathbb{C}^{N+1}}\exp(-\tilde{\mathbf{x}}^H\widetilde{\mathbf{K}}^{-1}\tilde{\mathbf{x}})\tilde{x}_n\tilde{x}_{n'}^*\mathrm{d}\tilde{\mathbf{x}} \\
&\qquad = \mathbb{E}[\tilde{x}_n\tilde{x}_{n'}^*] = [\widetilde{\mathbf{K}}]_{n,n'}
\end{aligned}$$

and hence we have

$$\begin{aligned}
&\mathbb{E}\big[\exp(-\gamma|\tilde{\mathbf{a}}_m^H\tilde{\mathbf{x}}|^2)\tilde{x}_n\tilde{x}_{n'}^*\big] \\
&\quad = \frac{|\widetilde{\mathbf{K}}|}{|\mathbf{C}_{\tilde{\mathbf{x}}}|}[\widetilde{\mathbf{K}}]_{n,n'} = \frac{1}{|\widetilde{\mathbf{K}}^{-1}||\mathbf{C}_{\tilde{\mathbf{x}}}|}[\widetilde{\mathbf{K}}]_{n,n'} \\
&\quad = \frac{1}{|\gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H + \mathbf{C}_{\tilde{\mathbf{x}}}^{-1}||\mathbf{C}_{\tilde{\mathbf{x}}}|}[\widetilde{\mathbf{K}}]_{n,n'} \\
&\quad = \frac{1}{|\gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}} + \mathbf{I}_{N+1}|}[\widetilde{\mathbf{K}}]_{n,n'}.
\end{aligned}$$

We can now use the matrix-determinant lemma to simplify

$$\begin{aligned}
|\gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}} + \mathbf{I}_{N+1}| &= \gamma\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}}\tilde{\mathbf{a}}_m + 1 \\
&= \gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1
\end{aligned}$$

and the matrix inversion lemma to simplify

$$\begin{aligned}
\widetilde{\mathbf{K}} &= (\gamma\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H + \mathbf{C}_{\tilde{\mathbf{x}}}^{-1})^{-1} \\
&= \mathbf{C}_{\tilde{\mathbf{x}}} - \frac{\gamma\mathbf{C}_{\tilde{\mathbf{x}}}\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}}}{\gamma\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}}\tilde{\mathbf{a}}_m + 1} \\
&= \mathbf{C}_{\tilde{\mathbf{x}}} - \frac{\gamma\mathbf{C}_{\tilde{\mathbf{x}}}\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}}}{\gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1}.
\end{aligned}$$

By using these two simplifications, we have

$$\begin{aligned}
&\mathbb{E}\big[\exp(-\gamma|\tilde{\mathbf{a}}_m^H\tilde{\mathbf{x}}|^2)\tilde{x}_n\tilde{x}_{n'}^*\big] \\
&= \frac{1}{\gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1} \\
&\qquad \times \left[\mathbf{C}_{\tilde{\mathbf{x}}} - \frac{\gamma\mathbf{C}_{\tilde{\mathbf{x}}}\tilde{\mathbf{a}}_m\tilde{\mathbf{a}}_m^H\mathbf{C}_{\tilde{\mathbf{x}}}}{\gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1}\right]_{n,n'}
\end{aligned}$$

and since we are only interested in the upper $N \times N$ part of the matrix $\widetilde{\mathbf{K}}$, we have

$$\begin{aligned}
&\mathbb{E}\big[\exp(-\gamma|\mathbf{a}_m^H\mathbf{x} + [\mathbf{e}^z]_m|^2)x_n x_{n'}^*\big] \\
&= \frac{1}{\gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1} \\
&\qquad \times \left[\sigma_x^2\mathbf{I}_N - \frac{\gamma\sigma_x^4\mathbf{a}_m\mathbf{a}_m^H}{\gamma(\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2) + 1}\right]_{n,n'} \\
&= \frac{1}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\left[\sigma_x^2\mathbf{I}_N - \frac{\gamma\sigma_x^4\mathbf{a}_m\mathbf{a}_m^H}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\right]_{n,n'}
\end{aligned}$$

since for our assumptions

$$\sigma_x^2\|\mathbf{a}_m\|^2 + \sigma_m^2 = [\mathbf{C}_{\mathbf{z}}]_{m,m}.$$

In compact matrix form, we have

$$\begin{aligned}
&\mathbb{E}\big[\exp(-\gamma|\mathbf{a}_m^H\mathbf{x} + [\mathbf{e}^z]_m|^2)\mathbf{x}\mathbf{x}^H\big] \\
&\qquad = \frac{1}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\left(\sigma_x^2\mathbf{I}_N - \frac{\gamma\sigma_x^4\mathbf{a}_m\mathbf{a}_m^H}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\right).
\end{aligned}$$

As a second step, we use definition (39) and obtain

$$\mathbb{E}[\exp(-\gamma[\mathbf{e}^y]_m)] = [\mathbf{p}_\gamma]_m.$$

By combining both steps, we obtain

$$\begin{aligned}
\mathbf{V}_m &= \frac{[\mathbf{p}_\gamma]_m}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\left(\sigma_x^2\mathbf{I}_N - \frac{\gamma\sigma_x^4\mathbf{a}_m\mathbf{a}_m^H}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\right) \\
&\qquad - \frac{[\mathbf{p}_\gamma]_m}{\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1}\sigma_x^2\mathbf{I}_N \\
&= -\frac{\gamma\sigma_x^4[\mathbf{p}_\gamma]_m}{(\gamma[\mathbf{C}_{\mathbf{z}}]_{m,m} + 1)^2}\mathbf{a}_m\mathbf{a}_m^H,
\end{aligned}$$

which is what we desperately wanted to show.

References

Bahmani, S. and Romberg, J. Phase retrieval meets statistical learning theory: A flexible convex relaxation. In *Proc. Intl. Conf. Artif. Intell. and Stat.*, pp. 252–260, May 2017.

Bendory, T. and Eldar, Y. C. Non-convex phase retrieval from STFT measurements. *IEEE Trans. Inf. Theory*, Aug. 2017.

Candès, E. J. and Li, X. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Found. Comput. Math.*, 14(5):1017–1026, Oct. 2014.

Candès, E. J., Thomas, S., and Voroninski, V. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Commun. Pure Appl. Math.*, 66(8):1241–1274, 2013.

Candès, E. J., Eldar, Y. C., Strohmer, T., and Voroninski, V. Phase retrieval via matrix completion. *SIAM Rev.*, 57(2): 225–251, Nov. 2015a.

Candès, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory*, 61(4):1985–2007, Feb. 2015b.

Chandra, R., Zhong, Z., Hontz, J., McCulloch, V., Studer, C., and Goldstein, T. PhasePack: A phase retrieval library. *arXiv preprint: 1711.10175*, Nov. 2017.

Chen, P., Fannjiang, A., and Liu, G. Phase retrieval with one or two diffraction patterns by alternating projections of the null vector. *arXiv preprint: 1510.07379*, Apr. 2015.

Chen, Y. and Candès, E. J. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Proc. Adv. Neural Info. Proc. Syst.*, pp. 739–747, Dec. 2015.

Dhifallah, O. and Lu, Y. Fundamental limits of PhaseMax for phase retrieval: A replica analysis. *arXiv preprint: 1708.03355*, Aug. 2017.

Dhifallah, O., Thrampoulidis, C., and Lu, Y. Phase retrieval via linear programming: Fundamental limits and algorithmic improvements. *arXiv preprint: 1710.05234*, Oct. 2017.

Faulkner, H. M. L. and Rodenburg, J. M. Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm. *Phys. Rev. Lett.*, 93(2):023903, July 2004.

Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, Aug. 1982.

Fogel, F., Waldspurger, I., and d'Aspremont, A. Phase retrieval for imaging problems. *Math. Prog. Comp.*, 8(3): 311–335, Sept. 2016.

Gerchberg, R. W. and Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, Aug. 1972.

Ghods, R., Lan, A. S., Goldstein, T., and Studer, C. PhaseLin: Linear phase retrieval. In *Proc. 52nd Ann. Conf. Info. Sci. Systems*, Mar. 2018.

Goldstein, T. and Studer, C. Convex phase retrieval without lifting via PhaseMax. In *Proc. Intl. Conf. Mach. Learn.*, pp. 1273–1281, Aug. 2017.

Hand, P. and Voroninski, V. An elementary proof of convex phase retrieval in the natural parameter space via the linear program PhaseMax. *arXiv preprint: 1611.03935*, Nov. 2016.

Harrison, R. W. Phase problem in crystallography. *J. Opt. Soc. Am. A*, 10(5):1046–1055, May 1993.

Holloway, J., Asif, M. S., Sharma, M. K., Matsuda, N., Horstmeyer, R., Cossairt, O., and Veeraraghavan, A. Toward long-distance subdiffraction imaging using coherent camera arrays. *IEEE Trans. Comput. Imag.*, 2(3):251–265, Sept. 2016.

Kan, R. and Robotti, C. On moments of folded and truncated multivariate normal distributions. *J. Comput. Graph Stat.*, 26(4):1–5, Apr. 2017.

Kou, S. S., Waller, L., Barbastathis, G., and Sheppard, C. J. R. Transport-of-intensity approach to differential interference contrast (TI-DIC) microscopy for quantitative phase imaging. *Opt. Lett.*, 35(3):447–449, Feb. 2010.

Lu, Y. and Li, G. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *arXiv preprint: 1702.06435*, Apr. 2017.

Ma, J., Xu, J., and Maleki, A. Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2-regularization. *arXiv preprint: 1801.01170*, Jan. 2018.

Metzler, C. A., Sharma, M. K., Nagesh, S., Baraniuk, R. G., Cossairt, O., and Veeraraghavan, A. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *Proc. IEEE Intl. Conf. Comput. Photograph.*, pp. 1–16, May. 2017.

Miao, J., Ishikawa, T., Shen, Q., and Earnest, T. Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes. *Ann. Rev. Phys. Chem.*, 59:387–410, Nov. 2008.

Mondelli, M. and Montanari, A. Fundamental limits of weak recovery with applications to phase retrieval. *arXiv preprint: 1708.05932*, Sep. 2017.

Netrapalli, P., Jain, P., and Sanghavi, S. Phase retrieval using alternating minimization. In *Proc. Adv. Neural Info. Proc. Syst.*, pp. 2796–2804, Dec. 2013.

Pfeiffer, F., Weitkamp, T., Bunk, O., and David, C. Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nat. Phys.*, 2(4):258–261, Apr. 2006.

Salehi, F., Abbasi, E., and Hassibi, B. A precise analysis of PhaseMax in phase retrieval. *arXiv preprint: 1801.06609*, Jan. 2018.

Schniter, P. and Rangan, S. Compressive phase retrieval via generalized approximate message passing. *IEEE Trans. Sig. Process.*, 63(4):1043–1055, Feb. 2015.

Sun, J., Qu, Q., and W., J. A geometric analysis of phase retrieval. *IEEE Intl. Symp. Info. Th.*, July 2016.

Waldspurger, I., d'Aspremont, A., and Mallat, S. Phase recovery, maxcut and complex semidefinite programming. *Math. Prog.*, 149(1-2):47–81, Feb. 2015.

Wang, G., Giannakis, G. B., and Eldar, Y. C. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv preprint: 1605.08285*, Aug. 2017a.

Wang, G., Giannakis, G. B., Saad, Y., and Chen, J. Solving almost all systems of random quadratic equations. *arXiv preprint: 1705.10407*, May 2017b.

Wei, K. Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study. *Inverse Probl.*, 31(12):125008, Nov. 2015.

Yeh, L., Dong, J., Zhong, J., Tian, L., Chen, M., Tang, G., Soltanolkotabi, M., and Waller, L. Experimental robustness of Fourier ptychography phase retrieval algorithms. *Opt. Express*, 23(26):33214–33240, Dec. 2015.

Yuan, Z. and Wang, H. Phase retrieval via reweighted Wirtinger flow. *Appl. Op.*, 56(9):2418, Mar. 2017.

Zeng, W. and So, H. C. Coordinate descent algorithms for phase retrieval. *arXiv preprint: 1706.03474*, June 2017.

Zhang, H. and Liang, Y. Reshaped Wirtinger flow for solving quadratic system of equations. In *Proc. Adv. Neural Info. Proc. Syst.*, pp. 2622–2630, Dec. 2016.

Zhang, H., Chi, Y., and Liang, Y. Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow. In *Proc. Intl. Conf. Mach. Learn.*, pp. 1022–1031, June 2016.