SIAM J. IMAGING SCIENCES
Vol. 12, No. 4, pp. 2119–2164

A Generalization of Wirtinger Flow for Exact Interferometric Inversion*

Bariscan Yonel† and Birsen Yazici†

Abstract. Interferometric inversion involves recovery of a signal from *cross*-correlations of its linear transformations. A close relative of interferometric inversion is the generalized phase retrieval problem, for which significant advancements were made in recent years despite the ill-posed and nonconvex nature of the problem. One such prominent phase retrieval method is Wirtinger flow (WF) [E. J. Candes, X. Li, and M. Soltanolkotabi, IEEE Trans. Inform. Theory, 61 (2015), pp. 1985–2007], a computationally efficient nonconvex optimization framework that provides high probability guarantees for exact recovery under certain measurement models, specifically coded diffraction patterns, and Gaussian sampling vectors. In this paper, we develop a generalization of WF for interferometric inversion, which we refer to as generalized Wirtinger flow (GWF). Our approach treats the theory of low rank matrix recovery and the nonconvex optimization approach of WF in a unified framework. Such a treatment facilitates the identification of a new sufficient condition on the lifted forward model for exact recovery via GWF and results in a deterministic framework based on geometric arguments for convergence. Thereby, GWF extends the model specific probabilistic guarantees in [12] to arbitrary measurement maps characterized over the equivalent lifted domain in the context of interferometric inversion, covering both random and deterministic measurement models. We then establish our sufficient condition for the cross-correlations of linear measurements collected by complex Gaussian sampling vectors. In the particular case of interferometric inversion with the Gaussian model, we show that the exact recovery theory of standard WF implies our sufficient condition when we have cross-correlations, and the regularity condition of WF is redundant. Finally, we demonstrate the effectiveness of GWF numerically in a deterministic, interferometric multistatic radar imaging scenario.

Key words. interferometric inversion, Wirtinger flow, phase retrieval, wave-based imaging, low rank matrix recovery, PhaseLift, interferometric imaging

AMS subject classifications. 46N10, 47N10, 47N30, 30E99

DOI. 10.1137/19M1238599

1. Introduction.

Interferometric inversion involves the recovery of a signal of interest from the cross-correlations of its linear measurements, each collected by a different sensing process. Let $\mathbf{L}_i^m, \mathbf{L}_j^m \in \mathbb{C}^N$ denote the mth sampling vectors of the ith and jth sensing processes and let $\boldsymbol{\rho}_t \in \mathbb{C}^N$ be the ground truth/signal of interest. We define

$$f_i^m = \langle \mathbf{L}_i^m, \boldsymbol{\rho}_t \rangle, \quad f_j^m = \langle \mathbf{L}_j^m, \boldsymbol{\rho}_t \rangle, \quad m = 1, \dots, M, \tag{1.1}$$

*Received by the editors January 14, 2019; accepted for publication (in revised form) August 12, 2019; published electronically December 18, 2019.
https://doi.org/10.1137/19M1238599
Funding: The work of the authors was supported by Air Force Office of Scientific Research (AFOSR) grant FA9550-16-1-0234, Office of Naval Research (ONR) grant N0001418-1-2068, and National Science Foundation (NSF) grant ECCS-1809234.
†Electrical and Computer Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 (yonelb@rpi.edu, yazici@ecse.rpi.edu).

as the linear measurements and describe the cross-correlated measurements as

$$d_{ij}^m = f_i^m \overline{f_j^m} = (\mathbf{L}_i^m)^H \boldsymbol{\rho}_t \boldsymbol{\rho}_t^H \mathbf{L}_j^m, \quad m = 1, \ldots, M, \tag{1.2}$$

where $\overline{(\cdot)}$ denotes complex conjugation. Thus, interferometric inversion involves recovery of $\boldsymbol{\rho}_t \in \mathbb{C}^N$ from $d_{ij}^m \in \mathbb{C}$, $m = 1, \ldots, M$, using the model in (1.2).

The interferometric inversion problem arises in many applications in different disciplines. These include radar and sonar interferometry [2, 27, 41], passive imaging in acoustic, electromagnetic, and geophysical applications [1, 33, 44, 45, 51, 52, 55, 56, 57, 58, 59, 60, 61, 63, 64, 65], interferometric microscopy [37], beamforming and sensor localization in large area networks [46], [26, 29, 36] among others. Additionally, cross-correlations were shown to provide robustness to statistical fluctuations in scattering media or incoherent sources in wave-based imaging [25, 32] and with respect to phase errors in the correlated linear transformations [6, 23, 28, 34]. Therefore, in applications such as passive imaging [1, 33, 44, 45, 51, 52, 55, 56, 57, 58, 59, 60, 61, 63, 64, 65] and interferometry [2, 27, 41], cross-correlations are formed as a part of the inference process after acquiring linear measurements by sensors that are configured differently in space, time, or frequency. Additionally, the cross-correlated measurement model arises naturally from the underlying physical sensing processes in certain applications such as optical and radio astronomy [19, 30] or quantum optical imaging [42].

A special case of the interferometric inversion problem is when $i = j$ in (1.2), in which the model becomes the autocorrelations of linear measurements collected by a single sensing process. In this case, the interferometric inversion problem reduces to the well-known *phase retrieval* problem. Notably, both problems are nonconvex due to the quadratic equality constraints enforced by the correlated measurement model. In recent years, several phase retrieval methods with exact recovery guarantees have been developed despite the nonconvex nature of the problem. These methods are characterized by either one or both of the following two principles: convexification of the solution set, which includes *lifting*-based approaches [9, 10, 53], or a provably accurate initialization, followed by an algorithmic map that refines the initial estimate which is most prominently established by Wirtinger flow (WF) [12] and its variants [4, 15, 16, 43, 54, 69, 70, 71].

In methods that deploy lifting, such as PhaseLift [9, 10], signal recovery from quadratic measurements is reformulated as a low rank matrix recovery (LRMR) problem. While the LRMR approach offers convergence guarantees via convexification, it has limited practical applicability in typical sensing and imaging problems since lifting increases the dimension of the inverse problem by an order of magnitude and requires demanding memory storage in implementation. The WF framework, on the other hand, avoids lifting and hence offers considerable advantages in computational complexity and memory requirements. Despite solving the non-convex problem directly, convergence to a global solution at a geometric rate is guaranteed by WF for coded diffraction patterns and Gaussian sampling vectors [12] and more recently for short time Fourier transforms [4]. These advantages promote WF as a theoretical framework suitable for large scale imaging problems.

Conventionally, interferometric inversion in imaging applications has been approached by Fourier-based techniques, such as time or frequency difference of arrival backprojection [50, 51, 52, 56, 58, 60, 63, 64]. While these methods are practical and computationally efficient, their applicability is limited to scenes composed of well-separated point targets due to underlying assumptions. As an alternative, LRMR theory has been explored for interferometric inversion [21, 33]. In particular, [21] assesses the robustness of solutions to several convex semidefinite programs (SDPs) in a deterministic framework after lifting the interferometric measurements. Notably, these SDP solvers are inspired by the PhaseLift method [9, 10, 20] and hence suffer from the same drawbacks in practice. In [33], an iterative optimization approach to LRMR was developed for interferometric passive imaging to circumvent the poor scaling properties of SDP approaches. The methodology in [33] is based on Uzawa's method for matrix completion [7, 38]. While this method is computationally more efficient than the SDP solvers, it still operates on the lifted domain and hence requires significant memory and computational resources. Additionally, these convexified lifting-based solvers require stringent theoretical conditions on the measurement model, which poses a major theoretical barrier for interferometric inversion problems with deterministic forward models.

In this paper, motivated by its advantages over lifting-based methods, we develop a generalization of WF applicable to interferometric inversion problems with exact recovery guarantees. We refer to this method as the generalized Wirtinger flow (GWF). Beyond the immediate extension of algorithmic principles of WF to the interferometric inversion, our mathematical framework differs from that of WF in two significant ways. (i) The GWF theory is established in the lifted domain. This lifting-based perspective allows us to develop a novel approach that bridges the theory between LRMR and the nonconvex methodology of WF and culminates in derivation of a new sufficient condition for exact recovery in the context of interferometric inversion. Namely, GWF guarantees exact recovery to a general class of problems that are characterized over the equivalent lifted domain by the restricted isometry property (RIP) on the set of rank-1, positive semidefinite (PSD) matrices. Specifically, we provide an upper bound for the *restricted isometry constant* (RIC) over the set of rank-1, PSD matrices so that RIP directly implies convergence to a global solution. (ii) Given the sufficient condition, we derive exact recovery guarantees for an arbitrary mapping over the lifted domain using solely geometric arguments in a deterministic framework, rather than model specific probabilistic arguments of standard WF. Notably, our analysis proves that the regularity condition, which forms the basis of convergence in standard WF, is redundant for the case of interferometric inversion.

In addition, developing the GWF framework through the equivalent lifted problem allows us to identify the key theoretical advantages of the nonconvex optimization approach over lifting based convex solvers beyond the immediate gains in computation and applicability. Our exact recovery condition for GWF proves to be a less stringent RIP than that of the sufficient conditions of LRMR methods. As a result of its fundamental differences from standard WF, and less stringent exact recovery requirements than those of LRMR methods, GWF offers a promising step toward exact recovery guarantees for interferometric inversion problems which are governed by deterministic measurement models. One such application is array imaging, for which RIP was shown to hold over rank-1 matrices for sufficiently high central frequencies, albeit asymptotically in the number of receivers [14]. Accordingly, we consider GWF as an exact interferometric imaging method alternative to LRMR for multistatic radar imaging, and we design imaging system parameters for the lifted forward model to satisfy of our sufficient condition in the nonasymptotic regime [66].

In our key results, we first prove that the RIP over rank-1, PSD matrices on the lifted forward model implies an accurate initialization by the spectral method and ensures that the regularity condition is satisfied in the ϵ-neighborhood defined by the initialization if RIC is less than 0.214. Next, although not applicable to the case of autocorrelations in [12], we show that this sufficient condition is satisfied for the cross-correlation of linear measurements collected by independent and identically distributed (i.i.d.) Gaussian sampling vectors. Essentially, our results establish that the probabilistic arguments used for the Gaussian sampling model in [12] are fully captured by the single event that the RIP is satisfied over rank-1, PSD matrices in the case $i \neq j$ in (1.2). To validate our analysis we conduct numerical experiments for the Gaussian sampling model by counting empirical probability of exact recovery. We then demonstrate the effectiveness of GWF in a realistic passive multistatic radar imaging scenario. Our preliminary numerical simulations confirm our theoretical results and show that the interferometric inversion is solved in an exact manner by GWF.

The rest of our paper is organized as follows. We first introduce the problem formulation for interferometric inversion on the signal domain and discuss lifting based prior art in section 2. We then formulate the GWF algorithm in section 3 and present key definitions and terminology followed by the main theorem statements in section 4. The proofs of the main theorems are provided in section 5. The numerical simulations for Gaussian sampling model and interferometric multistatic radar imaging are presented in section 6. Section 7 concludes our paper. Appendices A, B, and C include proofs of lemmas used in section 4. The notations used in the paper are provided in Table 1.

2. Problem formulation and prior art.

In this section, we introduce the nonconvex formulation of the interferometric inversion problem and its key challenges. Next we discuss lifting-based, convex formulations which address these key challenges in solving quadratic systems of equations via LRMR theory.

2.1. The Nonconvex objective function.

To address the interferometric inversion problem, we define the following objective function and set up the corresponding optimization problem:

$$\mathcal{J}(\boldsymbol{\rho}) := \frac{1}{2M} \sum_{m=1}^{M} |\,(\mathbf{L}_i^m)^H \boldsymbol{\rho}\boldsymbol{\rho}^H \mathbf{L}_j^m - d_{ij}^m|^2, \tag{2.1}$$

$$\hat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}}{\operatorname{argmin}}\ \mathcal{J}(\boldsymbol{\rho}). \tag{2.2}$$

Let $\mathbf{y} = [y_{ij}^1, y_{ij}^2, \cdots y_{ij}^M]^T$ and $\mathcal{L} : \mathbb{C}^N \to \mathbb{C}^M$ be the cross-correlated measurement map defined as

$$\mathcal{L}(\boldsymbol{\rho}) = \mathbf{y}, \text{ where } \boldsymbol{\rho} \in \mathbb{C}^N, \quad y_{ij}^m = (\mathbf{L}_i^m)^H \boldsymbol{\rho}\boldsymbol{\rho}^H \mathbf{L}_j^m. \tag{2.3}$$

The objective function in (2.1) is the ℓ_2 mismatch in the range of $\mathcal{L}$, i.e., the space of cross-correlated measurements, which is solved over the signal domain $\mathbb{C}^N$ in (2.2). The difficulty in (2.2) is that the objective function $\mathcal{J}$ is nonconvex over the variable $\boldsymbol{\rho}$ due to the invariance of cross-correlated measurements to global phase factors. Essentially (2.2) has a nonconvex solution set with infinitely many elements, which casts interferometric inversion as a challenging, ill-posed problem.

Table 1
Table of notation.

Symbol	Description
$\mathbf{L}_i^m$	mth measurement vector of ith receiver
$\mathbf{L}_j^m$	mth measurement vector of jth receiver
$\mathbf{L}^m$	mth measurement vector of a single receiver, i.e., $i = j$
d_{ij}^m	Interferometric data from correlating receivers i and j
d^m	Self-correlated (phaseless) measurements
M	Number of measurements
N	Number of unknowns
$\boldsymbol{\rho}$	Signal in $\mathbb{C}^N$
$\mathbf{X}$	Variable in the lifted domain in $\mathbb{C}^{N\times N}$
$\boldsymbol{\rho}_t$	Ground truth signal in $\mathbb{C}^N$
$\tilde{\boldsymbol{\rho}}$	Lifted signal $\boldsymbol{\rho}\boldsymbol{\rho}^H$
$\tilde{\boldsymbol{\rho}}_t$	Lifted ground truth $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$
$\mathcal{L}$	(Nonlinear) Operator of quadratic correlation mapping
$\mathbf{y}$	Element in the range of $\mathcal{L}$
$\mathcal{E}_{\boldsymbol{\rho}}$	Equivalence class of $\boldsymbol{\rho}$ under $\mathcal{L}$ over $\mathcal{R}$
$P_{\boldsymbol{\rho}}$	Equivalence set of $\boldsymbol{\rho}$ under $\mathcal{L}$ over $\mathcal{R}$
P	Global solution set with $P_{\boldsymbol{\rho}_t} = P$
$\mathcal{F}$	Lifted forward model
$\mathcal{F}^H$	Adjoint of $\mathcal{F}$/backprojection operator
$\mathcal{J}$	Objective function
$\nabla\mathcal{J}$	Gradient of the objective function with respect to $\boldsymbol{\rho}$
$\mathbf{e}$	Mismatch between correlated measurements
$\mathcal{X}$	Set of rank-1, PSD matrices
$\mathcal{P}$	Projection operator
S	Set of symmetric matrices
PSD	Positive semidefinite cone
$\circ$	Successive operation of projections
dist	Distance metric
$\mathbf{Y}$	Backprojection estimate of the lifted unknown
$\hat{\mathbf{X}}$	Spectral matrix of GWF
$\boldsymbol{\rho}_0$	Initialization from spectral method
$\tilde{\boldsymbol{\rho}}_0$	Initial lifted signal
$\lambda_0, \mathbf{v}_0$	Leading eigenvalue-eigenvector pair of $\hat{\mathbf{X}}$
$E(\epsilon)$	ϵ-neighborhood of the global solution set
$\mathbf{I}$	Identity operator on any domain
$\boldsymbol{\delta}$	Additive perturbation operator on $\mathbf{I}$
$\tilde{\mathbf{e}}$	Error in the lifted domain
RIP_+	Restricted isometry property over the PSD cone
$\text{RIC} - \delta_1$	Restricted isometry constant over rank-1 matrices

Definition 2.1 (global solution set). *We say that the points*

$$P := \left\{ e^{i\phi}\boldsymbol{\rho}_t : \phi \in [0, 2\pi) \right\}$$

form the global solution set for the interferometric inversion from the cross-correlated measurements (1.2).

More generally, for any $\boldsymbol{\rho} \in \mathbb{C}^N$, let $\mathcal{E}_{\boldsymbol{\rho}} = \{\mathbf{z} \in \mathbb{C}^N : \mathcal{L}(\mathbf{z}) = \mathcal{L}(\boldsymbol{\rho})\}$ be the equivalence class of $\boldsymbol{\rho}$ under $\mathcal{L}$. We then define the following collection of signals as the *equivalence set* of $\boldsymbol{\rho}$.

Definition 2.2 (equivalence set). *Let* $\boldsymbol{\rho} \in \mathbb{C}^N$ *and*

$$P_{\boldsymbol{\rho}} := \left\{ e^{\mathrm{i}\phi}\boldsymbol{\rho}, \phi \in [0, 2\pi) \right\}. \tag{2.4}$$

We refer to $P_{\boldsymbol{\rho}}$ *as the equivalence set of* $\boldsymbol{\rho}$.

Remark 2.3. Note that $P_{\boldsymbol{\rho}} \subset \mathcal{E}_{\boldsymbol{\rho}}$, and $P_{\boldsymbol{\rho}_t}$ is identical to the global solution set in Definition 2.1.

Alleviating the noninjectivity of the measurement map is a key step in formulating methods that guarantee exact recovery in phase retrieval literature [3] and offers us a blueprint in addressing (2.2). A key observation is that one can consider (2.3) as a mapping from a rank-1, PSD matrix $\boldsymbol{\rho}\boldsymbol{\rho}^H \in \mathbb{C}^{N\times N}$ instead of a quadratic map from the signal domain in $\mathbb{C}^N$ and attempt to recover $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$. This approach is known as the *lifting* technique, which is the main premise of LRMR-based phase retrieval [9, 10, 11, 53] and interferometric inversion methods [20, 21, 33].

2.2. Low rank matrix recovery via lifting. We adopt the concepts of the LRMR approach to the interferometric inversion problem based on PhaseLift [33], [10] and introduce the following definitions.

Definition 2.4 (lifting). *Each correlated measurement in* (1.2) *can be written in the form of an inner product of two rank-*1 *operators,* $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ *and* $\overline{\mathbf{F}}^m = \mathbf{L}_j^m(\mathbf{L}_i^m)^H$, *such that*[1]

$$d_{ij}^m = \langle \mathbf{F}^m, \tilde{\boldsymbol{\rho}}_t\rangle_F, \quad m = 1, \ldots, M, \tag{2.5}$$

where $\langle\cdot,\cdot\rangle_F$ *is the Frobenius inner product. We refer to the procedure of transforming interferometric inversion over* $\mathbb{C}^N$ *to the recovery of the rank-*1 *unknown* $\tilde{\boldsymbol{\rho}}_t$ *in* $\mathbb{C}^{N\times N}$ *as lifting.*

The lifting technique introduces a new linear measurement map which we define as follows.

Definition 2.5 (lifted forward model). *Let* $\mathbf{d} = [d_{ij}^1, d_{ij}^2, \cdots d_{ij}^M] \in \mathbb{C}^M$ *denote the vector obtained by stacking the cross-correlated measurements in* (1.2)*. Then using* (2.5)*, we define* $\mathcal{F} : \mathbb{C}^{N\times N} \to \mathbb{C}^M$ *as*

$$\mathbf{d} = \mathcal{F}(\tilde{\boldsymbol{\rho}}_t) \tag{2.6}$$

and refer to $\mathcal{F}$ *as the lifted forward model/map.*

Remark 2.6. The map $\mathcal{F}$ can be interpreted as an $M \times N^2$ matrix with $\bar{\mathbf{F}}^m$ as its rows in the vectorized problem in which $N \times N$ variable $\tilde{\boldsymbol{\rho}}_t$ is concatenated into a vector.

We refer to the problem of recovering of $\tilde{\boldsymbol{\rho}}_t$ from $\mathbf{d}$ using the model (2.6) as the lifted problem, or interferometric inversion in the lifted domain.

The main advantage of lifting is that $\forall\boldsymbol{\rho} \in \mathbb{C}^N$, each nonconvex equivalence set $P_{\boldsymbol{\rho}}$ is now mapped to a set with a single element $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}\boldsymbol{\rho}^H$. Using the definition of the lifted forward model, quadratic equality constraints reduce to affine equality constraints to define a convex manifold in $\mathbb{C}^{N\times N}$. Using the *rank*-1 *PSD* structure of the unknown $\tilde{\boldsymbol{\rho}}_t$, interferometric inversion in the lifted domain can be formulated as the following optimization problem:

[1] $\bar{(\cdot)}$ denotes elementwise complex conjugation.

$$\text{find: } \mathbf{X} \quad \text{s.t.} \quad \mathbf{d} = \mathcal{F}(\mathbf{X}),\ \mathbf{X} \succeq 0,\ \text{rank}(\mathbf{X}) = 1. \tag{2.7}$$

Here, we refer to $\mathbf{X} \in \mathbb{C}^{N \times N}$ as the lifted variable. Due to the fact that there surely exists a rank-1 solution from (2.6), (2.7) is equivalent to

$$\text{minimize: rank}(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{d} = \mathcal{F}(\mathbf{X}),\ \mathbf{X} \succeq 0, \tag{2.8}$$

which is known to be an NP-hard problem [7, 38]. Given its rank-minimization form, (2.8) is approached by LRMR theory analogous to compressive sensing [13, 22]. Most prominently, the nonconvex rank term of the objective function is relaxed by a convex surrogate, which under the PSD constraint corresponds to the trace norm. This results in the following formulation [9, 10, 14]:

$$\text{minimize: tr}(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{d} = \mathcal{F}(\mathbf{X}),\ \mathbf{X} \succeq 0, \tag{2.9}$$

which can be solved in polynomial time via semidefinite programming. To obtain a robust program, (2.9) is commonly *perturbed* using the least squares criterion, which under the additive i.i.d. noise assumption can be written as

$$\text{minimize: tr}(\mathbf{X}) \quad \text{s.t.} \quad \frac{1}{2}\|\mathcal{F}(\mathbf{X}) - \mathbf{d}\|_2^2 \leq \mathcal{O}(\sigma^2),\ \mathbf{X} \succeq 0, \tag{2.10}$$

where σ^2 corresponds to the variance in the case of Gaussian noise, and the order in the threshold can be tuned to obtain a desired lower bound on the log-likelihood of observing $\mathbf{d}$. Alternatively, one can equivalently formulate the Lagrangian of (2.10) with a proper choice of the regularization parameter λ in relation to the threshold, as follows:

$$\underset{\mathbf{X} \succeq 0}{\text{minimize:}}\ \frac{1}{2}\|\mathcal{F}(\mathbf{X}) - \mathbf{d}\|_2^2 + \lambda \text{tr}(\mathbf{X}). \tag{2.11}$$

(2.11) can be solved by Uzawa's method [7, 38] which is analogous to the singular value thresholding algorithm with a PSD constraint [33] with the following iterations:

$$\mathbf{X}_k = \mathcal{P}_+ \circ \mathcal{P}_{\tau_k}\left(\mathcal{F}^H \boldsymbol{\nu}_{k-1}\right), \tag{2.12}$$

$$\boldsymbol{\nu}_k = \boldsymbol{\nu}_{k-1} + \mu_k(\mathbf{d} - \mathcal{F}(\mathbf{X}_k)). \tag{2.13}$$

In (2.12) and (2.13), $\boldsymbol{\nu} = [\nu^1, \cdots \nu^M]^T$ denotes the Lagrange multipliers initialized as $\boldsymbol{\nu}_0 = 0$, μ_k is the step size, and subscript k denotes the iteration number. $\mathcal{P}_\tau$ is the shrinkage operator acting on the singular values of its argument with threshold $\tau_k = \mu_k \lambda$, which through the parameter λ of trace regularization enforces the low rank constraint. $\mathcal{P}_+$ is the projection operator onto the PSD cone.

As $M \ll N^2$ in typical estimation problems, the lifted forward model has a nontrivial null space. LRMR theory encapsulates identifying necessary and sufficient conditions on $\mathcal{F}$ in order to guarantee exact recovery despite having an underdetermined system of equations in (2.6). The key conditions on $\mathcal{F}$ are primarily characterized by its null space [35, 39, 40] or restricted isometry properties [5, 8, 38] on low rank matrices. Methods such as PhaseLift [9, 10] and PhaseCut [53] assert conditions on the mapping $\mathcal{F}$ such that there exists no feasible element in the PSD cone with a smaller trace than the true solution $\tilde{\boldsymbol{\rho}}_t$, from which exact recovery results to a unique minimizer are directly implied by the standard arguments of semidefinite programming [10]. Iterative optimization via Uzawa's method on the other hand requires RIP over rank-5 matrices with a sufficiently small RIC ($\leq 1/10$) for the convex problem in (2.11) to have the identical solution to the original nonconvex problem in (2.8). Notably, in this scenario the PSD constraint can simply be dropped due to the guaranteed uniqueness of the solution to the original problem and the convexity of the problem in (2.11). Furthermore, under the properties on $\mathcal{F}$ asserted by PhaseLift, it is observed in [20] that the lifted problem can be robustly solved as a convex feasibility problem by Douglas–Rachford splitting by eliminating the trace minimization step completely. These indicate a level of redundancy between trace regularization and the PSD constraint, which arises from the convexification of the problem.

Altogether, lifting-based approaches provide a profound perspective to the interferometric inversion problem. Our observation is that the key principles of lifting-based methods in establishing exact recovery guarantees are reciprocated in the nonconvex framework of WF. In fact, WF corresponds to solving a perturbed nonconvex feasibility problem over the lifted domain, and in this sense it is reminiscent of the optimizationless PhaseLift method of [20] and Uzawa's iterations in [33]. To observe this, we introduce the GWF iterations for interferometric inversion and develop the method as a solver in the lifted problem framework. The basis of our extension from WF to GWF framework is the identification of conditions on the lifted forward model for the exactness of a solution in the lifted domain.

It is worth noting that low rank models and LRMR theory cover a rich area of research with a wide range of applications in fields such as control theory and machine learning. Several of the LRMR methods discussed in this subsection were first proposed for matrix completion and relate strongly to the problem of low rank matrix factorization. Nonconvex optimization theory for low rank matrix factorization has undergone notable developments in recent years, with methods that offer exact recovery results for quadratic or bilinear equations of rank-r matrices in the random Gaussian model [73], or if the measurement map satisfies RIP over rank-$6r$ with a RIC $\leq 1/10$ [48]. More recently a primal-dual analysis was conducted in [72] to show that there exists no spurious local-minima in the nonconvex LRMR problem for restricted strongly convex and smooth objective functions. For further discussion on advances in nonconvex LRMR, we refer the reader to [18, 62].

3. Generalized Wirtinger flow framework for exact interferometric inversion.

In this section, we present the algorithmic principles of the GWF framework for exact interferometric inversion. We specifically identify and present the theoretical advantages obtained from viewing the algorithm in the lifted domain, despite operating on the signal domain. We first present the GWF iterations, then proceed with generalizing the spectral method for initialization using our lifted formulation of GWF. We finally provide an algorithm summary with specifications of computational complexity of each step.

3.1. GWF Iterations.

In presenting the GWF iterations, we begin by extending the iterative scheme of WF to the case of the interferometric measurement model in (1.2), with $i \neq j$. We next introduce an equivalent, novel interpretation of these iterations in the lifted domain, which yields the basis of our GWF framework for exact recovery with arbitrary lifted forward models.

3.1.1. Extending WF. In contrast to the lifting-based approaches, the nonconvex form of the problem in (2.2) is preserved in WF. Given an *accurate* initial estimate $\boldsymbol{\rho}_0$, WF involves using the following updates to refine the current estimate $\boldsymbol{\rho}_k$:

$$\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \frac{\mu_{k+1}}{\|\boldsymbol{\rho}_0\|^2} \nabla \mathcal{J}(\boldsymbol{\rho}_k). \tag{3.1}$$

Notably, $\mathcal{J}$ is a real-valued function of a complex variable $\boldsymbol{\rho} \in \mathbb{C}^N$ and, therefore, non-holomorphic. Hence, the gradient over $\mathcal{J}$ is defined by the means of Wirtinger derivatives

$$\nabla \mathcal{J} = \left(\frac{\partial \mathcal{J}}{\partial \boldsymbol{\rho}}\right)^H = \left(\frac{\partial \mathcal{J}}{\partial \overline{\boldsymbol{\rho}}}\right)^T, \text{ where} \tag{3.2}$$

$$\frac{\partial}{\partial \boldsymbol{\rho}} = \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{\rho}_R} - \mathrm{i}\frac{\partial}{\partial \boldsymbol{\rho}_I}\right), \quad \frac{\partial}{\partial \overline{\boldsymbol{\rho}}} = \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{\rho}_R} + \mathrm{i}\frac{\partial}{\partial \boldsymbol{\rho}_I}\right), \tag{3.3}$$

and $\boldsymbol{\rho} = \boldsymbol{\rho}_R + \mathrm{i}\boldsymbol{\rho}_I$, with $\boldsymbol{\rho}_R, \boldsymbol{\rho}_I \in \mathbb{R}^N$. Thus, the iterations in (3.1) correspond to that of the steepest descent method [12], where μ_{k+1} is the step size. For interferometric inversion by solving (2.2), $\nabla \mathcal{J}$ evaluated at $\boldsymbol{\rho}_k$ is given by

$$\nabla \mathcal{J}(\boldsymbol{\rho}_k) = \frac{1}{2M} \sum_{m=1}^{M} \left[\overline{e_{ij}^m}\left(\mathbf{L}_j^m \left(\mathbf{L}_i^m\right)^H \boldsymbol{\rho}_k\right) + e_{ij}^m \left(\mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H \boldsymbol{\rho}_k\right)\right], \tag{3.4}$$

where $e_{ij}^m = ((\mathbf{L}_i^m)^H \boldsymbol{\rho}_k \boldsymbol{\rho}_k^H \mathbf{L}_j^m - d_{ij}^m)$ is the mismatch between the synthesized and cross-correlated measurements. Note that in the case of phase retrieval we have $i = j$. Letting $\mathbf{L}^m = \mathbf{L}_i^m = \mathbf{L}_j^m$ and $d^m = |\langle \mathbf{L}^m, \boldsymbol{\rho}_t\rangle|^2$, for $m = 1, \ldots, M$, (3.4) reduces to the standard WF iterations of [12]. See Appendix A.1 for the derivation of $\nabla \mathcal{J}$ for $i \neq j$.

3.1.2. Interpretation of GWF updates in the lifted domain. Our formulation of the GWF framework resides in the lifted domain and reveals an illuminating interpretation of the updates in (3.4). Moving the common term of the current iterate $\boldsymbol{\rho}_k$ outside the summation, (3.4) can be expressed as

$$\nabla \mathcal{J}(\boldsymbol{\rho}_k) = \frac{1}{M}\left[\frac{1}{2}\left(\sum_{m=1}^{M} \overline{e_{ij}^m}\left(\mathbf{L}_j^m \left(\mathbf{L}_i^m\right)^H\right) + \sum_{m=1}^{M} e_{ij}^m \left(\mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H\right)\right)\right] \boldsymbol{\rho}_k. \tag{3.5}$$

From the definition of the lifted forward model in (2.6), the second term inside the brackets in (3.5) becomes the *backprojection* of the measurement error onto the adjoint space of $\mathcal{F}$.

Definition 3.1 (backprojection). *Let* $\mathbf{y} = [y^1, y^2, \cdots y^M] \in \mathbb{C}^M$. *For the lifted forward model* $\mathcal{F} : \mathbb{C}^{N \times N} \to \mathbb{C}^M$ *in Definition 2.5, we define the adjoint operator* $\mathcal{F}^H : \mathbb{C}^M \to \mathbb{C}^{N \times N}$ *as*

$$\mathcal{F}^H(\mathbf{y}) = \sum_{m=1}^{M} y^m \left(\mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H\right) \tag{3.6}$$

and refer to (3.6) *as the backprojection of* $\mathbf{y}$.

Since (3.5) consists of the average of terms that are Hermitian transposes of each other, the update term in (3.1) corresponds to backprojecting the mismatch between synthesized and true measurements, and then projecting it onto the set of symmetric matrices S, i.e.,

$$\nabla \mathcal{J}(\boldsymbol{\rho}_k) = \frac{1}{M} \mathcal{P}_S \left(\mathcal{F}^H(\mathbf{e}) \right) \boldsymbol{\rho}_k, \tag{3.7}$$

where $\mathbf{e} = [e_{ij}^1, e_{ij}^2, \cdots e_{ij}^M] \in \mathbb{C}^M$ is the measurement mismatch vector and $\mathcal{P}_S(\cdot)$ denotes the projection operator onto S.

The representation in (3.7) provides a novel perspective in interpreting GWF as a solver of the lifted problem. Consider the structure enforced by (3.1) for the nonrelaxed form of the low rank recovery problem in (2.8). In the GWF updates, the rank of the lifted variable $\mathbf{X}$ in (2.8) is merely fixed at one, and the rank minimization problem is converted to its original nonconvex feasibility problem in (2.7). Knowing that (2.1) corresponds to the ℓ_2 mismatch in the space of cross-correlated measurements, the GWF solver of (2.1) is equivalently the solver of the following lifted problem:

$$\text{minimize:} \quad \frac{1}{2M} \|\mathcal{F}(\mathbf{X}) - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}) = 1 \text{ and } \mathbf{X} \succeq 0. \tag{3.8}$$

Observe that the rank-1, PSD constraint precisely corresponds to minimization over the set of elements of $\mathcal{X} = \{\boldsymbol{\rho}\boldsymbol{\rho}^H, \boldsymbol{\rho} \in \mathbb{C}^N\}$ and that (3.8) can be equivalently cast as

$$\text{minimize:} \quad \frac{1}{2M} \|\mathcal{F}(\mathbf{X}) - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad \mathbf{X} = \boldsymbol{\rho}\boldsymbol{\rho}^H. \tag{3.9}$$

The updates to solve (3.9) then can be performed on the leading eigenspace of the lifted variable $\mathbf{X}$ directly by means of the *Jacobian* $\frac{\partial \bar{\mathbf{X}}}{\partial \bar{\boldsymbol{\rho}}}$. Since $\mathbf{X} = \mathbf{X}^H$, this yields[2]

$$\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \mu_{k+1} \left(\frac{\partial \overline{\mathbf{X}}}{\partial \bar{\boldsymbol{\rho}}} \left(\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{X}}} + \frac{\partial \mathcal{J}}{\partial \mathbf{X}} \right) \right)^T, \text{where} \tag{3.10}$$

(3.11)
$$\left(\frac{\partial \overline{\mathbf{X}}}{\partial \bar{\boldsymbol{\rho}}} \frac{\partial \mathcal{J}}{\partial \overline{\mathbf{X}}} \right)^T = \frac{1}{2M} \left(\mathcal{F}^H \mathcal{F}(\mathbf{X}) - \mathcal{F}^H \mathbf{d} \right) \boldsymbol{\rho}, \; \left(\frac{\partial \overline{\mathbf{X}}}{\partial \bar{\boldsymbol{\rho}}} \frac{\partial \mathcal{J}}{\partial \mathbf{X}} \right)^T = \frac{1}{2M} \left(\mathcal{F}^H \mathcal{F}(\mathbf{X}) - \mathcal{F}^H \mathbf{d} \right)^H \boldsymbol{\rho}.$$

Initializing the algorithm with a proper estimate in the constraint set $\mathcal{X}$, i.e., $\mathbf{X}_0 = \boldsymbol{\rho}_0 \boldsymbol{\rho}_0^H$, and substituting $\mathbf{X} = \boldsymbol{\rho}\boldsymbol{\rho}^H$, we precisely obtain $\nabla \mathcal{J}(\boldsymbol{\rho}_k)$ derived in (3.7) when (3.10) is evaluated at $\boldsymbol{\rho} = \boldsymbol{\rho}_k$.

3.2. The distance metric and equivalence of convergence in signal and lifted domains.

While the view of GWF in the lifted problem and the formulation in (3.9) are illuminating, the algorithmic map of GWF operates exclusively on the signal domain in $\mathbb{C}^N$. The duality between the lifted domain and the signal domain is established by the distance metric of WF framework, which is defined as follows [12].

[2]We use the property of Wirtinger derivatives in writing the update (3.10), such that the derivative of a real-valued function of a complex variable has the property that $\overline{(\frac{\partial \mathcal{J}}{\partial \mathbf{X}})} = (\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{X}}})$

Definition 3.2. *Let $\boldsymbol{\rho}_t \in \mathbb{C}^N$ be an element of the global solution set in* (2.1)*. The distance of an element $\boldsymbol{\rho} \in \mathbb{C}^N$ to $\boldsymbol{\rho}_t$ is defined as* [12]

$$\text{(3.12)} \qquad \text{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t) = \|\boldsymbol{\rho} - e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t\|, \ \textit{where} \ \ \Phi(\boldsymbol{\rho}) := \underset{\phi \in [0,2\pi]}{\text{argmin}} \|\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\phi}\|.$$

For the purpose of convergence, the metric implies that we are primarily interested in the distance of an estimate $\boldsymbol{\rho}$ to *any* of the elements in the nonconvex solution set P. In technical terms, invoking Definition 2.2, (3.12) is a measure of distance between the equivalence sets $P_{\boldsymbol{\rho}}$ and P. By Definition 3.2, the ambiguity due to the invariance of the cross-correlation map, $\mathcal{L}$, to the global phase factors is evaded on $\mathbb{C}^N$, without lifting the problem. Observe that the phase ambiguity is indeed removed analytically, since the ℓ_2 norm is minimized when $\text{Re}(\langle \boldsymbol{\rho}, e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t\rangle) = |\langle \boldsymbol{\rho}, e^{j\Phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t\rangle| = |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|$, which is achieved at $e^{\mathrm{i}\Phi(\boldsymbol{\rho})} = \frac{\langle \boldsymbol{\rho}_t, \boldsymbol{\rho}\rangle}{|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|}$. Hence, the squared distance becomes

$$\text{(3.13)} \qquad \text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) = \|\boldsymbol{\rho}\|^2 + \|\boldsymbol{\rho}_t\|^2 - 2|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|$$

and is independent of any global phase factor on $\boldsymbol{\rho}$ or $\boldsymbol{\rho}_t$. Geometrically, the metric suggests that the cosine of the angle θ between the elements $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_t$ is given by

$$\text{(3.14)} \qquad \cos\theta = \frac{|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|}{\|\boldsymbol{\rho}\|\|\boldsymbol{\rho}_t\|},$$

which can be viewed as the angle between the subspaces spanned by the elements $\boldsymbol{\rho}\boldsymbol{\rho}^H$ and $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ in the lifted domain $\mathbb{C}^{N\times N}$. This can be seen by evaluating the error between the lifted terms as follows:

$$\text{(3.15)} \qquad \|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F^2 = \|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2 + \|\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F^2 - 2\text{Re}\left(\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H, \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\rangle_F\right).$$

For two rank-1 arguments, the Frobenius inner product in (3.15) reduces to $|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|^2$, and the cosine of the angle between the elements becomes equal to $\cos^2(\theta)$. Since (3.14) is nonnegative, the relationship between the two angles is one-to-one. Therefore, WF distance metric can be interpreted as a measure of distance between the lifted variables, and the convergence with respect to the metric (3.12) in signal domain is equivalent to convergence with respect to (3.15) in the lifted space.

3.3. Theoretical advantages of GWF via the lifted perspective. The lifted perspective of GWF reveals its connection to convex LRMR methods, such as those proposed in [20] and [33], as the objective in (3.8) simply corresponds to solving the quadratic data-fit over a more exclusive search space. In the scope of lifting-based approaches, we discussed Uzawa's method as a solver for the trace regularized problem in (2.11). Consider solving (3.8) by Uzawa's method described in (2.12)–(2.13), which essentially reduces to projected gradient descent: a gradient step over the smooth ℓ_2 mismatch term is followed by a projection onto the intersection of the PSD cone and the set of rank-1 matrices. While in general projections to the intersection of two sets is an optimization problem on its own, and the rank-1 constraint constitutes a nonconvex manifold, there exists a simple projection onto the set $\mathcal{X} = \{\mathbf{X} \in \mathbb{C}^{N\times N} : \text{rank}(\mathbf{X}) = 1 \cap \ \mathbf{X} \succeq 0\}$ such that

$$\mathcal{P}_{\mathcal{X}} = \mathcal{P}_{r=1} \circ \mathcal{P}_{+}, \tag{3.16}$$

which is the successive operation of the projection onto the PSD cone, followed by a rank-1 approximation.

Although a solver can be formulated, recovery guarantees of Uzawa's method do not cover this case due to the nonconvexity of the projected set. Simply, a gradient step over the convex PSD cone is not guaranteed to improve the rank-1 approximation of an estimate, which is after all the original motivation behind pursuing the convex solvers. Essentially, the GWF framework circumvents convexification and presents an alternative update scheme to Uzawa's for minimizing the objective in (3.8). This alternative update form stems from the fact that the PSD rank-1 constraint of the lifted unknown can be enforced by a variable transformation in the update equation, rather than by projections given in (3.16).

Clearly, the immediate advantage of the GWF formulation in (3.9) over the convex relaxations is the dimensionality reduction of the search space, attributed to iterating in the signal domain. There is yet another significant advantage to the GWF formulation relating to exact recovery guarantees. By (3.9), an iterative scheme for the unrelaxed, nonconvex form of the lifted problem is formulated, which enforces the rank-1, PSD structure on the iterates. This allows the constraint set to be considerably *smaller* than that of the trace relaxation or the convex feasibility problems. Formally, the problem (3.9) has a unique solution if the equivalence set of any $\boldsymbol{\rho} \in \mathbb{C}^N$ is its equivalence class under the correlation map $\mathcal{L}$ as defined in Definition 2.2.

Condition 3.3 (uniqueness condition). *There exists a unique solution $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H \in \mathcal{X}$ for the problem in* (3.9) *if*

$$P_{\boldsymbol{\rho}} = \mathcal{E}_{\boldsymbol{\rho}} \quad \forall \boldsymbol{\rho} \in \mathbb{C}^N. \tag{3.17}$$

In other words, there should exist no element H in the null space of $\mathcal{F}$, such that $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H + H$ is a rank-1, PSD matrix. Therefore, for exact interferometric inversion by GWF, the null space condition of the lifted forward model has to hold over a much less restrictive set than any of the approaches discussed in section 2.2.

3.4. The spectral method for GWF initialization. Having to solve a nonconvex problem, exact recovery guarantees of the WF framework depend on the accuracy of the initial estimate $\boldsymbol{\rho}_0$. The initial estimate of the standard WF algorithm is computed by the spectral method which corresponds to the leading eigenvector of the following PSD matrix:

$$\mathbf{Y} = \frac{1}{M}\sum_{m=1}^{M} d^m \mathbf{L}^m (\mathbf{L}^m)^H, \tag{3.18}$$

where $\mathbf{L}^m = \mathbf{L}_i^m = \mathbf{L}_j^m$, and $d^m = |\langle \mathbf{L}^m, \boldsymbol{\rho}_t\rangle|^2$ for $i = j$. The leading eigenvector is scaled by the square root of the corresponding largest eigenvalue λ_0 of $\mathbf{Y}$. In [12], the spectral method is described from a stochastic perspective. By the strong law of large numbers, under the assumption that we have $\mathbf{L}^m \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I}) + \mathrm{i}\mathcal{N}(0, \frac{1}{2}\mathbf{I}), m = 1, \ldots, M$, the spectral matrix $\mathbf{Y}$ becomes equal to

$$\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M} d^m \mathbf{L}^m (\mathbf{L}^m)^H\right] = \|\boldsymbol{\rho}_t\|^2 \mathbf{I} + \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \tag{3.19}$$

as $M \to \infty$, which has the true solution $\boldsymbol{\rho}_t$ as its leading eigenvector. The concentration of the spectral matrix around its expectation is used to show that the leading eigenvector of $\mathbf{Y}$ is *sufficiently* accurate, such that the sequence of iterates $\{\boldsymbol{\rho}_k\}$ of (3.1) converges to an element in the global solution set P.

In developing the GWF framework, we view the spectral method as a procedure in the lifted domain. In fact, we observe that the spectral matrix of phase retrieval in (3.18) is the backprojection estimate of the lifted unknown $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$. Having different measurement vectors $\mathbf{L}_i^m$ and $\mathbf{L}_j^m$ in the cross-correlated measurement case, using the definition of the backprojection operator in (3.6), we extend (3.18) and redefine $\mathbf{Y}$ as follows:

$$\mathbf{Y} = \frac{1}{M}\mathcal{F}^H(\mathbf{d}) = \frac{1}{M}\sum_{m=1}^{M} d_{ij}^m \mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H. \tag{3.20}$$

As noted, the true solution $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ of the lifted problem lies in the PSD cone. In standard WF for phase retrieval, the spectral matrix $\mathbf{Y}$ is formed by summation of PSD outer products that are scaled by $\mathbb{R}^+$ valued measurements $\{d^m\}_{m=1}^M$, as autocorrelations are by definition squared magnitudes. Hence, the WF spectral method generates an estimate of the lifted unknown within the constraint set by default. This obviously is not the case for the backprojection estimate (3.20) with the cross-correlated measurement model. Therefore, the extension of the spectral method to cross-correlations includes a projection step onto the PSD cone. Since the PSD cone is convex, its projection operator is nonexpansive and yields a closer estimate to $\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ than that of (3.20). The GWF spectral matrix then becomes

$$\hat{\mathbf{X}} := \frac{1}{2M}\sum_{m=1}^{M} d_{ij}^m \mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H + \overline{d_{ij}^m}\mathbf{L}_j^m \left(\mathbf{L}_i^m\right)^H. \tag{3.21}$$

We discard the positive semidefinitivity in (3.21) and only project onto the set of symmetric matrices, which is also convex. This is simply because only the leading eigenvector will be kept from the generated lifted estimate, which is unaffected by the projection onto the PSD cone.[3] Letting $\lambda_0, \mathbf{v}_0$ denote the leading eigenvalue-eigenvector pair of $\hat{\mathbf{X}}$, the GWF initial estimate $\boldsymbol{\rho}_0$ is given by[4]

$$\boldsymbol{\rho}_0 = \sqrt{\lambda_0}\mathbf{v}_0. \tag{3.22}$$

Using the representation in the lifted problem in (3.20) and plugging in (2.6) for the measurements, the GWF spectral matrix $\hat{\mathbf{X}}$ can be written as

$$\hat{\mathbf{X}} := \frac{1}{M}\mathcal{P}_S\left(\mathcal{F}^H(\mathbf{d})\right) = \frac{1}{M}\mathcal{P}_S\left(\mathcal{F}^H\mathcal{F}(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H)\right). \tag{3.23}$$

[3]Unless the leading eigenvalue is negative, a scenario that is excluded due to the conditions for exact recovery in section 4.

[4]Note that (3.21) yields a symmetric matrix and hence has eigenvalues $\lambda_i \in \mathbb{R}$.

Hence, the leading eigenvalue-eigenvector extraction corresponds to keeping the rank-1, PSD approximation, $\tilde{\boldsymbol{\rho}}_0 := \boldsymbol{\rho}_0\boldsymbol{\rho}_0^H$, of the backprojection estimate in the lifted domain. Furthermore, by definition in (3.23), the accuracy of the spectral estimate fully hinges on the properties of the normal operator of the lifted problem, i.e., $\mathcal{F}^H\mathcal{F}$ over the set of rank-1, PSD matrices, $\mathcal{X}$.

As a comparison, consider the LRMR approach for the interferometric inversion by the PSD constrained singular value thresholding algorithm in [33]. It can be seen that, setting $\mu_k = 1/M$ in (2.13), the first iterate generated by (2.12) is $\hat{\mathbf{X}} = \mathcal{P}_\tau(\frac{1}{M}\mathcal{F}^H\mathbf{d})$, which, prior to the singular value thresholding, is identical to the backprojection estimate computed by the spectral method. Hence, the spectral method simply differs from the first Uzawa iteration by keeping the rank-1 approximation via the projection operator $\mathcal{P}_{\mathcal{X}}$ defined in (3.16), instead of a low rank approximation. This precisely corresponds to the first Uzawa iteration to solve the nonconvex rank-1 constrained problem in (3.8).

3.5. Algorithm summary and computational complexity. Compared to the lifting-based LRMR approaches, GWF provides significant reductions in computational complexity and memory requirements per iteration. As shown in section 3.4, GWF uses the first iteration of the Uzawa's method to compute an initial estimate $\boldsymbol{\rho}_0$ and replaces the following iterations over the lifted domain with iterations on the leading eigenspace. The GWF algorithm is summarized as follows:

- Input: Interferometric measurements d_{ij}^m and measurement vectors $\mathbf{L}_i^m, \mathbf{L}_j^m$ $\forall i \neq j$, $m = 1, \dots, M$.
- Initialization: Run Uzawa's method in (2.12) initialized with $\mathbf{y} = 0$, $\mu_0 = 1/M$, and $\lambda = 0$, i.e., trace regularization free, for 1 iteration, yielding
$$\hat{\mathbf{X}} = \frac{1}{M}\mathcal{P}_S\left(\mathcal{F}^H(\mathbf{d})\right).$$
Keep the rank-1 approximation $\lambda_0\boldsymbol{\rho}_0\boldsymbol{\rho}_0^H$. The initialization step consists of the outer product of the two measurement vectors for each of the M samples, resulting in $\mathcal{O}(MN^2)$ multiplications, followed by an eigenvalue decomposition with $\mathcal{O}(N^3)$ complexity.
- Iterations: Perform gradient descent updates as $\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \frac{\mu_{k+1}}{\|\boldsymbol{\rho}_0\|^2}\nabla\mathcal{J}(\boldsymbol{\rho}_k)$, with
$$\nabla\mathcal{J}(\boldsymbol{\rho}_k) = \frac{1}{M}\mathcal{P}_S\left(\mathcal{F}^H(\mathbf{e}_k)\right)\boldsymbol{\rho}_k,$$
where $(\mathbf{e}_k)^m = ((\mathbf{L}_i^m)^H\,\boldsymbol{\rho}_k\boldsymbol{\rho}_k^H\mathbf{L}_j^m - d_{ij}^m)$.
Each iteration requires the following operations:
    1. Computing and storing the linear terms $(\mathbf{L}_{i,j}^m)^H\boldsymbol{\rho}_k$, requiring M number of N multiplications for each, resulting in $\mathcal{O}(MN)$ multiplications.
    2. Computing the error by cross-correlating linear terms, requiring $\mathcal{O}(M)$ multiplications.
    3. Multiplication of the linear terms $(\mathbf{L}_{i,j}^m)^H\boldsymbol{\rho}_k$ and the error e_{ij}^m for each $m = 1, \dots, M$, requiring $\mathcal{O}(M)$ multiplications.
    4. Multiplication of the result in operation 3 with vectors $\{\mathbf{L}_i^m\}_{m=1}^M$ and $\{\mathbf{L}_j^m\}_{m=1}^M$, requiring $\mathcal{O}(MN)$ multiplications.

  These operations result in $\mathcal{O}(MN)$ multiplications for each iteration.

4. Theory of exact interferometric inversion via GWF. In this section we present our exact recovery guarantees for interferometric inversion by GWF. Notably, we merge the exact recovery conditions of standard Wirtinger flow that rely on statistical properties of the sampling vectors into a single condition on the lifted forward model in the context of interferometric inversion ($i \neq j$). In Theorem 4.6, we state that if $\mathcal{F}$ satisfies the RIP on the set of rank-1, PSD matrices with a sufficiently small RIC, GWF is guaranteed to recover the true solution up to a global phase factor for any $\boldsymbol{\rho}_t \in \mathbb{C}^N$. Following Theorem 4.6, we establish the validity of our condition in Theorem 4.9 for the case of $\mathcal{O}(N \log N)$ measurements that are cross-correlations of i.i.d. complex Gaussian sampling vectors.

We begin by introducing the definitions of some concepts that appear in our theorem statements.

4.1. The ϵ-neighborhood of P and the regularity condition.

Definition 4.1 (ϵ-neighborhood of P). *We denote the ϵ-neighborhood of the global solution set P in (2.1) by $E(\epsilon)$ and define it as follows [12]:*

$$E(\epsilon) = \left\{ \boldsymbol{\rho} \in \mathbb{C}^N : \text{dist}(\boldsymbol{\rho}, P) \leq \epsilon \right\}. \tag{4.1}$$

The set $E(\epsilon)$ is determined by the distance of the spectral initialization to the global solution set, i.e., $\epsilon = \text{dist}(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t)$. The main result of the standard WF framework is that, for Gaussian and coded diffraction pattern measurement models [12], ϵ is sufficiently small so that the objective function $\mathcal{J}$ satisfies the following regularity condition.

Condition 4.2 (regularity condition). *The objective function $\mathcal{J}$ satisfies the regularity condition if $\forall \boldsymbol{\rho} \in E(\epsilon)$, the following holds:*

$$\text{Re}\left(\left\langle \nabla \mathcal{J}(\boldsymbol{\rho}), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})} \right) \right\rangle \right) \geq \frac{1}{\alpha} \text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) + \frac{1}{\beta} \|\nabla \mathcal{J}(\boldsymbol{\rho})\|^2 \tag{4.2}$$

for fixed $\alpha > 0$ and $\beta > 0$ such that $\alpha\beta > 4$.

The regularity condition guarantees that the iterations in (3.4) are contractions with respect to the distance metric (3.12), which ensures that all gradient descent iterates remain in $E(\epsilon)$, which is established by the following lemma from [12].

Lemma 4.3. *Assume that $\mathcal{J}$ obeys the regularity condition in (4.2) for some fixed α, β $\forall \boldsymbol{\rho} \in E(\epsilon)$. Having $\boldsymbol{\rho}_0 \in E(\epsilon)$, and assuming $\mu \leq 2/\beta$, consider the following update:*

$$\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \mu \nabla \mathcal{J}(\boldsymbol{\rho}_k). \tag{4.3}$$

Then, $\forall k$ we have $\boldsymbol{\rho}_k \in E(\epsilon)$ and

$$\text{dist}^2(\boldsymbol{\rho}_k, \boldsymbol{\rho}_t) \leq \left(1 - \frac{2\mu}{\alpha}\right)^k \text{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t).$$

Proof. See [12, Lemma 7.10]. ∎

Furthermore, from the definition of $\nabla \mathcal{J}$ in (3.4), the regularity condition implies that there exists no $\boldsymbol{\rho} \in E(\epsilon)$ that belongs to the equivalence class of $\boldsymbol{\rho}_t$ under the mapping $\mathcal{L}$.

Hence, the uniqueness condition for exact recovery is satisfied locally, and (4.2) is a sufficient condition by Lemma 7.1 of [12] such that the algorithm iterates $\{\boldsymbol{\rho}_k\}$ converge to P at a geometric rate. The spectral initialization is said to be sufficiently accurate, if (4.2) holds $\forall \boldsymbol{\rho} \in E(\epsilon)$.

4.2. Sufficient conditions for exact recovery. The condition we assert on the lifted forward model $\mathcal{F}$ is the RIP on the set of rank-1, PSD matrices, $\mathcal{X}$.

Definition 4.4 (restricted isometry property). *Let $\mathcal{A} : \mathbb{C}^{K\times N} \to \mathbb{C}^M$ denote a linear operator. Without loss of generality assume $K \leq N$. For every $1 \leq r \leq K$, the r-restricted isometry constant is defined as the smallest $\delta_r < 1$ such that*

$$(1-\delta_r)\|\mathbf{X}\|_F^2 \leq \|\mathcal{A}(\mathbf{X})\|^2 \leq (1+\delta_r)\|\mathbf{X}\|_F^2 \tag{4.4}$$

holds for all matrices $\mathbf{X}$ of rank at most r, where $\|\mathbf{X}\|_F = \sqrt{Tr(\mathbf{X}^H\mathbf{X})}$ denotes the Frobenius norm.

Suppose (4.4) holds $\forall \mathbf{X} \in \mathcal{D} \subset \mathbb{C}^{K\times N}$ that have rank-r with some constant $0 < \delta_r < 1$; then $\mathcal{A}$ is said to satisfy the RIP$_{\mathcal{D}}$ with RIC-δ_r. For the interferometric inversion problem, having $K = N$, if there exists $\delta_1 < 1$ such that $\mathcal{F}$ in (2.6) satisfies the RIP on the PSD cone, we say that the lifted forward model satisfies RIP$_+$ with RIC-δ_1, i.e., RIP over the set of rank-1, PSD matrices, $\mathcal{X}$.

(4.4) quantifies how close $\mathcal{F}^H\mathcal{F}$ is to an identity over rank-1, PSD matrices through the following lemma.

Lemma 4.5. *Suppose $\mathcal{F}$ satisfies RIP on the set rank-1, PSD matrices of size $N \times N$, i.e., $\mathcal{X} = \{\boldsymbol{\rho}\boldsymbol{\rho}^H : \boldsymbol{\rho} \in \mathbb{C}^N\}$ with RIC-δ_1. Then, for any $\mathbf{X} \in \mathcal{X}$ we have*

$$\left(\mathcal{F}^H\mathcal{F} - \mathbf{I}\right)(\mathbf{X}) = \boldsymbol{\delta}(\mathbf{X}), \tag{4.5}$$

where $\boldsymbol{\delta} : \mathcal{X} \to \mathbb{C}^M$ is a bounded operator such that

$$\|\boldsymbol{\delta}\|_{\mathcal{X}} := \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\setminus\{\mathbf{0}\}} \frac{\|\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|} = \delta_1,$$

where $\|\cdot\|$ denotes the spectral norm.

Proof. See Appendix A.2. ■

In the GWF framework, we identify RIP over rank-1, PSD matrices with RIC-$\delta_1 \leq 0.214$ as a sufficient condition for an arbitrary lifted forward model $\mathcal{F}$, which guarantees that the spectral initialization of GWF provides an initial estimate that is sufficiently accurate, i.e., GWF iterates are guaranteed to converge to a global solution in P starting from $\boldsymbol{\rho}_0$.

Theorem 4.6 (exact recovery by GWF). *Assume the lifted forward model $\mathcal{F}$[5] satisfies the RIP condition over rank-1, PSD matrices with RIC-δ_1. Then, the initial estimate $\boldsymbol{\rho}_0$ obtained from the spectral method by (3.21) and (3.22) satisfies*

[5]Up to a normalization factor, such as $\frac{1}{\sqrt{M}}$.

$$\operatorname{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t) \leq \epsilon^2 \|\boldsymbol{\rho}_t\|^2, \tag{4.6}$$

where ϵ *is an* $\mathcal{O}(1)$ *constant with* $\epsilon^2 = (2+\delta_1)(1-\sqrt{1-\frac{\delta_1}{1-\delta_1}})+\frac{\delta_1^2}{8}$*, and the regularity condition in* (4.2) *surely holds for the objective function in* (2.1) *if* $\delta_1 \leq 0.214$*, with any* α*,* β *satisfying*

$$\frac{1}{\alpha\|\boldsymbol{\rho}_t\|^2} + \frac{c^2(\delta_1)\|\boldsymbol{\rho}_t\|^2}{\beta} \leq h(\delta_1) := (1-\delta_2)(1-\epsilon)(2-\epsilon), \tag{4.7}$$

where $\delta_2 = \frac{\sqrt{2}(2+\epsilon)\delta_1}{\sqrt{(1-\epsilon)(2-\epsilon)}}$ *and* $c(\delta_1) = (2+\epsilon)(1+\epsilon)(1+\delta_1)$*. Thus, for the iterations of* (4.3) *with the update term in* (3.4) *and the fixed step size* $\mu \leq 2/\beta$*, we have*

$$\operatorname{dist}^2(\boldsymbol{\rho}_k, \boldsymbol{\rho}_t) \leq \epsilon^2 \left(1 - \frac{2\mu}{\alpha}\right)^k \|\boldsymbol{\rho}_t\|^2. \tag{4.8}$$

Proof. See section 5.1. ∎

We refer to δ_2 as the RIC of the *local* RIP-2 condition by Lemma 5.5 which is exclusively over elements of the form $\{\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H : \boldsymbol{\rho} \in E(\epsilon)\}$ and c as the local Lipschitz constant of $\nabla\mathcal{J}$ by Lemma 5.6, both stated in section 5.

Remark 4.7. We summarize the implications of Theorem 4.6 by the following remarks.

1. Theorem 4.6 establishes a regime in which the regularity condition holds by default as a result of the RIP over rank-1, PSD matrices, captured by (4.7). This regime also defines the range of values the RIC-δ_1 can attain, through the quantity δ_2 which must satisfy $\delta_2 < 1$, as shown in Figure 1. Notably, having $\delta_2 < 1$ guarantees the uniqueness of a solution locally in $E(\epsilon)$ and the restricted strong convexity of $\mathcal{J}$ since there exists a fixed α, β satisfying (4.7). Numerically, plugging in the ϵ constant defined by δ_1, this is satisfied for $\delta_1 \leq 0.214$.
2. Figure 1 demonstrates the values the constants c and h attain in the valid range for δ_1. These $\mathcal{O}(1)$ constants directly impact the convergence rate of the algorithm, as α and β are required to be sufficiently large values for (4.7) to hold. Observe that (4.7) implies setting $\alpha = \alpha'/\|\boldsymbol{\rho}_t\|^2$, and $\beta = \beta'\|\boldsymbol{\rho}_t\|^2$, where $\alpha', \beta' = \mathcal{O}(1)$, hence $\alpha\beta = \mathcal{O}(1)$. Since

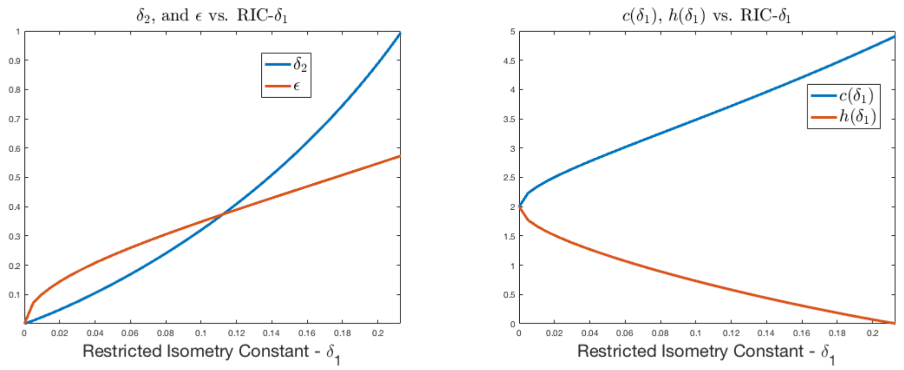


Figure 1. *Numerical evaluation of* ϵ*, RIC-*δ_2 *and* c*,* h *values with respect to the RIC-*$\delta_1 \leq 0.214$ *of the RIP on the set of rank-*1 *PSD matrices.*

clearly $h < 2$, and $c^2 > 4$, and the regularity condition can at best hold when $\alpha'\beta' > 4$ by definition. Therefore, the RIP over rank-1, PSD matrices with RIC-$\delta_1 \leq 0.214$ is a sufficient condition for the exact recovery via GWF.

3. Since $\|\boldsymbol{\rho}_t\|$ is unknown a priori, a suitable approximate scaling in setting constants α, β is $\|\boldsymbol{\rho}_0\|^2$. This scaling factor is merely the leading eigenvalue λ_0 of the spectral matrix in (3.21). From Lemma 5.2, for $\mathcal{F}$ satisfying RIP over rank-1, PSD matrices with the RIC-δ_1, λ_0 is lower bounded by $1 - \delta_1$; hence the condition in (4.7) can be enforced by setting $\alpha', \beta' = \mathcal{O}(1)$ such that

$$\frac{1+\delta_1}{\alpha'} + \frac{c^2(\delta_1)}{(1-\delta_1)\beta'} \leq h(\delta_1).$$

Similarly, picking $\mu \leq \frac{2}{\beta}$ to yield convergence rate of $\frac{2\mu}{\alpha} \leq \frac{4}{\alpha'\beta'}$, the iterations in (4.3) must have a step size μ that is $\mathcal{O}(1/\|\boldsymbol{\rho}_t\|^2)$. This is essentially where the normalization term in (3.1) originates from, with $\|\boldsymbol{\rho}_0\|^2$ serving as an approximation to $\|\boldsymbol{\rho}_t\|^2$. As a result, Lemma 4.3 for (3.1) simply holds for $\mu_k \leq (1-\delta_1)\frac{2}{\beta'}$ due to the effect of the mismatch in the scaling factors, in agreement with what is noted in [12].

4. Overall, Theorem 4.6 establishes that the convergence speed of the GWF algorithm is controlled by RIC-δ_1. As δ_1 approaches the critical limit of 0.214, c and δ_2 values increase superlinearly as shown in Figure 1. This has strong implications on the convergence speed, as β' is inversely proportional to the choice of step sizes μ_k and is quadratically related to the magnitude of c.

5. A consequence of our result is a universal upper bound on ϵ under our sufficient condition of RIP over rank-1, PSD matrices. As depicted in Figure 1, Theorem 4.6 determines what sufficiently close means numerically.

Remark 4.8. Despite solving the identical perturbed problem, GWF iterations provably converge, whereas the convergence guarantees of Uzawa's method vanish due to inclusion of a nonconvex constraint. Similarly, the special structure of the constraint set and the GWF iterates suffice the RIP condition to be satisfied only over rank-1 matrices in the PSD cone, whereas the uniqueness condition of Uzawa's method requires RIP over the set of rank-2 matrices in the nonconvex rank minimization problem.

4.3. Restricted isometry property for cross-correlation of Gaussian measurements. In standard WF for phase retrieval with the i.i.d. complex Gaussian model, i.e., $\mathbf{L}_m \sim \mathcal{N}(0, \mathbf{I}/2) + \mathrm{i}\mathcal{N}(0, \mathbf{I}/2)$, the accuracy of the spectral estimate (3.18) is established as

$$\|\mathbf{Y} - (\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H + \|\boldsymbol{\rho}_t\|^2\mathbf{I})\| \leq \delta\|\boldsymbol{\rho}_t\|^2 \tag{4.9}$$

with probability $1 - 10e^{\gamma N} - 8/N^2$, where γ is a fixed positive numerical constant. This result is derived from the concentration bound of the Hessian of the objective function around its expectation at a global minimizer $\boldsymbol{\rho}_t$ such that

$$\|\nabla^2\mathcal{J}(\boldsymbol{\rho}_t) - \mathbb{E}[\nabla^2\mathcal{J}(\boldsymbol{\rho}_t)]\| \leq \delta\|\boldsymbol{\rho}_t\|^2, \tag{4.10}$$

where δ is the concentration bound.

In the problem of phase retrieval, plugging the definition of phaseless measurements into (3.18), the autocorrelations yield the fourth moments of the elements of the Gaussian measurement vectors. This introduces a bias of $\|\boldsymbol{\rho}_t\|^2\mathbf{I}$ in the spectral estimate $\mathbf{Y}$, as can be seen in (4.9). Moving from the autocorrelations to cross-correlations removes this bias component from the spectral matrix of GWF. Hence, for the Gaussian model, cross-correlated measurement map, i.e., $\mathcal{F}$ when $i \neq j$, satisfies the RIP over the set of rank-1, PSD matrices.

Without loss of generality and following [12], we present our result for the case $\|\boldsymbol{\rho}_t\| = 1$.

Theorem 4.9 (RIP over rank-1, PSD matrices for cross-correlated Gaussian measurements). *Let the measurement vectors $\mathbf{L}_i^m, \mathbf{L}_j^m$ in (1.2) follow the i.i.d. complex Gaussian model, i.e., $\mathbf{L}_i^m, \mathbf{L}_j^m \sim \mathcal{N}(0, \mathbf{I}/2) + \mathrm{i}\mathcal{N}(0, \mathbf{I}/2)$. Then, the lifted forward model $\frac{1}{\sqrt{M}}\mathcal{F}$ in (2.6) for cross-correlated measurements (when $i \neq j$) satisfies RIP defined in (4.4) for $r = 1$ over the PSD cone, with probability $1 - 8e^{-\gamma N} - 5/N^2$ given $\mathcal{O}(N \log N)$ measurements, where γ is a fixed positive numerical constant. Moreover, the spectral matrix $\hat{\mathbf{X}}$ defined in (3.21) satisfies*

$$\|\hat{\mathbf{X}} - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\| \leq \delta_1, \tag{4.11}$$

where $\boldsymbol{\rho}_t$ is the ground truth signal with $\|\boldsymbol{\rho}_t\| = 1$ and δ_1 is the RIC of $\frac{1}{\sqrt{M}}\mathcal{F}$.

Proof. See section 5.2. ■

Remark 4.10. Outcomes of Theorem 4.9 are explained with the following remarks.

1. Theorem 4.9 establishes the relationship between the concentration bound of the spectral matrix and the RIP-1 condition for interferometric inversion. This indicates that the regularity condition of the WF framework is *redundant* for our problem if δ_1 is picked properly, since by Theorem 4.6, RIP$_+$ with RIC-$\delta_1 \leq 0.214$ directly implies the regularity condition.
2. Note that the equivalent linear model in the lifted domain actually has N^2 unknowns. By our formulation of GWF in the lifted problem, having measurements of the order of $N \log N$, therefore, corresponds to an underdetermined system of equations in which exact recovery guarantees of GWF hold.
3. Note that our measurement complexity is identical to that of standard WF. This is an expected result, as the cross-correlations impact only the removal of the diagonal bias in the spectral matrix $\mathbf{Y}$, not the concentration of $\mathbf{Y}$ around its expectation.
4. It should be reiterated that when $i = j$, the backprojection estimate $\mathbf{Y}$ contains a diagonal bias of the form $\|\boldsymbol{\rho}_t\|^2\mathbf{I}$, which breaks our RIP condition over rank-1, PSD matrices. In such a case, the concentration bound then holds around the expectation term that includes this diagonal bias. Establishing a sufficient condition for the phaseless case under our framework is not in the scope of this paper, and we leave it for future work.

Remark 4.11. It should be noted that phaseless measurements in the Gaussian model are known to satisfy a RIP under the ℓ_1 norm in the range of the lifted forward map. This is referred to as the RIP-1 condition in literature and is prominently featured in exact recovery theory of standard WF, more recently in [43] for nonconvex optimization via gradient descent, even in the presence of nonconvex regularizers.

A critical difference of our work is that such a RIP-type condition isn't used as a tool to obtain local curvature and local smoothness conditions to assert the regularity condition with a high probability. Instead, under the RIP studied in this paper, a regime is derived in which the regularity condition is guaranteed to hold, deterministically. Notably, for the interferometric inversion problem with $i \neq j$, δ_1 values tested in [12] are confidently within the range in which our GWF theory applies.

5. Proofs of Theorems 4.6 and 4.9. In this section, we present the proofs of Theorems 4.6 and 4.9. We first present key lemmas, and next prove the theorems using these lemmas. We provide the detailed proofs of the lemmas in Appendices B and C.

To prove Theorem 4.6, we begin by showing that the RIC-δ_1 of our RIP condition determines the distance ϵ of the spectral initialization in a one-to-one manner. We then establish that the regularity condition is directly implied by RIP$_+$ with RIC-$\delta_1 \leq 0.214$. In achieving this result we first show that the structure of the rank-1, PSD set allows for RIP to hold locally for the difference of two rank-1 PSD matrices, i.e., a local RIP-2 condition similar to the one in [31], with a RIC-δ_2. The upper bound on δ_1 ensures that RIC-δ_2 of the local RIP-2 satisfies $\delta_2 < 1$. The local RIP-2 condition, in turn, ensures that restricted strong convexity holds in the ϵ-neighborhood of the global solution set, which leads to exact recovery conditions of GWF.

For Theorem 4.9, we first show that the bias term in (4.9) resulting from the fourth moments of the random Gaussian entries disappears when we have cross-correlations instead of autocorrelations of measurements. We then establish that the spectral matrix is concentrated around its expectation, using the machinery in [12], adapted for cross-correlations. Finally, we use the definition of the spectral matrix to derive the RIP over rank-1, PSD matrices from the concentration bound, which yields the RIC-δ_1.

5.1. Proof of Theorem 4.6. Without loss of generality, we assume $\boldsymbol{\rho}_t$ is a solution with $\|\boldsymbol{\rho}_t\| = 1$. In establishing the exact recovery guarantees for GWF, we take a two-step approach. For a lifted forward map $\mathcal{F}$ satisfying RIP$_+$ with RIC-δ_1, we first show that the initialization by spectral method yields an estimate that is in the set $E(\epsilon)$. We then establish the regularity condition (4.2) for the objective function (2.1) in the ϵ-neighborhood defined by the initialization. These two results culminate into convergence to a global solution at a geometric rate as stated in Theorem 4.6.

5.1.1. ϵ-neighborhood of spectral initialization. Rather than the law of large numbers approach in [12], we take the geometric point of view of [54] in establishing the ϵ-neighborhood of the spectral initialization. We begin by evaluating the distance of the leading eigenvector $\mathbf{v}_0 \in \mathbb{C}^N$ of the spectral matrix in (3.21) to the global solution set (2.1). Recall the definition of the distance metric

$$\text{dist}(\mathbf{v}_0, \boldsymbol{\rho}_t) = \left\| \mathbf{v}_0 - e^{\mathrm{i}\Phi(\mathbf{v}_0)} \boldsymbol{\rho}_t \right\|, \tag{5.1}$$

which is essentially the Euclidean distance of $\mathbf{v}_0$ to the closest point in the solution set P in (2.1). Without loss of generality, we fix $\Phi(\mathbf{v}_0) = \Phi_0$ and incorporate it into $\boldsymbol{\rho}_t$ such that $\hat{\boldsymbol{\rho}}_t = e^{\mathrm{i}\Phi_0}\boldsymbol{\rho}_t$ represents the closest solution to $\mathbf{v}_0$ in P. We break down the key arguments of our proof into the following three lemmas.

Lemma 5.1. *Let $\boldsymbol{\rho}_t$ be a solution with $\|\boldsymbol{\rho}_t\| = 1$ and $\hat{\boldsymbol{\rho}}_t$ is the closest solution in P to $\mathbf{v}_0$. Then,* $\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle = \langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle$, *and*

$$\hat{\boldsymbol{\rho}}_t = \cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^{\perp}, \tag{5.2}$$

where $\|\mathbf{v}_0\| = 1$, $\cos(\theta) = \mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle$, *and $\mathbf{v}_0^{\perp}$ is a unit vector lying in a plane whose normal is $\mathbf{v}_0$. Similarly, there exists a perpendicular unit vector, $\hat{\boldsymbol{\rho}}_t^{\perp}$, to $\hat{\boldsymbol{\rho}}_t$ such that*

$$\hat{\boldsymbol{\rho}}_t^{\perp} = -\sin(\theta)\mathbf{v}_0 + \cos(\theta)\mathbf{v}_0^{\perp}. \tag{5.3}$$

Proof. See Appendix B.1. ■

Lemma 5.2. *Consider the spectral matrix $\hat{\mathbf{X}}$ given by (3.23), and denote the spectral matrix projected onto the PSD cone as $\hat{\mathbf{X}}_{PSD}$. Then, for a lifted mapping $\mathcal{F}$ satisfying RIP$_+$ with RIC-$\delta_1 = \delta$, $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}_{PSD}$ have the identical leading eigenvalue-eigenvector pair λ_0, $\mathbf{v}_0$ such that*

$$1 - \delta \le \lambda_0 \le 1 + \delta.$$

Furthermore, $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}_{PSD}$ generate identical spectral initialization, $\boldsymbol{\rho}_0$.

Proof. See Appendix B.2. ■

Lemma 5.2 allows us to analyze the distance of the initial estimate $\boldsymbol{\rho}_0$ to the solution set by the convenience of either the PSD $\hat{\mathbf{X}}_{PSD}$ or the symmetric spectral estimate $\hat{\mathbf{X}}$, since they generate the same initial estimate $\boldsymbol{\rho}_0$.

Next, using Lemmas 5.1 and 5.2 we reach the following key result.

Lemma 5.3. *In the setup of Lemmas 5.1 and 5.2, for the angle θ between the one-dimensional subspaces spanned by $\hat{\boldsymbol{\rho}}_t$ and $\mathbf{v}_0$ we have*

$$\sin^2(\theta) \le \frac{\delta}{1-\delta}, \tag{5.4}$$

where δ is the RIC-δ_1 of the lifted map $\mathcal{F}$ satisfying RIP$_+$.

Proof. See Appendix B.3. ■

From Lemma 5.3, we can now lower bound the inner product of $\hat{\boldsymbol{\rho}}_t$ and $\mathbf{v}_0$ such that

$$(\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle)^2 = \cos^2(\theta) = 1 - \sin^2(\theta) \ge 1 - \kappa,$$

where $\kappa = \frac{\delta}{1-\delta}$. Writing the distance of the spectral initialization $\boldsymbol{\rho}_0 = \sqrt{\lambda_0}\mathbf{v}_0$ to the solution set, we have

$$\mathrm{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t) = \lambda_0 + 1 - 2\mathrm{Re}\left\langle e^{\mathrm{i}\Phi(\boldsymbol{\rho}_0)}\boldsymbol{\rho}_t, \sqrt{\lambda_0}\mathbf{v}_0\right\rangle.$$

It is easy to see that $\mathrm{Re}\langle e^{\mathrm{i}\Phi(\boldsymbol{\rho}_0)}\boldsymbol{\rho}_t, \sqrt{\lambda_0}\mathbf{v}_0\rangle$ is maximized when $\Phi(\boldsymbol{\rho}_0) = \Phi(\mathbf{v}_0)$, hence we get

$$\mathrm{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t) = \lambda_0 + 1 - 2\sqrt{\lambda_0}\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle \le \lambda_0 + 1 - 2\sqrt{\lambda_0}\sqrt{1-\kappa}. \tag{5.5}$$

From Lemma 5.2, we know that $1 - \delta \le \lambda_0 \le 1 + \delta$. Moreover, the upper bound on the right-hand side of (5.5) is simply a quadratic term with respect to $\sqrt{\lambda_0}$ since $\sqrt{\lambda_0}(\sqrt{\lambda_0} - 2\sqrt{1-\kappa}) + 1$, which is maximized at the boundary of the domain of values $\sqrt{\lambda_0}$ takes. Since the quadratic equation is minimized at $\sqrt{1-\kappa}$ and we have $1-\kappa \leq 1-\delta$ for the domain of possible values of $0 \leq \delta < 1$, $\lambda_0 = 1+\delta$ is an upper bound for the right-hand side of (5.5). Hence, we obtain

$$\mathrm{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t) \leq 2 + \delta - 2\sqrt{1+\delta}\sqrt{1-\kappa}.$$

Writing the Taylor series expansion of $\sqrt{1+\delta}$ around 0, and discarding the components of order $\mathcal{O}(\delta^3)$ and higher, we have the final upper bound

$$\mathrm{dist}^2(\boldsymbol{\rho}_0, \boldsymbol{\rho}_t) \leq (2+\delta)\left(1-\sqrt{1-\kappa}\right) + \frac{\delta^2}{8}, \tag{5.6}$$

which sets the ϵ-neighborhood as

$$\epsilon^2 = (2+\delta)\left(1-\sqrt{1-\kappa}\right) + \frac{\delta^2}{8}. \tag{5.7}$$

5.1.2. Proof of the regularity condition. Recall that we seek a solution to the interferometric inversion problem by minimizing the loss function

$$\mathcal{J}(\boldsymbol{\rho}) = \frac{1}{2M}\sum_{m=1}^{M}\left|(\mathbf{L}_i^m)^H \boldsymbol{\rho}\boldsymbol{\rho}^H \mathbf{L}_j^m - d_{ij}^m\right|^2 \tag{5.8}$$

and address the optimization by forming the steepest descent iterates

$$\boldsymbol{\rho}^{k+1} = \boldsymbol{\rho}^k - \mu \nabla \mathcal{J}(\boldsymbol{\rho}^k), \tag{5.9}$$

where μ is the learning rate and $\nabla\mathcal{J}$ is the complex gradient defined by the Wirtinger derivatives.

As shown in section 3.1, the gradient evaluated at a point $\boldsymbol{\rho}$ can be expressed as

$$\nabla\mathcal{J}(\boldsymbol{\rho}) = \mathcal{Y}(\boldsymbol{\rho})\boldsymbol{\rho}, \tag{5.10}$$

where

$$\mathcal{Y}(\boldsymbol{\rho}) = \mathcal{P}_S\left(\mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}} - \tilde{\boldsymbol{\rho}}_t)\right) \tag{5.11}$$

with $\tilde{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\rho}}_t$ denoting the lifted variables $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}\boldsymbol{\rho}^H$ and $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$, respectively. Invoking Lemma 4.5 and the linearity of $\mathcal{F}^H\mathcal{F}$ and $\boldsymbol{\delta}$ of (4.5), (5.11) can be represented as

$$\mathcal{Y}(\boldsymbol{\rho}) = \mathcal{P}_S\left(\tilde{\boldsymbol{\rho}} - \tilde{\boldsymbol{\rho}}_t + \boldsymbol{\delta}(\tilde{\mathrm{e}})\right),$$

where $\tilde{\mathbf{e}} = \tilde{\boldsymbol{\rho}} - \tilde{\boldsymbol{\rho}}_t$ is the error in the lifted problem. Since the lifted variables are already symmetric we can take them out of the projection operator due to its linearity. Hence, for the update term we obtain

$$\nabla\mathcal{J}(\boldsymbol{\rho}) = \mathcal{Y}(\boldsymbol{\rho})\boldsymbol{\rho} = \tilde{\boldsymbol{\rho}}\boldsymbol{\rho} - \tilde{\boldsymbol{\rho}}_t\boldsymbol{\rho} + \mathcal{P}_S(\boldsymbol{\delta}(\tilde{\mathrm{e}}))\boldsymbol{\rho} \tag{5.12}$$

$$= \|\boldsymbol{\rho}\|^2\boldsymbol{\rho} - (\boldsymbol{\rho}_t^H\boldsymbol{\rho})\boldsymbol{\rho}_t + \mathcal{P}_S(\boldsymbol{\delta}(\tilde{\mathrm{e}}))\boldsymbol{\rho}. \tag{5.13}$$

Reprising the regularity condition under consideration, we need to establish that there exists constants α and β such that $\alpha\beta > 4$ $\forall \boldsymbol{\rho} \in E(\epsilon)$ and

$$\text{Re}\left(\left\langle \nabla \mathcal{J}(\boldsymbol{\rho}), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right\rangle\right) \geq \frac{1}{\alpha}\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) + \frac{1}{\beta}\|\nabla \mathcal{J}(\boldsymbol{\rho})\|^2. \tag{5.14}$$

To show the existence of constants α and β that satisfy (5.14), we upper bound the gradient term, which converts the regularity condition to a *restricted strong convexity condition* [17, 47]. We begin the proof by introducing the following key lemmas.

Lemma 5.4. *Let $\boldsymbol{\rho}_t$ be the ground truth signal with $\|\boldsymbol{\rho}_t\| = 1$, and let $\hat{\boldsymbol{\rho}}_t$ denote the global solution closest to $\boldsymbol{\rho}$ such that $\hat{\boldsymbol{\rho}}_t = e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t$. Then, for any $\boldsymbol{\rho} \in E(\epsilon)$, we have*

$$\left(\sqrt{(1-\epsilon)(2-\epsilon)}\right) \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\| \leq \|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F \leq (2+\epsilon)\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\|. \tag{5.15}$$

Proof. See Appendix B.4. ■

Lemma 5.5. *Let $\boldsymbol{\rho}_t$ be the ground truth signal with $\|\boldsymbol{\rho}_t\| = 1$, and let the linear map $\mathcal{F}$ satisfy RIP$_+$ with RIC-δ_1. Then, for $\delta_1 \leq 0.214$ and any $\boldsymbol{\rho} \in E(\epsilon)$, we have $\delta_2 < 1$ such that*

$$(1-\delta_2)\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F^2 \leq \|\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\|^2 \leq (1+\delta_2)\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F^2, \tag{5.16}$$

where $\delta_2 = \frac{\sqrt{2}(2+\epsilon)}{\sqrt{(1-\epsilon)(2-\epsilon)}}\delta_1$.

We refer to (5.16) as the local RIP-2 condition in the lifted domain with RIC-δ_2 for the mapping $\mathcal{F}$. The two lemmas culminate into the local Lipschitz continuity of $\nabla\mathcal{J}$.

Proof. See Appendix B.5. ■

Lemma 5.6. *In the setup of Lemmas 5.4 and 5.5, for any $\boldsymbol{\rho} \in E(\epsilon)$, the objective function $\mathcal{J}$ in (5.8) is Lipschitz differentiable with*

$$\|\nabla \mathcal{J}(\boldsymbol{\rho})\| \leq c \cdot \text{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t), \tag{5.17}$$

where $c = (1+\epsilon)(2+\epsilon)(1+\delta_1)$ is the Lipschitz constant. Furthermore, to establish the regularity condition for $\mathcal{J}$, it is sufficient to show that

$$\text{Re}\left(\left\langle \nabla \mathcal{J}(\boldsymbol{\rho}), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right\rangle\right) \geq \left(\frac{1}{\alpha} + \frac{c^2}{\beta}\right)\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) \tag{5.18}$$

for any $\boldsymbol{\rho} \in E(\epsilon)$.

Proof. See Appendix B.6. ■

We finally utilize the following lemma to obtain an alternative form of the restricted strong convexity condition [68].

Lemma 5.7. *For the objective function $\mathcal{J}$ in (5.8), the condition in (5.18) is satisfied if*

$$\mathcal{J}(\boldsymbol{\rho}) \geq \frac{\eta}{2}\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t), \tag{5.19}$$

where $\eta = \frac{1}{\alpha} + \frac{c^2}{\beta}$.

Proof. See Appendix B.7. ■

Writing the objective function explicitly in terms of the lifted terms, and applying the lower bound from the local RIP-2 condition of the lifted forward model, we can express the regularity condition simply as

$$\frac{1}{2}\left\|\mathcal{F}\left[\boldsymbol{\rho}\boldsymbol{\rho}^H-\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right]\right\|^2 \geq \frac{(1-\delta_2)}{2}\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H-\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F^2 \geq \frac{\eta}{2}\mathrm{dist}^2(\boldsymbol{\rho},\boldsymbol{\rho}_t),$$

where δ_2 is as defined in Lemma 5.5. From Lemma 5.4, the regularity condition is then satisfied by identifying α,β with $\alpha\beta > 4$ such that

$$(1-\delta_2)\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H-\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F^2 \geq (1-\delta_2)(1-\epsilon)(2-\epsilon)\mathrm{dist}^2(\boldsymbol{\rho},\boldsymbol{\rho}_t) \geq \left(\frac{1}{\alpha}+\frac{c^2}{\beta}\right)\mathrm{dist}^2(\boldsymbol{\rho},\boldsymbol{\rho}_t),$$

where, from Lemma 5.6, $c=(1+\epsilon)(2+\epsilon)(1+\delta_1)$.

5.2. Proof of Theorem 4.9.

Lemma 5.8 (expectation of spectral matrix). *Let measurement vectors $\mathbf{L}_i^m,\mathbf{L}_j^m$ be statistically independent and distributed according to the complex Gaussian model as $\mathbf{L}_i,\mathbf{L}_j^m \sim \mathcal{N}(0,\frac{1}{2}\mathbf{I})+\mathrm{i}\mathcal{N}(0,\frac{1}{2}\mathbf{I})$. Let $\boldsymbol{\rho}_t$ be independent of the measurement vectors, and let $\mathbf{Y}$ denote the backprojection estimate of the lifted signal generated by the spectral method given as*

$$\mathbf{Y}=\frac{1}{M}\mathcal{F}^H\mathcal{F}(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H),$$

where $\mathcal{F}$ is the lifted forward map in (2.6). Then,

$$\mathbb{E}[\mathbf{Y}]=\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H.$$

Proof. See Appendix C.1. ■

Lemma 5.9 (concentration around expectation). *In the setup of Lemma 5.8, assume that the number of measurements is $M=C(\delta)\cdot N\log N$, where C is a constant that depends on δ. Then,*

$$\|\mathbf{Y}-\mathbb{E}[\mathbf{Y}]\| \leq \delta \tag{5.20}$$

holds with probability at least $p=1-8e^{-\gamma N}-5N^{-2}$, where γ is a fixed positive constant.

Proof. See Appendix C.2. ■

Plugging the expectation from Lemma 5.8 into (5.20), and using the definition of $\boldsymbol{\delta}$ from Lemma 4.5, Lemmas 5.8 and 5.9 culminate to

$$\left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\| \leq \delta \tag{5.21}$$

for *any* $\boldsymbol{\rho}_t$ with $\|\boldsymbol{\rho}_t\|=1$ with probability at least p. From the definition of the spectral norm on $\mathbb{C}^{N\times N}$, we can write the left-hand side of (5.21) equivalently as

$$\max_{\boldsymbol{\rho}\in C^N,\|\boldsymbol{\rho}\|=1}\left|\boldsymbol{\rho}^H\boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\boldsymbol{\rho}\right| \leq \delta. \tag{5.22}$$

Since the spectral norm corresponds to the maximum over the unit sphere in $\mathbb{C}^N$, we have

$$\left|\boldsymbol{\rho}_t^H \boldsymbol{\delta}\left(\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\right) \boldsymbol{\rho}_t\right| \leq \left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\right)\right\| \leq \delta. \tag{5.23}$$

Observe that the left-hand side can equivalently be represented as a Frobenius inner product via lifting as

$$\left|\boldsymbol{\rho}_t^H \boldsymbol{\delta}\left(\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\right) \boldsymbol{\rho}_t\right| = \left|\left\langle \boldsymbol{\delta}\left(\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\right), \boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\right\rangle_F\right|. \tag{5.24}$$

Having Lemmas 5.8 and 5.9 hold for any element $\boldsymbol{\rho}_t \in \mathbb{C}^N$ with $\|\boldsymbol{\rho}_t\| = 1$ via unitary invariance, for any $\boldsymbol{\rho} \in \mathbb{C}^N$ we obtain

$$\frac{\left|\left\langle \left(\frac{1}{M}\mathcal{F}^H\mathcal{F} - \mathbf{I}\right)\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F\right|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2} \leq \delta, \tag{5.25}$$

which yields

$$(1-\delta)\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2 \leq \frac{1}{M}\left\|\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|^2 \leq (1+\delta)\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2 \tag{5.26}$$

for any $\boldsymbol{\rho} \in \mathbb{C}^N$. Therefore, RIP$_+$ with RIC-δ is established with probability at least p for mapping $\frac{1}{\sqrt{M}}\mathcal{F}$ with $\mathbf{L}_i^m, \mathbf{L}_j^m \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I}) + \mathrm{i}\mathcal{N}(0, \frac{1}{2}\mathbf{I})$, where $M = C(\delta) \cdot N \log N$.

Furthermore, we know that the true lifted unknown $\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H$ lies in the PSD cone, which is a convex set. Since the spectral matrix $\hat{\mathbf{X}}$ is the projection of $\frac{1}{M}\mathcal{F}^H\mathcal{F}(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H)$ onto the set of Hermetian symmetric matrices, from the nonexpansiveness property of projections onto convex sets we have

$$\left\|\hat{\mathbf{X}} - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\| \leq \left\|\frac{1}{M}\mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right) - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\| \leq \delta,$$

which completes the proof of Theorem 4.9.

6. Numerical simulations.

6.1. Signal recovery from random Gaussian measurements. We begin by considering recovery of random signals from cross-correlations of complex random Gaussian measurements, $\mathbf{L}_i^m, \mathbf{L}_j^m \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I}) + \mathrm{i}\mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$. For our numerical evaluations of the Gaussian model, we conduct an experiment similar to that of [12]. We set $N = 128$ and run 100 instances of interferometric inversion by GWF with independently sampled Gaussian measurement vectors on two types of signals: random low-pass signals, $\boldsymbol{\rho}^{LP}$, and random Gaussian signals, $\boldsymbol{\rho}^G$. The entries of the signals are generated independently of the measurement vectors at each instance by

$$\rho_l^{LP} = \sum_{p=-\frac{P}{2}+1}^{\frac{P}{2}} (X_l + \mathrm{i}Y_l) e^{\frac{2\pi \mathrm{i}(p-1)(l-1)}{N}}, \quad \rho_l^{G} = \sum_{p=-\frac{N}{2}+1}^{\frac{N}{2}} \frac{1}{\sqrt{8}} (X_l + \mathrm{i}Y_l) e^{\frac{2\pi \mathrm{i}(p-1)(l-1)}{N}}, \tag{6.1}$$

where $P = N/8$, and X_l and Y_l are i.i.d. $\mathcal{N}(0,1)$. As described in [12], a random low-pass signal corresponds to a bandlimited version of this random model and variances are adjusted so that the expected signal power is the same.

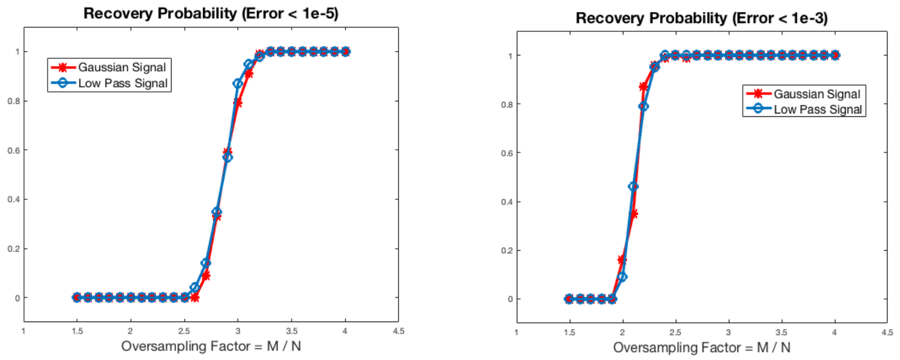


Figure 2. *Empirical recovery probabilities based on* 100 *random trials versus oversampling factor of the number of measurements* M/N*. The red curves correspond to empirical recovery probability of the Gaussian signal, whereas the blue curve corresponds to that of realization of the random low-pass signal model. The two figures vary with respect to the values of success criterion assumed for successful recovery, with a relative error of* 10^{-5} *and* 10^{-3}*, respectively.*

We implement the GWF algorithm with the learning rate heuristic of WF in [12] such that the descent algorithm takes smaller steps initially due to higher inaccuracy of the iterates. The step size is gradually increased such that $\mu_k = \min(1 - e^{k/\tau_0}, \mu_{max})$, where $\tau_0 = 33000$, and $\mu_{max} = 0.2$. For 2500 iterations, the learning parameter corresponds to a nearly linear regime and attains the maximum value of 0.073.

In the experimentation, we compute the empirical probability of success after 2500 iterations by counting the exact recovery instances of GWF recovery from different realizations of Gaussian measurements for $\{\mathbf{L}_i^m, \mathbf{L}_j^m\}_{m=1}^M$. We evaluate the exact recovery by the relative normalized error of the final estimate, $\boldsymbol{\rho}_{GWF}$, such that $\text{dist}(\boldsymbol{\rho}_{GWF}, \boldsymbol{\rho}_t)/\|\boldsymbol{\rho}_t\| \leq err = 10^{-5}$. In addition, we evaluate the probability of moderately precise recovery by setting $err = 10^{-3}$. As shown in Figure 2, our experimentation indicates that beginning with $3N$ interferometric Gaussian measurements, GWF achieves exact recovery with high probabilities. Furthermore, the method provides robust recovery with as low as $2.3N$ interferometric Gaussian measurements, with a relative error of 10^{-3} and below at over 95 percent.

6.2. Multistatic passive radar imaging. An interferometric inversion problem of great interest is multistatic passive radar imaging. We consider an imaging setup in which several static, terrestrial receivers are placed in a circle of radius around the scene of interest, which is illuminated by a transmitter of opportunity. An exemplary multistatic imaging geometry is illustrated in Figure 3. At receiver i, the backscattered signal is collected at a fixed location by a linear map $\mathbf{L}_i$ parameterized by the temporal frequency variable ω. The linear measurements collected at two receivers i and j are then pairwise correlated in time to yield the cross-correlation model (1.2) defined in the temporal frequency domain.

The key advantage of the interferometric model is the elimination of the dependence of measurements on the transmitter location and phase of the transmitted waveform, both of which are unknown in the passive scenario. In prior studies, the interferometric wave-based imaging was approached by low rank recovery methods [33, 44, 67]. We postulate that the GWF framework provides a computationally and memorywise efficient alternative to LRMR-based passive radar imaging.

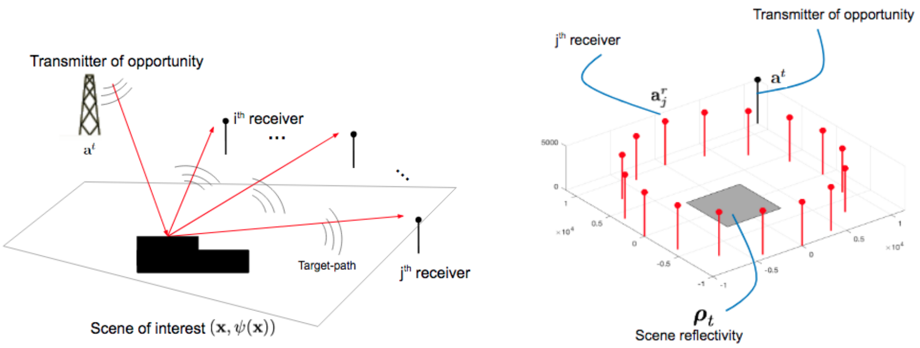


Figure 3. *An illustration of a multistatic imaging configuration. A scene is illuminated by a stationary illuminator of opportunity, located at* $\mathbf{a}^t$*. The backscattered signal is measured by a collection of stationary receivers, encircling the scene of interest at locations* $\mathbf{a}_j^r$*.*

6.2.1. Received signal model. Let $\mathbf{a}_i^r \in \mathbb{R}^3$ denote the spatial locations of the receivers, and assume S number of receivers such that $i = 1, 2, \ldots, S$. We assume that scattered signals are due to a single source of opportunity located at $\mathbf{a}^t$. The location on the surface of the earth is denoted by $\mathbf{x} = (\boldsymbol{x}, \psi(\boldsymbol{x})) \in \mathbb{R}^3$, where $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$ is a known ground topography, and $\rho : \mathbb{R}^2 \to \mathbb{R}$ denotes the ground reflectivity.

Under the flat topography assumption and Born approximation, and assuming waves propagate in free space, the fast-time temporal Fourier transform of the received signal at the ith receiver can be modeled as [64]

$$f_i(\omega) \approx \mathcal{L}_i[\rho](\omega, s) := p(\omega, s) \int_D \mathrm{e}^{-\mathrm{i}\omega \frac{\phi_i(\mathbf{x})}{c_0}} \alpha_i(\boldsymbol{x}, \mathbf{a}^t)\rho(\boldsymbol{x}) d\boldsymbol{x}, \tag{6.2}$$

where ω is the temporal frequency variable, c_0 is the speed of light in free space, $p(\omega, s)$ is the transmitted waveform, $\alpha_i(\boldsymbol{x}, \mathbf{a}^t)$ is the azimuth beam pattern, and

$$\phi_i(\boldsymbol{x}) = |\mathbf{x} - \mathbf{a}_i^r| + |\mathbf{x} - \mathbf{a}^t|, \tag{6.3}$$

is the bi-static delay term.

Following the derivations in [33] and discretizing the domain D of ρ into N samples, the cross-correlated data model for the multistatic scenario is obtained as

$$d_{ij}(\omega) = \sum_{k=1}^{N} e^{-\mathrm{i}\omega\left(|\mathbf{x}_k - \mathbf{a}_i^r| + \hat{\mathbf{a}}^t \cdot \mathbf{x}_k\right)/c_0} \boldsymbol{\rho}_k \sum_{k'=1}^{N} e^{\mathrm{i}\omega\left(|\mathbf{x}_{k'} - \mathbf{a}_j^r| + \hat{\mathbf{a}}^t \cdot \mathbf{x}_{k'}\right)} \overline{\boldsymbol{\rho}_{k'}}, \tag{6.4}$$

where $\boldsymbol{\rho}_k = \rho(\mathbf{x}_k)$ is the kth entry of the discretized scene reflectivity vector $\boldsymbol{\rho} \in \mathbb{C}^N$, and $\hat{\mathbf{a}}^t$ is the unit vector in the direction of the $\mathbf{a}^t$. We next discretize the temporal frequency domain Ω into M' samples and define the measurement vectors with entries

$$[\mathbf{L}_i^{m'}]_k = e^{\mathrm{i}\omega_{m'}(|\mathbf{x}_k - \mathbf{a}_i^r| - \hat{\mathbf{a}}^t \cdot \mathbf{x}_k)/c_0}, \quad k = 1, \ldots, N. \tag{6.5}$$

Rewriting (6.4) using (6.5), we obtain the interferometric measurement model as

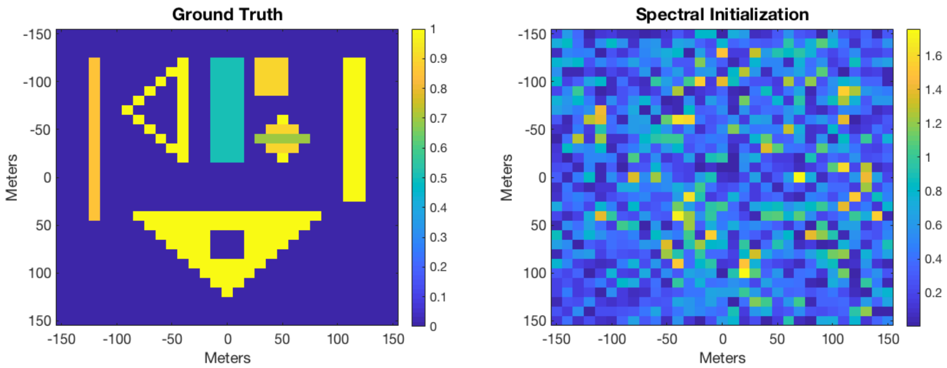


Figure 4. *The ground truth image and the initial estimate by the spectral method of GWF by* (3.21) *and* (3.22).

$$d^{ij}(\omega_{m'}) = \langle \mathbf{L}_i^{m'}, \ \boldsymbol{\rho}\rangle \overline{\langle \mathbf{L}_j^{m'}, \ \boldsymbol{\rho}\rangle}, \quad m' = 1, \dots, M', \ i = 1, \dots, S, \ j \neq i, \tag{6.6}$$

which corresponds to total of $M = M'\binom{S}{2}$ cross-correlated measurements[6] In [66], we show that the model in (6.4) satisfies the sufficient condition of Theorem 4.6, and hence GWF can provide exact image reconstruction for multistatic passive radar.

6.2.2. Simulation setup and results. We assume isotropic transmit and receive antennas, and we simulate a transmitted signal with 20MHz bandwidth and 1GHz center frequency. We assume isotropic transmit and receive antennas, and we simulate a transmitted signal with 20MHz bandwidth and 1GHz center frequency. We place the center of the scene at the origin of the coordinate system and generate a phantom, which consists of multiple point and extended targets as depicted in Figure 4. The transmitter is fixed and located at coordinates $\mathbf{a}^t = [11.5, 11.5, 0.5]$ km. We simulate a flat spectrum waveform and sample the temporal frequency domain into 32 samples.

(i) *GWF reconstruction:* To evaluate the performance of GWF for interferometric multistatic radar imaging, we simulate a $300 \times 300\text{m}^2$ scene and discretize it by 10m, which corresponds to 31×31 pixels, hence, $N = 961$. We use 16 receiver antennas that are placed in a circle of radius 10km around the scene at height of 0.5km, which corresponds to 120 unique cross-correlations at each temporal frequency sample. We generate the backscattered signals at each receiver by the linear measurement map of (6.2) and correlate linear measurements of each pairwise combination of receivers to generate interferometric data. In reconstruction, we deploy the approximate measurement vectors in which the transmitter distance is removed as defined in (6.4) and (6.5), hence only the transmitter look-direction is used at recovery by GWF.

In Figure 5, we demonstrate the reconstruction obtained by GWF after 10000 iterations, using the gradually increasing step size heuristic described in section 6.1. In addition, we plot the relative error of GWF iterates, which displays a geometric rate of convergence as stated

[6]Note that since we have multiple receivers to cross-correlate, the $m = 1, \dots, M$ summation of the objective function in (2.1) now consists of two sums, one over all unique cross-correlations $i \neq j$, and one over the temporal frequency index m'.

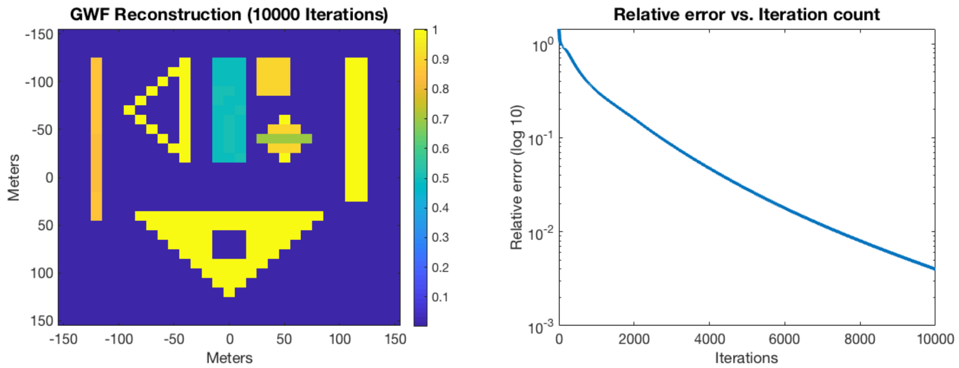


Figure 5. *Reconstructed image by GWF after* 10000 *iterations of* (3.1) *and the relative error with respect to the ground truth in log-scale versus iterations.*

in Theorem 4.6. The numerical results on our phantom image indicate GWF has the capacity to form accurate imagery of complex extended scenes.

(ii) *Comparison to LRMR by Uzawa's method:* We next perform smaller sized experiments to compare with Uzawa's method for LRMR by lifting. Namely, we compare GWF to three different formulations that yield Uzawa's iterations of the form in (2.12). These are the convex-relaxed trace regularization problem solved by [33], the rank-1 constrained nonconvex version with projections defined in (3.16), and the convex formulation obtained by dropping the rank constraint fully, i.e., a projected gradient descent analogue of [20].[7] Notably, we assess the reconstruction performance with respect to amount of computations, and we run the algorithms for identical number of flops. For our problem size, we run GWF for 8000 iterations, which roughly corresponds to 56 iterations of Uzawa's method. As complexity is known only up to an order, to avoid positive bias toward GWF we run the variations of the Uzawa's method for 110 iterations, i.e., double what the mismatch in computation from lifting the unknown indicates.

The simulated scene corresponds to 144×144m, discretized uniformly by 12m, to yield a 12×12 unknown, hence $N = 144$. We use 12 receiver antennas that are placed in a circle of radius 10km around the scene at a height of 0.5km, yielding 66 unique cross-correlations at each temporal frequency sample. The transmitted waveform bandwidth is set at 10MHz, around 1.9GHz central frequency, which corresponds to imaging below the range resolution limit of the system [66]. We provide mean squared error with respect to number of flops in computation over the lifted and the signal domains, and the final images reconstructed by each algorithm in Figures 6 and 7, respectively. Overall, our results indicate that GWF achieves exact reconstruction considerably faster than LRMR by Uzawa's method when compared at an identical number of flops. Furthermore, the convergent behavior of the convex LRMR methods under consideration supports our observation that their exact recovery guarantees

[7]Note that the original proposed method in [20] follows the Douglas–Rachford splitting framework to solve the nontrace regularized convex program in the lifted domain, which has different exact recovery arguments to Uzawa's method. To avoid the computational burden required to solve a linear system in the lifted domain, we use the projected gradient approach under the identical problem formulation.

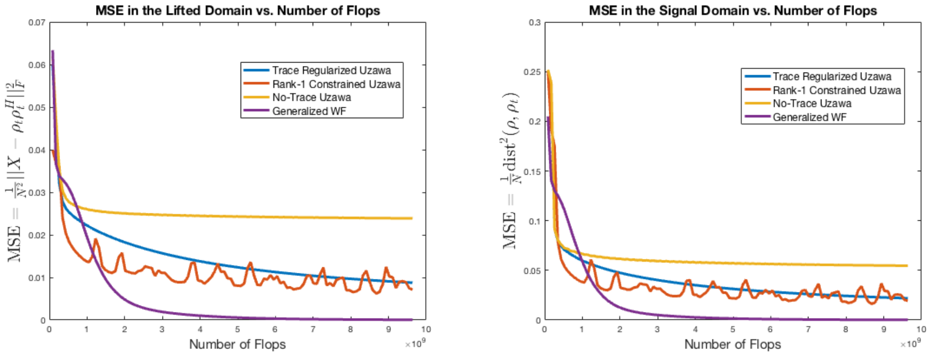
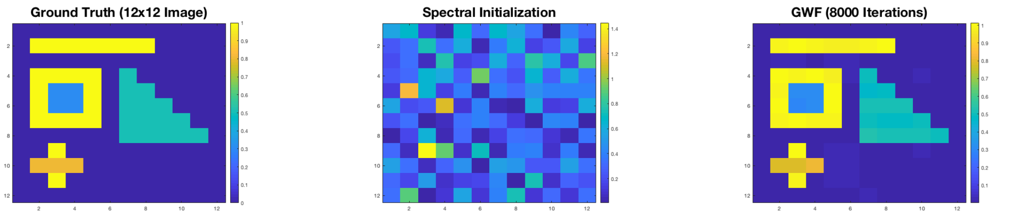
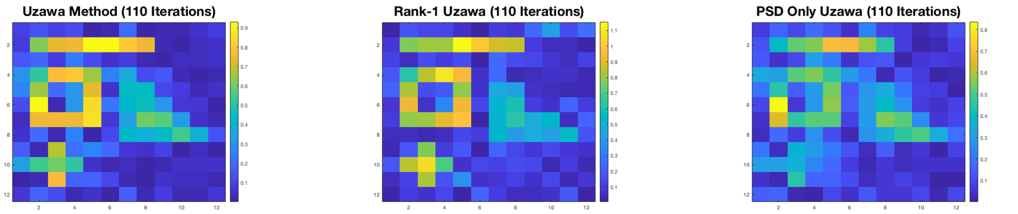


Figure 6. *Mean squared error curves in the lifted and signal domain versus number of flops, respectively.*



(a) Ground truth scene. (b) Spectral initialization. (c) GWF reconstruction.



(d) Uzawa's method with trace regularization. (e) Uzawa's method with rank-1 constraint. (f) Uzawa's method with only PSD constraint.

Figure 7. *Reconstruction results of GWF and variations of Uzawa's method at identical number of flops (*8000 *iterations for GWF,* 110 *iterations for Uzawa's methods).*

are more stringent than those of GWF. Thereby, our numerical results demonstrate its superior performance at an identical number of flops to lifting-based approaches and promote GWF as a favorable alternative to state-of-the-art interferometric wave-based imaging methods.

7. Conclusion. In this paper, we present a novel framework for exact interferometric inversion. We approach the interferometric inversion problem from the perspective of phase retrieval techniques. We examine two of the most prominent phase retrieval methods, namely LRMR-based PhaseLift, and nonconvex optimization based WF, and bridge the theory between the two frameworks. We then generalize WF and formulate the GWF framework for interferometric inversion and extend the exact recovery guarantees to arbitrary measurement maps with properties that are characterized in the equivalent lifted problem. Thereby, we establish exact recovery conditions for a larger class of problems than that of standard WF.

We identify the sufficient conditions for exact interferometric inversion on the lifted forward model as the RIP on rank-1, PSD matrices, with a RIC of $\delta \leq 0.214$. In developing our theory, we use the special structure of the rank-1, PSD set of matrices to show that the RIP directly implies the regularity condition of WF. Furthermore, we show that the concentration bound of the spectral matrix directly implies the RIP over rank-1, PSD matrices for cross-correlations of the complex Gaussian model. Hence, in generalizing the theory of WF for interferometric inversion in the complex Gaussian case, we demonstrate that the regularity condition becomes redundant. We illustrate that the empirical probability of exact interferometric inversion by GWF requires smaller oversampling factors than that of phase retrieval in the Gaussian model. Finally, we demonstrate the applicability of GWF in a deterministic, passive multistatic radar imaging problem using realistic imaging parameters. In conclusion, our paper shows that the computational and theoretical advantages promote GWF as a practical technique in real-world imaging applications.

Appendix A. Derivations.

A.1. Derivation of $\nabla\mathcal{J}$. Recall the objective function $\mathcal{J}$ in (2.1),

$$\mathcal{J}(\boldsymbol{\rho}) = \frac{1}{2M}\sum_{m=1}^{M} |\left(\mathbf{L}_i^m\right)^H \boldsymbol{\rho}\boldsymbol{\rho}^H \mathbf{L}_j^m - d_m^{ij}|^2. \tag{A.1}$$

Letting $e^m = \left(\mathbf{L}_i^m\right)^H \boldsymbol{\rho}\boldsymbol{\rho}^H \mathbf{L}_j^m - d_{ij}^m$, we write

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\rho}} = \frac{1}{2M}\sum_{m=1}^{M} \frac{\partial}{\partial \boldsymbol{\rho}}(e^m \bar{e}^m) = \frac{1}{2M}\sum_{m=1}^{M} \frac{\partial e^m}{\partial \boldsymbol{\rho}}\bar{e}^m + \frac{\partial \bar{e}^m}{\partial \boldsymbol{\rho}} e^m. \tag{A.2}$$

Having $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^H$ independent by properties of Wirtinger derivatives, we compute the partial derivatives in (A.2) as

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\rho}} = \frac{1}{2M}\sum_{m=1}^{M} \bar{e}^m \left(\boldsymbol{\rho}^H \mathbf{L}_j^m \left(\mathbf{L}_i^m\right)^H\right) + e^m \left(\boldsymbol{\rho}^H \mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H\right). \tag{A.3}$$

Using the definition of complex gradient provided in (3.2), we finally get

$$\nabla\mathcal{J} = \frac{1}{2M}\sum_{m=1}^{M} \bar{e}^m \mathbf{L}_j^m \left(\mathbf{L}_i^m\right)^H \boldsymbol{\rho} + e^m \mathbf{L}_i^m \left(\mathbf{L}_j^m\right)^H \boldsymbol{\rho}. \tag{A.4}$$

A.2. Proof of Lemma 4.5. Assuming the RIP on $\mathcal{F}$ over rank-1, PSD matrices with RIC-δ, we write

$$(1-\delta)\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H\right\|_F^2 \leq \left\|\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|^2 \leq (1+\delta)\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H\right\|_F^2. \tag{A.5}$$

Equivalently, from the definition of the Frobenius inner product and the adjoint operator $\mathcal{F}^H$, we reexpress (A.5) as

$$(1-\delta)\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H, \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F \leq \left\langle \mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F \leq (1+\delta)\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H, \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F, \tag{A.6}$$

$$\left|\left\langle \left(\mathcal{F}^H\mathcal{F} - \mathbf{I}\right)\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F\right| \leq \delta\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H\right\|_F^2; \tag{A.7}$$

hence for any $\boldsymbol{\rho} \in \mathbb{C}^N$ we have

$$\frac{\left|\left\langle \left(\mathcal{F}^H\mathcal{F} - \mathbf{I}\right)\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H \right\rangle_F\right|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2} \leq \delta. \tag{A.8}$$

Now consider the definition of the spectral norm with

$$\left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\| = \max_{\|\mathbf{v}\|=1}\left|\mathbf{v}^H\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\mathbf{v}\right| = \max_{\|\mathbf{v}\|=1}\left|\left\langle \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \mathbf{v}\mathbf{v}^H\right\rangle_F\right|. \tag{A.9}$$

From the Hermitian property of $\boldsymbol{\delta}$, we have

$$\left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\| = \max_{\|\mathbf{v}\|=1}\left|\left\langle \sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \sqrt{\boldsymbol{\delta}}^H\left(\mathbf{v}\mathbf{v}^H\right)\right\rangle_F\right| \leq \max_{\|\mathbf{v}\|=1}\left\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|_F\left\|\sqrt{\boldsymbol{\delta}}\left(\mathbf{v}\mathbf{v}^H\right)\right\|_F, \tag{A.10}$$

where $\sqrt{\boldsymbol{\delta}} : \mathbb{C}^{N\times N} \to \mathbb{C}^{N\times N}$ with $\sqrt{\boldsymbol{\delta}}\sqrt{\boldsymbol{\delta}} = \boldsymbol{\delta}$, and the inequality follows from the matrix Cauchy–Schwarz property and the fact that the Frobenius norm is unaltered by conjugation. Observe that using the Hermitian property of $\boldsymbol{\delta}$ via Definition 6.1 in [24], for any $\boldsymbol{\rho} \in \mathbb{C}^N$, we have $\|\sqrt{\boldsymbol{\delta}}(\boldsymbol{\rho}\boldsymbol{\rho}^H)\|_F^2/\|\boldsymbol{\rho}\boldsymbol{\rho}\|_F^2 \leq \delta$, hence

$$\left\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|_F \max_{\|\mathbf{v}\|=1}\left\|\sqrt{\boldsymbol{\delta}}\left(\mathbf{v}\mathbf{v}^H\right)\right\|_F \leq \sqrt{\delta}\left\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|_F \leq \delta\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H\right\|_F. \tag{A.11}$$

Using the definition of the spectral norm, we simply obtain

$$\frac{\left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|} \leq \delta \tag{A.12}$$

for any $\boldsymbol{\rho} \in \mathbb{C}^N$ if (A.6) holds. Furthermore since δ is the smallest constant such that (A.6) is satisfied, let

$$\delta := \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\frac{\left|\left\langle \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F\right|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2} = \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\frac{\left\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\|_F^2}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2}. \tag{A.13}$$

Revisiting (A.10), since from (A.12) we know for any $\boldsymbol{\rho}$, $\|\boldsymbol{\delta}(\boldsymbol{\rho}\boldsymbol{\rho}^H)\| \leq \delta\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F$, the maximal value of δ is reached at $\mathbf{v} = \boldsymbol{\rho}/\|\boldsymbol{\rho}\|$, as

$$\begin{aligned}\max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\|\boldsymbol{\delta}(\boldsymbol{\rho}\boldsymbol{\rho}^H)\| &\leq \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\max_{\|\mathbf{v}\|=1}\left\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\|_F\|\sqrt{\boldsymbol{\delta}}\left(\mathbf{v}\mathbf{v}^H\right)\right\|_F \\ &= \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\frac{\|\sqrt{\boldsymbol{\delta}}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\|_F^2}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F} = \delta\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F,\end{aligned} \tag{A.14}$$

and by definition

$$\begin{aligned}\max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\frac{\|\boldsymbol{\delta}(\boldsymbol{\rho}\boldsymbol{\rho}^H)\|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F} &\geq \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\max_{\|\mathbf{v}\|=1}\frac{\left|\left\langle \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \mathbf{v}\mathbf{v}^H\right\rangle_F\right|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F} \\ &\geq \max_{\boldsymbol{\rho}\in\mathbb{C}^{\mathbf{N}}\backslash\{\mathbf{0}\}}\frac{\left|\left\langle \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right), \boldsymbol{\rho}\boldsymbol{\rho}^H\right\rangle_F\right|}{\|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2} = \delta.\end{aligned} \tag{A.15}$$

Hence the proof is complete.

Appendix B. Lemmas for Theorem 4.6.

B.1. Proof of Lemma 5.1. Since $\hat{\boldsymbol{\rho}}_t = e^{\mathrm{i}\Phi(\mathbf{v}_0)}\boldsymbol{\rho}_t$, we have

$$\mathrm{dist}^2(\mathbf{v}_0, \boldsymbol{\rho}_t) = \|\mathbf{v}_0 - \hat{\boldsymbol{\rho}}_t\|^2 = \|\mathbf{v}_0\|^2 + \|\hat{\boldsymbol{\rho}}_t\|^2 - 2\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle. \tag{B.1}$$

Knowing that $\Phi(\mathbf{v}_0)$ achieves $\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle = |\langle\boldsymbol{\rho}_t, \mathbf{v}_0\rangle|$, we have

$$\mathrm{Re}\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle = \langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle = |\langle e^{\mathrm{i}\Phi(\mathbf{v}_0)}\boldsymbol{\rho}_t, \mathbf{v}_0\rangle|. \tag{B.2}$$

Since $\hat{\boldsymbol{\rho}}_t$ and $\mathbf{v}_0$ are unit norm, the geometric angle between them can be written as

$$\cos(\theta) = \langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle. \tag{B.3}$$

Invoking the representation theorem in Hilbert spaces, there exists a unit vector $\mathbf{v}_0^\perp$ that lies in the plane whose normal is $\mathbf{v}_0$ such that

$$\langle\hat{\boldsymbol{\rho}}_t, \mathbf{v}_0\rangle = \cos(\theta)\langle\mathbf{v}_0, \mathbf{v}_0\rangle + \sin(\theta)\langle\mathbf{v}_0^\perp, \mathbf{v}_0\rangle, \tag{B.4}$$

$$\langle\hat{\boldsymbol{\rho}}_t - (\cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^\perp), \mathbf{v}_0\rangle = 0. \tag{B.5}$$

The inner product is zero only when (1) $\hat{\boldsymbol{\rho}}_t = \cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^\perp$, 2) $\hat{\boldsymbol{\rho}}_t - (\cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^\perp)$ is perpendicular to $\mathbf{v}_0$. The latter case occurs iff $\hat{\boldsymbol{\rho}}_t - \cos(\theta)\mathbf{v}_0 = 0$, which is true only for $\theta = 0$, which indicates the identical solution as the former; hence we have the unique representation

$$\hat{\boldsymbol{\rho}}_t = \cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^\perp. \tag{B.6}$$

Using the same representation, it is straightforward to see that the unit length element $\hat{\boldsymbol{\rho}}_t^\perp = -\sin(\theta)\mathbf{v}_0 + \cos(\theta)\mathbf{v}_0^\perp$ satisfies

$$\langle\hat{\boldsymbol{\rho}}_t^\perp, \hat{\boldsymbol{\rho}}_t,\rangle = \left\langle -\sin(\theta)\mathbf{v}_0 + \cos(\theta)\mathbf{v}_0^\perp, \cos(\theta)\mathbf{v}_0 + \sin(\theta)\mathbf{v}_0^\perp \right\rangle = 0 \tag{B.7}$$

and hence lies in the plane whose normal is $\hat{\boldsymbol{\rho}}_t$.

B.2. Proof of Lemma 5.2. Recalling the representation of the spectral estimate in the lifted problem we have

$$\hat{\mathbf{X}} = \mathcal{P}_S\left(\mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right), \tag{B.8}$$

where $\mathcal{P}_S$ is the projection onto the set of symmetric matrices. $\hat{\mathbf{X}}_{PSD}$ is similarly written as

$$\hat{\mathbf{X}}_{PSD} = \mathcal{P}_{PSD}(\hat{\mathbf{X}}) = \mathcal{P}_{PSD}\left(\mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right) \tag{B.9}$$

since the PSD cone is a subset of the set of symmetric matrices S, and $\mathcal{P}_{PSD}$ projection consists of projection onto S by $\mathcal{P}_S$, followed by suppression of negative eigenvalues. From Lemma 4.5, denoting $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ for the solution $\boldsymbol{\rho}_t$ obeying $\|\boldsymbol{\rho}_t\| = 1$ we have

$$\|\hat{\mathbf{X}} - \tilde{\boldsymbol{\rho}}_t\| \leq \left\|\frac{1}{2}\mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}}_t) + \frac{1}{2}\left(\mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}}_t)\right)^H - \tilde{\boldsymbol{\rho}}_t\right\| \tag{B.10}$$

$$\leq \frac{1}{2}\left\|\mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}}_t) - \tilde{\boldsymbol{\rho}}_t\right\| + \frac{1}{2}\left\|\left(\mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}}_t)\right)^H - \tilde{\boldsymbol{\rho}}_t\right\| \tag{B.11}$$

$$\leq \frac{1}{2}\delta + \frac{1}{2}\left\|((\tilde{\boldsymbol{\rho}}_t) + \boldsymbol{\delta}(\tilde{\boldsymbol{\rho}}_t))^H - \tilde{\boldsymbol{\rho}}_t\right\|.$$

And since $\tilde{\boldsymbol{\rho}}_t$ is Hermitian symmetric and the spectral norm is unaffected by adjoint operation, we have

$$\frac{1}{2}\left\|((\tilde{\boldsymbol{\rho}}_t)+\boldsymbol{\delta}(\tilde{\boldsymbol{\rho}}_t))^H-\tilde{\boldsymbol{\rho}}_t\right\|=\frac{1}{2}\left\|(\boldsymbol{\delta}(\tilde{\boldsymbol{\rho}}_t))^H\right\|=\frac{1}{2}\|\boldsymbol{\delta}(\tilde{\boldsymbol{\rho}}_t)\|\leq\frac{1}{2}\delta. \tag{B.12}$$

Hence the spectral norm of the lifted error is upper bounded as follows:

$$\|\hat{\mathbf{X}}-\tilde{\boldsymbol{\rho}}_t\|\leq\delta. \tag{B.13}$$

For the maximal eigenvalue λ_0 of $\hat{\mathbf{X}}$, we can write

$$\lambda_0\geq\boldsymbol{\rho}_t^H\hat{\mathbf{X}}\boldsymbol{\rho}_t=\boldsymbol{\rho}_t^H(\hat{\mathbf{X}}-\tilde{\boldsymbol{\rho}}_t)\boldsymbol{\rho}_t+1\geq 1-\delta. \tag{B.14}$$

On the other hand, we have

$$\left|\mathbf{v}_0^H(\hat{\mathbf{X}}-\tilde{\boldsymbol{\rho}}_t)\mathbf{v}_0\right|\leq\|\hat{\mathbf{X}}-\tilde{\boldsymbol{\rho}}_t\|\leq\delta\ \rightarrow\ \left|\lambda_0-\left|\mathbf{v}_0^H\boldsymbol{\rho}_t\right|^2\right|\leq\delta, \tag{B.15}$$
$$\lambda_0\leq\delta+\left|\mathbf{v}_0^H\boldsymbol{\rho}_t|^2\right|\leq 1+\delta.$$

Hence we obtain

$$1-\delta\leq\lambda_0\leq 1+\delta. \tag{B.16}$$

Since the PSD estimate $\hat{\mathbf{X}}_{PSD}$ only differs from $\hat{\mathbf{X}}$ by the suppression of negative eigenvalues, and since spectral initialization only preserves the leading eigenvalue-eigenvector pair λ_0, $\mathbf{v}_0$ and $\lambda_0\geq 1-\delta$ where $\delta>0$, we have the identical $\boldsymbol{\rho}_0=\sqrt{\lambda_0}\mathbf{v}_0$.

B.3. Proof of Lemma 5.3. We consider the case where the spectral estimate (3.21) is projected onto the PSD cone as in LRMR. In this case, the estimate obtained in spectral initialization is projected onto the PSD cone to obtain $\hat{\mathbf{X}}_{PSD}$, which yields the identical initial point $\boldsymbol{\rho}_0$ per Lemma 5.2. Since its a symmetric, PSD matrix, we can decompose $\hat{\mathbf{X}}_{PSD}$ such that

$$\hat{\mathbf{X}}_{PSD}=\mathbf{S}_0\mathbf{S}_0,$$

where $\mathbf{S}_0$ is a PSD matrix with its eigenvalues as $\sqrt{\lambda_{\hat{\mathbf{X}}_{PSD}}}$. Since

$$\mathbf{v}_0:=\underset{\|\mathbf{v}\|=1}{\operatorname{argmax}}\ \mathbf{v}^H\hat{\mathbf{X}}_{PSD}\mathbf{v}, \tag{B.17}$$

where $\mathbf{v}^H\hat{\mathbf{X}}_{PSD}\mathbf{v}=\mathbf{v}^H\mathbf{S}_0\mathbf{S}_0\mathbf{v}=\|\mathbf{S}_0\mathbf{v}\|^2$. Then using the definitions of $\hat{\boldsymbol{\rho}}_t$ and $\hat{\boldsymbol{\rho}}_t^\perp$ from Lemma 5.1, we have

$$\mathbf{S}_0\hat{\boldsymbol{\rho}}_t=\cos(\theta)\mathbf{S}_0\mathbf{v}_0+\sin(\theta)\mathbf{S}_0\mathbf{v}_0^\perp, \tag{B.18}$$
$$\mathbf{S}_0\hat{\boldsymbol{\rho}}_t^\perp=-\sin(\theta)\mathbf{S}_0\mathbf{v}_0+\cos(\theta)\mathbf{S}_0,\mathbf{v}_0^\perp \tag{B.19}$$

and from the Pythogorean theorem, since we have orthogonal components, we get

$$\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t\|^2=\cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0\|^2+\sin^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^\perp\|^2, \tag{B.20}$$
$$\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t^\perp\|^2=\sin^2(\theta)\|\mathbf{S}_0\mathbf{v}_0\|^2+\cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^\perp\|^2. \tag{B.21}$$

Consider the expression $f = \sin^2(\theta)\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t\|^2 - \|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t^{\perp}\|^2$. Following the algebra in [54], we have

$$\begin{aligned} f &= \sin^2(\theta)\left(\cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0\|^2 + \sin^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2\right) - \left(\sin^2(\theta)\|\mathbf{S}_0\mathbf{v}_0\|^2 + \cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2\right) \\ &= \sin^2(\theta)\left(\cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0\|^2 - \|\mathbf{S}_0\mathbf{v}_0\|^2 + \sin^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2\right) - \cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2, \end{aligned}$$

which finally yields

$$f = \sin^4(\theta)\left(\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2 - \|\mathbf{S}_0\mathbf{v}_0\|^2\right) - \cos^2(\theta)\|\mathbf{S}_0\mathbf{v}_0^{\perp}\|^2. \tag{B.22}$$

Since $\mathbf{v}_0$ is the unit vector that maximizes $\|\mathbf{S}_0\mathbf{v}\|^2$, with the fact that $\mathbf{v}_0^{\perp}$ is also a unit vector, this expression is always nonpositive, hence

$$\sin^2(\theta)\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t\|^2 - \|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t^{\perp}\|^2 \leq 0,$$

$$\sin^2(\theta) \leq \frac{\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t^{\perp}\|^2}{\|\mathbf{S}_0\hat{\boldsymbol{\rho}}_t\|^2}. \tag{B.23}$$

Equivalently, expressing the upper bound with the spectral estimate, we obtain

$$\sin^2(\theta) \leq \frac{(\hat{\boldsymbol{\rho}}_t^{\perp})^H\hat{\mathbf{X}}_{PSD}\hat{\boldsymbol{\rho}}_t^{\perp}}{\hat{\boldsymbol{\rho}}_t^H\hat{\mathbf{X}}_{PSD}\hat{\boldsymbol{\rho}}_t}. \tag{B.24}$$

Now we consider the spectral estimate $\hat{\mathbf{X}}_{PSD}$. We know that $\hat{\mathbf{X}}_{PSD}$ is obtained by projecting the intermediate estimate $\mathbf{Y} = \mathcal{F}^H\mathcal{F}(\tilde{\boldsymbol{\rho}}_t)$ onto the feasible set of PSD matrices as defined in Lemma 5.2. From Lemma 4.5, we know that $\mathcal{F}^H\mathcal{F}$ is approximately an identity on the domain of rank-1, PSD matrices, hence we can write the perturbation model

$$\hat{\mathbf{X}}_{PSD} = \mathcal{P}_{PSD}\left(\tilde{\boldsymbol{\rho}}_t + \boldsymbol{\delta}[\tilde{\boldsymbol{\rho}}_t]\right). \tag{B.25}$$

Since projection onto the PSD cone is nonexpansive under the spectral norm, we have the following:

$$\|\hat{\mathbf{X}}_{PSD} - \tilde{\boldsymbol{\rho}}_t\| \leq \|\mathbf{Y} - \tilde{\boldsymbol{\rho}}_t\|. \tag{B.26}$$

Setting $\hat{\mathbf{X}}_{PSD} = \tilde{\boldsymbol{\rho}}_t + \tilde{\mathbf{e}}$, the upper bound in (B.24) can be written as

$$\sin^2(\theta) \leq \frac{(\hat{\boldsymbol{\rho}}_t^{\perp})^H\tilde{\boldsymbol{\rho}}_t\hat{\boldsymbol{\rho}}_t^{\perp} + (\hat{\boldsymbol{\rho}}_t^{\perp})^H\tilde{\mathbf{e}}\hat{\boldsymbol{\rho}}_t^{\perp}}{\hat{\boldsymbol{\rho}}_t^H\tilde{\boldsymbol{\rho}}_t\hat{\boldsymbol{\rho}}_t + \hat{\boldsymbol{\rho}}_t^H\tilde{\mathbf{e}}\hat{\boldsymbol{\rho}}_t}. \tag{B.27}$$

Since we have that $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$, the bound reduces to

$$\sin^2(\theta) \leq \frac{(\hat{\boldsymbol{\rho}}_t^{\perp})^H\tilde{\mathbf{e}}\hat{\boldsymbol{\rho}}_t^{\perp}}{1 + \hat{\boldsymbol{\rho}}_t^H\tilde{\mathbf{e}}\hat{\boldsymbol{\rho}}_t} \tag{B.28}$$

as $\hat{\boldsymbol{\rho}}_t^{\perp}$ is orthogonal to $\hat{\boldsymbol{\rho}}_t$ and the global phase component in $\hat{\boldsymbol{\rho}}_t = e^{j\Phi(v_0)}\boldsymbol{\rho}_t$ vanishes in the quadratic form. Moreover, we know that $(\hat{\boldsymbol{\rho}}_t^{\perp})^H\tilde{\mathbf{e}}\hat{\boldsymbol{\rho}}_t^{\perp}$ is nonnegative from positive semidefinitivity of $\hat{\mathbf{X}}_{PSD}$. Hence we further upper bound the numerator by the spectral norm of $\tilde{\mathbf{e}}$ and follow with the series of bounds

$$\sin^2(\theta) \leq \frac{\|\tilde{\mathbf{e}}\|}{1+\hat{\boldsymbol{\rho}}_t^H \tilde{\mathbf{e}} \hat{\boldsymbol{\rho}}_t} \leq \frac{\|\boldsymbol{\delta}(\tilde{\boldsymbol{\rho}}_t)\|}{1+\hat{\boldsymbol{\rho}}_t^H \tilde{\mathbf{e}} \hat{\boldsymbol{\rho}}_t} \leq \frac{\delta \|\tilde{\boldsymbol{\rho}}_t\|_F}{1+\hat{\boldsymbol{\rho}}_t^H \tilde{\mathbf{e}} \hat{\boldsymbol{\rho}}_t}, \tag{B.29}$$

where the last two inequalities follow from the nonexpansiveness property of the projection operator onto the PSD cone, and the definition of $\boldsymbol{\delta}$ from Lemma 4.5, such that the spectral norm of $\boldsymbol{\delta}$ is upper bounded by the RIC-δ of $\mathcal{F}$. For the term in the denominator, since we established that $\|\tilde{\mathbf{e}}\|_2 \leq \delta$, we have that $-\delta \leq \hat{\boldsymbol{\rho}}_t^H \tilde{\mathbf{e}} \hat{\boldsymbol{\rho}}_t \leq \delta$ to finally obtain

$$\sin^2(\theta) \leq \frac{\delta}{1+\hat{\boldsymbol{\rho}}_t^H \tilde{\mathbf{e}} \hat{\boldsymbol{\rho}}_t} \leq \frac{\delta}{1-\delta}. \tag{B.30}$$

B.4. Proof of Lemma 5.4.

B.4.1. The upper bound. Noting that $\hat{\boldsymbol{\rho}}_t$ is the closest solution in P to an estimate $\boldsymbol{\rho}$, we define $\boldsymbol{\rho}_e = \boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t$. Since $\boldsymbol{\rho} \in E(\epsilon)$, $\mathrm{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t) = \|\boldsymbol{\rho}_e\| \leq \epsilon$, and from reverse triangle inequality we have

$$|\|\boldsymbol{\rho}\| - \|\hat{\boldsymbol{\rho}}_t\|| \leq \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\| \leq \epsilon.$$

Since $\|\hat{\boldsymbol{\rho}}_t\| = \|\boldsymbol{\rho}_t\| = 1$, we have that $1-\epsilon \leq \|\boldsymbol{\rho}\| \leq 1+\epsilon$. Having $\hat{\boldsymbol{\rho}}_t \hat{\boldsymbol{\rho}}_t^H = \boldsymbol{\rho}_t \boldsymbol{\rho}_t^H$, we let $\tilde{\mathbf{e}} = \boldsymbol{\rho}\boldsymbol{\rho}^H - \hat{\boldsymbol{\rho}}_t \hat{\boldsymbol{\rho}}_t^H$ denote the error in the lifted problem. Then, writing the lifted error $\tilde{\boldsymbol{\rho}}_e = \boldsymbol{\rho}_e \boldsymbol{\rho}_e^H$ as

$$\begin{aligned} \boldsymbol{\rho}_e \boldsymbol{\rho}_e^H &= (\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t)(\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t)^H = \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}\hat{\boldsymbol{\rho}}_t^H - \hat{\boldsymbol{\rho}}_t \boldsymbol{\rho}^H + \hat{\boldsymbol{\rho}}_t \hat{\boldsymbol{\rho}}_t^H + \left(2\hat{\boldsymbol{\rho}}_t \hat{\boldsymbol{\rho}}_t^H - 2\hat{\boldsymbol{\rho}}_t \hat{\boldsymbol{\rho}}_t^H\right) \\ &= \tilde{\mathbf{e}} - (\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t)\hat{\boldsymbol{\rho}}_t^H - \hat{\boldsymbol{\rho}}_t \left(\boldsymbol{\rho}^H - \hat{\boldsymbol{\rho}}_t^H\right) \end{aligned} \tag{B.31}$$

finally yields the expression

$$\tilde{\mathbf{e}} = \boldsymbol{\rho}_e \boldsymbol{\rho}_e^H + \boldsymbol{\rho}_e \hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t \boldsymbol{\rho}_e^H \tag{B.32}$$

for the error in the lifted domain. We can then write the upper bound for $\|\tilde{\mathbf{e}}\|_F$ as

$$\|\tilde{\mathbf{e}}\|_F \leq \|\boldsymbol{\rho}_e \boldsymbol{\rho}_e^H\|_F + \|\boldsymbol{\rho}_e \hat{\boldsymbol{\rho}}_t^H\|_F + \|\hat{\boldsymbol{\rho}}_t \boldsymbol{\rho}_e^H\|_F. \tag{B.33}$$

Since all the arguments of the Frobenius norm in the right-hand side have rank-1, we have $\|\cdot\|_2 = \|\cdot\|_F$. Knowing that $\|\boldsymbol{\rho}_t\| = 1$ we get

$$\|\tilde{\mathbf{e}}\|_F \leq 2\|\hat{\boldsymbol{\rho}}_t\|\|\boldsymbol{\rho}_e\| + \|\boldsymbol{\rho}_e\|^2 \leq (2+\epsilon)\|\boldsymbol{\rho}_e\|. \tag{B.34}$$

Hence we obtain the upper bound

$$\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\|_F \leq (2+\epsilon)\mathrm{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t). \tag{B.35}$$

B.4.2. The lower bound. Expanding the error in the lifted domain, we get

$$\|\tilde{\mathbf{e}}\|_F^2 = \|\boldsymbol{\rho}\boldsymbol{\rho}^H\|_F^2 + \|\boldsymbol{\rho}_t \boldsymbol{\rho}_t^H\|_F^2 - 2\mathrm{Re}\langle \boldsymbol{\rho}\boldsymbol{\rho}^H, \boldsymbol{\rho}_t \boldsymbol{\rho}_t^H \rangle_F. \tag{B.36}$$

Since we have the rank-1 lifted signals, the Frobenius norms and the inner product reduce to

$$\begin{aligned} \|\tilde{\mathbf{e}}\|_F^2 &= \|\boldsymbol{\rho}\|^4 + \|\boldsymbol{\rho}_t\|^4 - 2|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|^2 = (\|\boldsymbol{\rho}\|^4 - |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|^2) + (\|\boldsymbol{\rho}_t\|^4 - |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|^2) \\ &= (\|\boldsymbol{\rho}\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|)(\|\boldsymbol{\rho}\|^2 - |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|) + (\|\boldsymbol{\rho}_t\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|)(\|\boldsymbol{\rho}_t\|^2 - |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t \rangle|). \end{aligned} \tag{B.37}$$

Having $\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) = \|\boldsymbol{\rho}\|^2 + \|\boldsymbol{\rho}_t\|^2 - 2|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle| = \|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\| \geq 0$, we can lower bound (B.36) using (B.37) as

$$\|\tilde{\mathbf{e}}\|_F^2 \geq \min\left((\|\boldsymbol{\rho}\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|), (\|\boldsymbol{\rho}_t\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|)\right)\left(\|\boldsymbol{\rho}\|^2 + \|\boldsymbol{\rho}_t\|^2 - 2|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|\right). \tag{B.38}$$

Knowing that $\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t) \leq \epsilon^2$ and the result from the reverse triangle inequality on $\|\boldsymbol{\rho}\|$, the terms within the minimization are further lower bounded using

$$\begin{aligned} 2|\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle| &\geq \|\boldsymbol{\rho}\|^2 + \|\boldsymbol{\rho}_t\|^2 - \epsilon^2, \\ |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle| &\geq (1-\epsilon). \end{aligned} \tag{B.39}$$

We then get the bound on the scalar multiplying $\text{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t)$ as

$$\min\left((\|\boldsymbol{\rho}\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|), (\|\boldsymbol{\rho}_t\|^2 + |\langle \boldsymbol{\rho}, \boldsymbol{\rho}_t\rangle|)\right) \geq (1-\epsilon)^2 + (1-\epsilon), \tag{B.40}$$

which yields the final lower bound as

$$\left\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F \geq \sqrt{(1-\epsilon)(2-\epsilon)}\ \text{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t). \tag{B.41}$$

B.5. Proof of Lemma 5.5. From the adjoint property of the inner product we have

$$\left\|\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\|^2 = \left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F.$$

Splitting the linear operator $\mathcal{F}^H\mathcal{F}$ over the rank-1 inputs, we can use the perturbation model from Lemma 4.5 such that

$$\begin{aligned} &\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right)\right\rangle_F - \left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \mathcal{F}^H\mathcal{F}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F \\ &= \left\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F^2 + \left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F. \end{aligned} \tag{B.42}$$

Using the representation of $\tilde{\mathbf{e}} = \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$ in (B.32) and the linearity of $\boldsymbol{\delta}$ we get

$$\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right) = \boldsymbol{\delta}\left(\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\right) + \boldsymbol{\delta}\left(\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H\right). \tag{B.43}$$

Notably, the domain of $\boldsymbol{\delta}$ is by definition the set of rank-1, PSD matrices. Having $\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H$, a symmetric matrix of at most rank-2 in the argument of $\boldsymbol{\delta}$ on the right-hand side, we can represent (B.43) by an eigenvalue decomposition and use the linearity of $\boldsymbol{\delta}$ to obtain

$$\boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right) = \boldsymbol{\delta}\left(\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\right) + \lambda_1\boldsymbol{\delta}\left(\mathbf{v}_1\mathbf{v}_1^H\right) + \lambda_2\boldsymbol{\delta}\left(\mathbf{v}_2\mathbf{v}_2^H\right), \tag{B.44}$$

where $\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H = \sum_{i=1}^2 \lambda_i\mathbf{v}_i\mathbf{v}_i^H$, with $\|\mathbf{v}_i\| = 1$. Plugging (B.44) into (B.42) and applying the triangle inequality, we have

$$\begin{aligned} \left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F\right| \leq &\left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}\left(\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\right)\right\rangle_F\right| \\ &+ |\lambda_1|\left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}\left(\mathbf{v}_1\mathbf{v}_1^H\right)\right\rangle_F\right| + |\lambda_2|\left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}(\mathbf{v}_2\mathbf{v}_2^H)\right\rangle_F\right|. \end{aligned} \tag{B.45}$$

Furthermore, knowing that $\tilde{\mathbf{e}}$ is symmetric and at most rank-2, let $\tilde{\mathbf{e}} = \sum_{i=1}^{2} \sigma_i \mathbf{u}_i \mathbf{u}_i^H$, where $\|\mathbf{u}_i\| = 1$. Then, using the triangle inequality on the right-hand-side terms in (B.45), and the outcome of Lemma 4.5 to RIP over rank-1, PSD matrices with RIC-δ_1, we have

$$\left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F\right| \leq \left(\sum_{i=1}^{2} |\sigma_i|\right) \delta_1 \left(\|\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\|_F + |\lambda_1| + |\lambda_2|\right). \tag{B.46}$$

Furthermore, since $\tilde{\mathbf{e}}$ is rank-2 by definition, and $\sum_{i=1}^{2} |\sigma_i| = \|\tilde{\mathbf{e}}\|_* \leq \sqrt{2}\|\tilde{\mathbf{e}}\|_F$, we obtain

$$\begin{aligned} \left|\left\langle \tilde{\mathbf{e}}, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F\right| &\leq \delta_1\sqrt{2}\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F \left(\|\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\|_F + |\lambda_1| + |\lambda_2|\right) \\ &\leq \delta_1\sqrt{2}\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F \left(\|\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\|_F + \|\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H\|_* + \|\hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H\|_*\right), \end{aligned} \tag{B.47}$$

where again we use the fact that $\|\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H\|_* = |\lambda_1| + |\lambda_2|$ by definition, from which the triangle inequality follows. Invoking the rank-1 property on the terms inside the parentheses in (B.47), the Frobenius norm and the nuclear norms can be computed by the spectral norm, since $\|\cdot\| = \|\cdot\|_F = \|\cdot\|_*$ for a rank-1 argument. Then, having $\|\boldsymbol{\rho}_t\| = 1$, for the right-hand side we obtain

$$\begin{aligned} \left|\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F\right| &\leq \delta_1\sqrt{2}\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F \left(2\|\boldsymbol{\rho}_t\|\|\boldsymbol{\rho}_e\| + \|\boldsymbol{\rho}_e\|^2\right) \\ &\leq \delta_1\sqrt{2}(2+\epsilon)\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F \operatorname{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t). \end{aligned} \tag{B.48}$$

Using the bound

$$\sqrt{(1-\epsilon)(2-\epsilon)}\operatorname{dist}(\boldsymbol{\rho}, \boldsymbol{\rho}_t) \leq \left\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right\|_F \tag{B.49}$$

from Lemma 5.4, we finally obtain the upper bound on the perturbation on $\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F^2$ as

$$\left|\left\langle \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H, \boldsymbol{\delta}\left(\boldsymbol{\rho}\boldsymbol{\rho}^H\right) - \boldsymbol{\delta}\left(\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\right)\right\rangle_F\right| \leq \delta_1 \frac{(2+\epsilon)\sqrt{2}}{\sqrt{(1-\epsilon)(2-\epsilon)}}\|\boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\|_F^2. \tag{B.50}$$

Hence, setting $\delta_2 = \delta_1 \frac{(2+\epsilon)\sqrt{2}}{\sqrt{(1-\epsilon)(2-\epsilon)}}$, we have the local RIP-2 condition satisfied with RIC-δ_2.

B.6. Proof of Lemma 5.6. Having $\tilde{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H = \hat{\boldsymbol{\rho}}_t\hat{\boldsymbol{\rho}}_t^H$, we rewrite the gradient term in (5.12) as

$$\nabla\mathcal{J}(\boldsymbol{\rho}) = \|\boldsymbol{\rho}\|^2\boldsymbol{\rho} - \left(\hat{\boldsymbol{\rho}}_t^H\boldsymbol{\rho}\right)\hat{\boldsymbol{\rho}}_t + \mathcal{P}_S(\boldsymbol{\delta}(\tilde{\mathbf{e}}))\boldsymbol{\rho},$$

where $\tilde{\mathbf{e}} = \boldsymbol{\rho}\boldsymbol{\rho}^H - \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$. Then, simply from the triangle inequality and the definition of projection operator on the set of symmetric matrices, we get

$$\begin{aligned} \|\nabla\mathcal{J}(\boldsymbol{\rho})\| &\leq \left\|\|\boldsymbol{\rho}\|^2\boldsymbol{\rho} - \left(\hat{\boldsymbol{\rho}}_t{}^H\boldsymbol{\rho}\right)\hat{\boldsymbol{\rho}}_t\right\| + \frac{1}{2}\|\boldsymbol{\delta}(\tilde{\mathbf{e}})\|_2\|\boldsymbol{\rho}\| + \frac{1}{2}\|\boldsymbol{\delta}(\tilde{\mathbf{e}})^H\|_2\|\boldsymbol{\rho}\| \\ \|\nabla\mathcal{J}(\boldsymbol{\rho})\| &\leq \left\|\|\boldsymbol{\rho}\|^2\boldsymbol{\rho} - \left(\hat{\boldsymbol{\rho}}_t{}^H\boldsymbol{\rho}\right)\hat{\boldsymbol{\rho}}_t\right\| + \|\boldsymbol{\delta}(\tilde{\mathbf{e}})\|_2\|\boldsymbol{\rho}\|, \end{aligned} \tag{B.51}$$

since the spectral norm is unchanged by the Hermitian transpose operation. For the spectral norm of $\boldsymbol{\delta}(\tilde{\mathbf{e}})$, we use the representation in (B.44) and apply the triangle inequality such that

$$\|\boldsymbol{\delta}(\tilde{\mathbf{e}})\| \leq \left\|\boldsymbol{\delta}\left(\boldsymbol{\rho}_e\boldsymbol{\rho}_e^H\right)\right\| + |\lambda_1|\left\|\boldsymbol{\delta}\left(\mathbf{v}_1\mathbf{v}_1^H\right)\right\| + |\lambda_2|\left\|\boldsymbol{\delta}\left(\mathbf{v}_2\mathbf{v}_2^H\right)\right\|, \tag{B.52}$$

where again $\boldsymbol{\rho}_e\hat{\boldsymbol{\rho}}_t^H + \hat{\boldsymbol{\rho}}_t\boldsymbol{\rho}_e^H = \sum_{i=1}^2 \lambda_i \mathbf{v}_i\mathbf{v}_i^H$, with $\|\mathbf{v}_i\| = 1$. We obtain the identical form in (B.48) such that

$$\|\boldsymbol{\delta}(\tilde{\mathbf{e}})\| \leq \delta_1\left(\|\boldsymbol{\rho}_e\|^2 + 2\|\boldsymbol{\rho}_t\|\|\boldsymbol{\rho}_e\|\right) \leq \delta_1(2+\epsilon)\|\boldsymbol{\rho}_e\|, \tag{B.53}$$

which overall yields

$$\|\nabla\mathcal{J}(\boldsymbol{\rho})\| \leq \left\|\|\boldsymbol{\rho}\|^2\boldsymbol{\rho} - \left(\hat{\boldsymbol{\rho}}_t{}^H\boldsymbol{\rho}\right)\hat{\boldsymbol{\rho}}_t\right\| + \delta_1(2+\epsilon)\|\boldsymbol{\rho}_e\|\|\boldsymbol{\rho}\| \tag{B.54}$$

$$\leq \|\boldsymbol{\rho}\|\left((\|\boldsymbol{\rho}\| + \|\boldsymbol{\rho}_t\|)\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\| + \delta_1(2+\epsilon)\|\boldsymbol{\rho}_e\|\right). \tag{B.55}$$

Using the fact that $\boldsymbol{\rho} \in E(\epsilon)$, and $\hat{\boldsymbol{\rho}}_t = e^{\mathrm{i}\phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t$, we know that $1-\epsilon \leq \|\boldsymbol{\rho}\| \leq 1+\epsilon$ from triangle and reverse triangle inequalities. Hence,

$$\|\nabla\mathcal{J}(\boldsymbol{\rho})\| \leq (1+\epsilon)\left((2+\epsilon)\|\boldsymbol{\rho}_e\| + \delta_1(2+\epsilon)\|\boldsymbol{\rho}_e\|\right) \tag{B.56}$$

$$\leq (1+\epsilon)(1+\delta_1)(2+\epsilon)\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\|. \tag{B.57}$$

Thereby, setting $c = (1+\epsilon)(1+\delta_1)(2+\epsilon)$ the right-hand side of the regularity condition is upper bounded as

$$\frac{1}{\alpha}\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\|^2 + \frac{1}{\beta}\|\nabla\mathcal{J}(\boldsymbol{\rho})\|^2 \leq \left(\frac{1}{\alpha} + \frac{c^2}{\beta}\right)\|\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}_t\|^2 \tag{B.58}$$

and the regularity condition (5.14) is established if the following condition holds:

$$\mathrm{Re}\left(\left\langle\nabla\mathcal{J}(\boldsymbol{\rho}), \left(\boldsymbol{\rho} - e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\boldsymbol{\rho}_t\right)\right\rangle\right) \geq \left(\frac{1}{\alpha} + \frac{c^2}{\beta}\right)\,\mathrm{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t). \tag{B.59}$$

B.7. Proof of Lemma 5.7. Indeed, the form of the regularity condition is nothing but the *restricted strong convexity condition.* Since $\nabla\mathcal{J}(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}) = 0$ for any $\Phi(\boldsymbol{\rho}) = \Phi_0 \in [0, 2\pi)$, by definition, one can equivalently write (5.18) as

$$\mathrm{Re}\left(\left\langle\left(\nabla\mathcal{J}(\boldsymbol{\rho}) - \nabla\mathcal{J}\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right\rangle\right) \geq \eta\|\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\|^2, \tag{B.60}$$

where $\eta = (\frac{1}{\alpha} + \frac{c^2}{\beta})$. Reorganizing the terms, we have

$$\mathrm{Re}\left(\left\langle\left(\nabla\mathcal{J}(\boldsymbol{\rho}) - \nabla\mathcal{J}\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right) - \eta\left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right\rangle\right) \geq 0. \tag{B.61}$$

Letting $g(\boldsymbol{\rho}) = \mathcal{J}(\boldsymbol{\rho}) - \frac{\eta}{2}\|\boldsymbol{\rho}\|^2$, we can write (B.61) as

$$\mathrm{Re}\left(\left\langle\left(\nabla g(\boldsymbol{\rho}) - \nabla g\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right), \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right\rangle\right) \geq 0. \tag{B.62}$$

For any $\boldsymbol{\rho}$ in the ϵ-ball of $\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}$, (B.62) is merely the local convexity condition for g at point $\hat{\boldsymbol{\rho}}_t = \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}$. Since the ϵ-ball around $\hat{\boldsymbol{\rho}}_t$ is a convex set, we can use the equivalent condition

$$g(\boldsymbol{\rho}) \geq g\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right) + \mathrm{Re}\left(\nabla g\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)^H\left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right). \tag{B.63}$$

Plugging in the definition for g, we obtain

$$\mathcal{J}(\boldsymbol{\rho}) \geq \mathcal{J}\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right) + \mathrm{Re}\left(\nabla\mathcal{J}\left(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)^H \left(\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\right)\right) + \frac{\eta}{2}\|\boldsymbol{\rho} - \boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}\|^2. \tag{B.64}$$

Since $\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}$ is a global minimizer of $\mathcal{J}$, it satisfies the first order optimality condition with $\nabla\mathcal{J}(\boldsymbol{\rho}_t e^{\mathrm{i}\Phi(\boldsymbol{\rho})}) = 0$, and the minimum it attains is 0. Hence the condition reduces to

$$\mathcal{J}(\boldsymbol{\rho}) \geq \frac{\eta}{2}\mathrm{dist}^2(\boldsymbol{\rho}, \boldsymbol{\rho}_t). \tag{B.65}$$

Appendix C. Lemmas for Theorem 4.9.

C.1. Proof of Lemma 5.8. For the intermediate stage, $\mathbf{Y}$, of the spectral estimate, we write

$$\mathbf{Y} = \frac{1}{M}\mathcal{F}^H\mathcal{F}(\boldsymbol{\rho}_t\boldsymbol{\rho}_t) = \frac{1}{M}\sum_{m=1}^{M}\left((\mathbf{L}_i^m)^H\boldsymbol{\rho}_t\boldsymbol{\rho}_t^H\mathbf{L}_j^m\right)\mathbf{L}_i^m(\mathbf{L}_j^m)^H. \tag{C.1}$$

Reorganizing the terms in (C.1) and taking the expectation, we have

$$\mathbb{E}[\mathbf{Y}] = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[(\mathbf{L}_i^m(\mathbf{L}_i^m)^H\boldsymbol{\rho}_t)(\mathbf{L}_j^m(\mathbf{L}_j^m)^H\boldsymbol{\rho}_t)^H]. \tag{C.2}$$

For fixed m, and having $\boldsymbol{\rho}_t$ is independent of sampling vectors, the $N \times N$ matrix inside the summation has the entries of the form

$$\sum_{n=1}^{N}\sum_{n'=1}^{N}\mathbb{E}[(\mathbf{L}_i^m)_k(\overline{\mathbf{L}_i^m})_n(\overline{\mathbf{L}_j^m})_l(\mathbf{L}_j^m)_{n'}]\boldsymbol{\rho}_{tn}\overline{\boldsymbol{\rho}}_{tn'}, \tag{C.3}$$

where k, l denote the row and column indexes, respectfully. Since $\mathbf{L}_i^m$ and $\mathbf{L}_j^m$ are independent of each other and have i.i.d. entries, the fourth moments of Gaussian entries are removed as $\mathbb{E}[(\mathbf{L}_i^m)_k(\overline{\mathbf{L}_i^m})_n(\overline{\mathbf{L}_j^m})_l(\mathbf{L}_j^m)_{n'}] = \mathbb{E}[(\mathbf{L}_i^m)_k(\overline{\mathbf{L}_i^m})_n]\mathbb{E}[(\overline{\mathbf{L}_j^m})_l(\mathbf{L}_j^m)_{n'}]$, in which the expectations are only nonzero for $n = k$, $n' = l$, yielding

$$= \mathbb{E}\left[|(\mathbf{L}_i^m)_k|^2\right]\mathbb{E}\left[|(\mathbf{L}_j^m)_l|^2\right]\boldsymbol{\rho}_{tk}\overline{\boldsymbol{\rho}_{tl}}, \tag{C.4}$$

where $\mathbf{L}_i^m, \mathbf{L}_j^m \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I}) + \mathrm{i}\mathcal{N}(0, \frac{1}{2}\mathbf{I})$ have unit variance. Hence,

$$\mathbb{E}[\mathbf{Y}]_{k,l} = \boldsymbol{\rho}_{tk}\overline{\boldsymbol{\rho}_{tl}}, \tag{C.5}$$

which is precisely equal to $\mathbb{E}[\mathbf{Y}] = \boldsymbol{\rho}_t\boldsymbol{\rho}_t^H$.

C.2. Proof of Lemma 5.9. We use the machinery in the proof of the concentration bound of the Hessian in [12].[8] By unitary invariance, we take $\boldsymbol{\rho}_t = \mathbf{e}_1$, where $\mathbf{e}_1$ is the first standard basis vector. We want to establish that

[8]Corresponds to Lemma 7.4 in the source material.

$$\left\| \frac{1}{M} \sum_{m=1}^{M} \overline{(\mathbf{L}_i^m)_1} (\mathbf{L}_j^m)_1 \mathbf{L}_i^m (\mathbf{L}_j^m)^H - \mathbf{e}_1 \mathbf{e}_1^T \right\| \leq \delta. \tag{C.6}$$

The inequality in (C.6) is equivalent to

$$I_0(\mathbf{y}) := \left| \mathbf{y}^H \left(\frac{1}{M} \sum_{m=1}^{M} \overline{(\mathbf{L}_i^m)_1} (\mathbf{L}_j^m)_1 \mathbf{L}_i^m (\mathbf{L}_j^m)^H - \mathbf{e}_1 \mathbf{e}_1^T \right) \mathbf{y} \right| \tag{C.7}$$

$$= \left| \frac{1}{M} \sum_{m=1}^{M} \overline{(\mathbf{L}_i^m)_1} (\mathbf{L}_j^m)_1 (\mathbf{y}^H \mathbf{L}_i^m (\mathbf{L}_j^m)^H \mathbf{y}) - |\mathbf{y}_1|^2 \right| \leq \delta \tag{C.8}$$

for any $\mathbf{y} \in \mathbb{C}^N$ obeying $\|\mathbf{y}\| = 1$. Letting $\mathbf{y} = [\mathbf{y}_1, \tilde{\mathbf{y}}]$ where $\tilde{\mathbf{y}} \in \mathbb{C}^{N-1}$ and similarly partitioning the sampling vectors as $\mathbf{L}_i^m = [(\mathbf{L}_i^m)_1, \tilde{\mathbf{L}}_i^m]$ we have

$$\begin{aligned} \mathbf{y}^H \mathbf{L}_i^m (\mathbf{L}_j^m)^H \mathbf{y} = (\mathbf{L}_i^m)_1 \overline{\mathbf{y}_1} \overline{(\mathbf{L}_j^m)_1} \mathbf{y}_1 + (\mathbf{L}_i^m)_1 \overline{\mathbf{y}_1} (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \\ + \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m \overline{(\mathbf{L}_j^m)_1} \mathbf{y}_1 + \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}}. \end{aligned} \tag{C.9}$$

This yields

$$\begin{aligned} I_0(\mathbf{y}) &= \left| \frac{1}{M} \sum_{m=1}^{M} (|(\mathbf{L}_i^m)_1|^2 |(\mathbf{L}_j^m)_1|^2 - 1) |\mathbf{y}_1|^2 + |(\mathbf{L}_i^m)_1|^2 (\mathbf{L}_j^m)_1 \overline{\mathbf{y}_1} (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \cdots \right. \\ &\quad \left. \cdots + \; |(\mathbf{L}_j^m)_1|^2 \overline{(\mathbf{L}_i^m)_1} \mathbf{y}_1 \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m + \overline{(\mathbf{L}_i^m)_1} (\mathbf{L}_j^m)_1 \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right| \\ &\leq \left| \frac{1}{M} \sum_{m=1}^{M} (|(\mathbf{L}_i^m)_1|^2 |(\mathbf{L}_j^m)_1|^2 - 1) |\mathbf{y}_1|^2 \right| + \left| \frac{1}{M} \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^2 (\mathbf{L}_j^m)_1 \overline{\mathbf{y}_1} (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right| \\ &\quad + \left| \frac{1}{M} \sum_{m=1}^{M} |(\mathbf{L}_j^m)_1|^2 \overline{(\mathbf{L}_i^m)_1} \mathbf{y}_1 \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m \right| + \left| \frac{1}{M} \sum_{m=1}^{M} \overline{(\mathbf{L}_i^m)_1} (\mathbf{L}_j^m)_1 \tilde{\mathbf{y}}^H \tilde{\mathbf{L}}_i^m (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right|. \end{aligned} \tag{C.10}$$

Due to the independence $\mathbf{L}_i^m$ and $\mathbf{L}_j^m$, and from the fact that they are zero mean, unit variance i.i.d. Gaussian entry vectors, all four terms on the right-hand side are measures of distance to expected values.

For the second and third terms, we have

$$\begin{aligned} \left| \frac{1}{M} \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^2 (\mathbf{L}_j^m)_1 \overline{\mathbf{y}_1} (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right| &\leq \frac{1}{M} \sqrt{\sum_{m=1}^{M} \left| |(\mathbf{L}_i^m)_1|^2 (\mathbf{L}_j^m)_1 \overline{\mathbf{y}_1} \right|^2} \sqrt{\sum_{m=1}^{M} \left| (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right|^2} \\ &\leq |\mathbf{y}_1| \frac{1}{M} \sqrt{N \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^4 |(\mathbf{L}_j^m)_1|^2} \left(\frac{1}{\sqrt{N}} \sum_{m=1}^{M} \left| (\tilde{\mathbf{L}}_j^m)^H \tilde{\mathbf{y}} \right| \right), \end{aligned} \tag{C.11}$$

where the inequalities follow from Cauchy–Schwarz and $\ell_2 \leq \ell_1$. We then invoke Hoeffding's inequality from Proposition 10 in [49]. For any δ_0 and γ, there exists a constant $C(\delta_0, \gamma)$ such that $M \geq C(\delta_0, \gamma) \sqrt{N \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^4 |(\mathbf{L}_j^m)_1|^2}$, where

(C.12) $$\left|\frac{1}{M}\sum_{m=1}^{M}|(\mathbf{L}_i^m)_1|^2(\mathbf{L}_j^m)_1\overline{\mathbf{y}_1}(\tilde{\mathbf{L}}_j^m)^H\tilde{\mathbf{y}}\right| \leq \delta_0\|\tilde{\mathbf{y}}\||\mathbf{y}_1| \leq \delta_0,$$

with probability $1-3e^{-2\gamma N}$. For the final term, we invoke the Bernstein-type inequality of Proposition 16 in [49] per [12], such that for any positive constants δ_0, γ, there exists the constant $C(\delta_0,\gamma)$ with

(C.13) $$M \geq C(\delta_0,\gamma)\left(\sqrt{N\sum_{m=1}^{M}|(\mathbf{L}_i^m)_1|^2|(\mathbf{L}_j^m)_1|^2} + N\max_{m=1\cdots M}|(\mathbf{L}_i^m)_1||(\mathbf{L}_j^m)_1|\right),$$

where

(C.14) $$\left|\frac{1}{M}\sum_{m=1}^{M}\overline{(\mathbf{L}_i^m)_1}(\mathbf{L}_j^m)_1\tilde{\mathbf{y}}^H\tilde{\mathbf{L}}_i^m(\tilde{\mathbf{L}}_j^m)^H\tilde{\mathbf{y}}\right| \leq \delta_0\|\tilde{\mathbf{y}}\|^2 \leq \delta_0$$

with probability $1-2e^{-2\gamma N}$.

To control the remaining terms, we use Chebyshev's inequality, per [12]. For any $\epsilon_0 > 0$ there exists a constant C with $M \geq C\cdot N$ such that the following hold:

(C.15) $$\left|\frac{1}{M}\sum_{m=1}^{M}(|(\mathbf{L}_i^m)_1|^2|(\mathbf{L}_j^m)_1|^2-1)|\mathbf{y}_1|^2\right| \leq \epsilon_0|\mathbf{y}_1|^2,$$

(C.16) $$\left|\frac{1}{M}\sum_{m=1}^{M}(|(\mathbf{L}_i^m)_1|^4|(\mathbf{L}_j^m)_1|^2-1)\right| \leq \epsilon_0, \quad \left|\frac{1}{M}\sum_{m=1}^{M}(|(\mathbf{L}_i^m)_1|^2|(\mathbf{L}_j^m)_1|^4-1)\right| \leq \epsilon_0$$

with probability at least $1-3N^{-2}$. Moreover, from union bound we have

(C.17) $$\max_{m=1\cdots M}|(\mathbf{L}_{i,j}^m)_1| \leq \sqrt{10\log M}$$

with probability at least $1-2N^{-2}$. As in [12], we denote the event that the results from Chebyshev's inequality hold by E_0. Then, in the event E_0, combining all the terms, the inequality

(C.18) $$I_0(\mathbf{y}) \leq \epsilon_0|\mathbf{y}_1|^2 + \delta_0|y_1|\|\tilde{\mathbf{y}}\| + \delta_0\|\tilde{\mathbf{y}}\|^2 \leq \epsilon_0 + 2\delta_0$$

holds with probability at least $1-8e^{-2\gamma N}$. We then follow by the ϵ-net argument of [12] via Lemma 5.4 in [49] to bound the operator norm such that

(C.19) $$\max_{\mathbf{y}\in\mathcal{S}_{\mathbb{C}^N}} I_0(\mathbf{y}) \leq 2\max_{\mathbf{y}\in\mathcal{N}} I_0(\mathbf{y}) \leq 2\epsilon_0 + 4\delta_0,$$

where $\mathcal{S}_{\mathbb{C}^N}$ is the unit sphere in $\mathbb{C}^N$ and $\mathcal{N}$ is an $1/4$-net of $\mathcal{S}_{\mathbb{C}^N}$. Then, choosing appropriate ϵ_0, δ_0, and γ and applying the union bound we have

(C.20) $$\left\|\frac{1}{M}\sum_{m=1}^{M}\overline{(\mathbf{L}_i^m)_1}(\mathbf{L}_j^m)_1\mathbf{L}_i^m(\mathbf{L}_j^m)^H - \mathbf{e}_1\mathbf{e}_1^T\right\| \leq \delta$$

with probability $1-8e^{-\gamma N}$ for $\delta = 2\epsilon_0 + 4\delta_0$, and

$$M \geq C' \left(\sqrt{N \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^4 |(\mathbf{L}_j^m)_1|^2} + \sqrt{N \sum_{m=1}^{M} |(\mathbf{L}_i^m)_1|^2 |(\mathbf{L}_j^m)_1|^2} + N \max_{m=1\cdots M} |(\mathbf{L}_i^m)_1||(\mathbf{L}_j^m)_1| \right). \tag{C.21}$$

From E_0, we have $M \geq C \cdot N$, which gives $M = \mathcal{O}(N \log N)$, where an overall event holds with probability at least $1 - 8e^{-\gamma N} - 5N^{-2}$, hence the proof is complete.

REFERENCES

[1] H. Ammari, J. Garnier, and W. Jing, *Passive array correlation-based imaging in a random waveguide*, Multiscale Model. Simul., 11 (2013), pp. 656–681.

[2] R. Bamler and P. Hartl, *Synthetic aperture radar interferometry*, Inverse Problems, 14 (1998), R1.

[3] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, *Saving phase: Injectivity and stability for phase retrieval*, Appl. Comput. Harmon. Anal., 37 (2014), pp. 106–125.

[4] T. Bendory, Y. C. Eldar, and N. Boumal, *Non-convex phase retrieval from stft measurements*, IEEE Trans. Inform. Theory, 64 (2018), pp. 467–484.

[5] S. Bhojanapalli, B. Neyshabur, and N. Srebro, *Global optimality of local search for low rank matrix recovery*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, UK, 2016, pp. 3873–3881.

[6] P. Blomgren, G. Papanicolaou, and H. Zhao, *Super-resolution in time-reversal acoustics*, J. Acoust. Soc. Amer., 111 (2002), pp. 230–248.

[7] J.-F. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.

[8] T. T. Cai, *Sharp RIP bound for sparse signal and low-rank matrix recovery*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 74–93.

[9] E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM J. Imaging Sci., 6 (2013), pp. 199–225, https://doi.org/10.1137/110848074.

[10] E. Candes and T. Strohmer, *Phaselift: Exact and stable recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math., 66 (2013), pp. 1241–1274.

[11] E. J. Candes, X. Li, and M. Soltanolkotabi, *Phase retrieval from coded diffraction patterns*, Appl. Comput. Harmon. Anal., 39 (2015), pp. 277–299.

[12] E. J. Candes, X. Li, and M. Soltanolkotabi, *Phase retrieval via wirtinger flow: Theory and algorithms*, IEEE Trans. Inform. Theory, 61 (2015), pp. 1985–2007.

[13] E. J. Candes, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[14] A. Chai, M. Moscoso, and G. Papanicolaou, *Array imaging using intensity-only measurements*, Inverse Problems, 27 (2011), 015005, http://stacks.iop.org/0266-5611/27/i=1/a=015005.

[15] J. Chen, L. Wang, X. Zhang, and Q. Gu, *Robust Wirtinger Flow for Phase Retrieval with Arbitrary Corruption*, https://arxiv.org/abs/1704.06256, 2017.

[16] Y. Chen and E. Candes, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2015, pp. 739–747.

[17] Y. Chen and E. J. Candes, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, Comm. Pure Appl. Math., 70 (2017), pp. 822–883.

[18] Y. Chi, Y. M. Lu, and Y. Chen, *Nonconvex Optimization Meets Low-Rank Matrix Factorization: An Overview*, https://arxiv.org/abs/1809.09573, 2018.

[19] E. J. Copeland, M. Sami, and S. Tsujikawa, *Dynamics of dark energy*, Internat. J. Modern Phys. D, 15 (2006), pp. 1753–1935.
[20] L. Demanet and P. Hand, *Stable optimizationless recovery from phaseless linear measurements*, J. Fourier Anal. Appl., 20 (2014), pp. 199–221, https://doi.org/10.1007/s00041-013-9305-2.
[21] L. Demanet and V. Jugnon, *Convex recovery from interferometric measurements*, IEEE Trans. Comput. Imaging, 3 (2017), pp. 282–295.
[22] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
[23] S. Flax and M. O'Donnell, *Phase-aberration correction using signals from point reflectors and diffuse scatterers: Basic principles*, IEEE Trans. Ultrasonics Ferroelectrics Frequency Control, 35 (1988), pp. 758–767.
[24] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Vol. 1, Birkhäuser, Basel, 2013.
[25] J. Garnier, *Imaging in randomly layered media by cross-correlating noisy signals*, Multiscale Model. Simul., 4 (2005), pp. 610–640.
[26] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, *Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks*, IEEE Signal Process. Mag., 22 (2005), pp. 70–84.
[27] R. M. Goldstein and H. Zebker, *Interferometric radar measurement of ocean surface currents*, Nature, 328 (1987), pp. 707–709.
[28] P. Gough and M. Miller, *Displaced ping imaging autofocus for a multi-hydrophone sas*, IEE Proc. Radar Sonar Navigation, 151 (2004), pp. 163–170.
[29] J. E. Guivant and E. M. Nebot, *Optimization of the simultaneous localization and map-building algorithm for real-time implementation*, IEEE Trans. Robotics Automation, 17 (2001), pp. 242–257.
[30] B. Jain and A. Taylor, *Cross-correlation tomography: Measuring dark energy evolution with weak lensing*, Phys. Rev. Lett., 91 (2003), 141302.
[31] X. Li, S. Ling, T. Strohmer, and K. Wei, *Rapid, robust, and reliable blind deconvolution via nonconvex optimization*, Appl. Comput. Harmon. Anal., 47 (2019), pp. 893–934.
[32] O. I. Lobkis and R. L. Weaver, *On the emergence of the green's function in the correlations of a diffuse field*, J. Acoust. Soc. Amer., 110 (2001), pp. 3011–3017.
[33] E. Mason, I. Y. Son, and B. Yazici, *Passive synthetic aperture radar imaging using low-rank matrix recovery methods*, IEEE J. Selected Topics Signal Process., 9 (2015), pp. 1570–1582, https://doi.org/10.1109/JSTSP.2015.2465361.
[34] E. Mason and B. Yazici, *Robustness of lrmr based passive radar imaging to phase errors*, in Proceedings of EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, VDE, 2016, pp. 1–4.
[35] S. Oymak, K. Mohan, M. Fazel, and B. Hassibi, *A simplified approach to recovery conditions for low rank matrices*, in IEEE International Symposium on Information Theory Proceedings, IEEE, 2011, pp. 2318–2322.
[36] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, *Locating the nodes: Cooperative localization in wireless sensor networks*, IEEE Signal Process. Mag., 22 (2005), pp. 54–69.
[37] T. S. Ralston, D. L. Marks, P. S. Carney, and S. A. Boppart, *Interferometric synthetic aperture microscopy*, Nature Phys., 3 (2007), pp. 129–134.
[38] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
[39] B. Recht, W. Xu, and B. Hassibi, *Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization*, in 47th IEEE Conference on Decision and Control, IEEE, 2008, pp. 3065–3070.
[40] B. Recht, W. Xu, and B. Hassibi, *Null space conditions and thresholds for rank minimization*, Math. Program., 127 (2011), pp. 175–202.
[41] T. O. Saebo, *Seafloor Depth Estimation by Means of Interferometric Synthetic Aperture Sonar*, Doctoral thesis, Arctic University of Norway, 2010.
[42] J. C. Schotland, *Quantum imaging and inverse scattering*, Opt. Lett., 35 (2010), pp. 3309–3311.
[43] M. Soltanolkotabi, *Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization*, IEEE Trans. Inform. Theory, 65 (2019), pp. 2374–2400.

[44] I.-Y. Son and B. Yazici, *Passive imaging with multistatic polarimetric radar*, in Proceedings of RadarCon IEEE, 2015, pp. 1584–1589.

[45] I.-Y. Son and B. Yazici, *Passive Polarimetric Multistatic Radar Detection of Moving Targets*, https://arxiv.org/abs/1706.09369, 2017.

[46] P. Stoica, J. Li, and Y. Xie, *On probing signal design for mimo radar*, IEEE Trans. Signal Process., 55 (2007), pp. 4151–4161.

[47] J. Sun, Q. Qu, and J. Wright, *A geometric analysis of phase retrieval*, Found. Comput. Math., 18 (2018), pp. 1131–1198.

[48] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, *Low-Rank Solutions of Linear Matrix Equations via Procrustes Flow*, https://arxiv.org/abs/1507.03566, 2015.

[49] R. Vershynin, *Introduction to the Non-asymptotic Analysis of Random Matrices*, https://arxiv.org/abs/1011.3027, 2010.

[50] S. Wacks and B. Yazici, *Doppler-DPCA and Doppler-ATI: Novel sar modalities for imaging of moving targets using ultra-narrowband waveforms*, IEEE Trans. Comput. Imaging, 4 (2018), pp. 125–136.

[51] S. Wacks and B. Yazici, *Passive synthetic aperture hitchhiker imaging of ground moving targets—part* 1*: Image formation and velocity estimation*, IEEE Trans. Image Process., 23 (2014), pp. 2487–2500, https://doi.org/10.1109/TIP.2014.2302682.

[52] S. Wacks and B. Yazici, *Passive synthetic aperture hitchhiker imaging of ground moving targets—part* 2*: Performance analysis*, IEEE Trans. Image Process., 23 (2014), pp. 4126–4138, https://doi.org/10.1109/TIP.2014.2336543.

[53] I. Waldspurger, A. d'Aspremont, and S. Mallat, *Phase recovery, maxcut and complex semidefinite programming*, Math. Program., 149 (2015), pp. 47–81, https://doi.org/10.1007/s10107-013-0738-9.

[54] G. Wang, G. B. Giannakis, and Y. C. Eldar, *Solving systems of random quadratic equations via truncated amplitude flow*, IEEE Trans. Inform. Theory, 64 (2018), pp. 773–794.

[55] L. Wang, I.-Y. Son, and B. Yazici, *Passive imaging using distributed apertures in multiple-scattering environments*, Inverse Problems, 26 (2010), 065002.

[56] L. Wang, C. Yarman, and B. Yazici, *Doppler-Hitchhiker: A novel passive synthetic aperture radar using ultranarrowband sources of opportunity*, IEEE Trans. Geosci. Remote Sensing, 49 (2011), pp. 3521–3537, https://doi.org/10.1109/TGRS.2011.2142420.

[57] L. Wang and B. Yazici, *Bistatic synthetic aperture radar imaging using narrowband continuous waveforms*, IEEE Trans. Image Process., 21 (2012), pp. 3673–3686, https://doi.org/10.1109/TIP.2012.2193134.

[58] L. Wang and B. Yazici, *Passive imaging of moving targets exploiting multiple scattering using sparse distributed apertures*, Inverse Problems, 28 (2012), 125009.

[59] L. Wang and B. Yazici, *Passive imaging of moving targets using sparse distributed apertures*, SIAM J. Imaging Sci., 5 (2012), pp. 769–808.

[60] L. Wang and B. Yazici, *Ground moving target imaging using ultranarrowband continuous wave synthetic aperture radar*, IEEE Trans. Geosci. Remote Sensing, 51 (2013), pp. 4893–4910, https://doi.org/10.1109/TGRS.2013.2272603.

[61] L. Wang and B. Yazici, *Bistatic synthetic aperture radar imaging of moving targets using ultra-narrowband continuous waveforms*, SIAM J. Imaging Sci., 7 (2014), pp. 824–866, https://doi.org/10.1137/130906714.

[62] L. Wang, X. Zhang, and Q. Gu, *A Unified Computational and Statistical Framework for Nonconvex Low-Rank Matrix Estimation*, https://arxiv.org/abs/1610.05275, 2016.

[63] C. Yarman, L. Wang, and B. Yazici, *Doppler synthetic aperture hitchhiker imaging*, Inverse Problems, 26 (2010), 065006, http://stacks.iop.org/0266-5611/26/i=6/a=065006.

[64] C. Yarman and B. Yazici, *Synthetic aperture hitchhiker imaging*, IEEE Trans. Image Process., 17 (2008), pp. 2156–2173.

[65] C. E. Yarman, B. Yazici, and M. Cheney, *Bistatic synthetic aperture radar imaging for arbitrary flight trajectories*, IEEE Trans. Image Process., 17 (2008), pp. 84–93.

[66] B. Yonel, I.-Y. Son, and B. Yazici, *Exact Imaging of Extended Targets Using Multistatic Interferometric Measurements*, https://arxiv.org/abs/1905.10459, 2019.

[67] B. Yonel, I.-Y. Son, and B. Yazici, *Generalized wirtinger flow for passive polarimetric reconstruction of extended dipole targets*, in Proceedings of RadarCon18, IEEE, 2018, pp. 1353–1358.

[68] H. Zhang and L. Cheng, *Restricted strong convexity and its applications to convergence analysis of gradient-type methods in convex optimization*, Optim. Lett., 9 (2015), pp. 961–979.

[69] H. Zhang, Y. Chi, and Y. Liang, *Median-Truncated Nonconvex Approach for Phase Retrieval with Outliers*, Preprint, 2017.

[70] H. Zhang and Y. Liang, *Reshaped Wirtinger flow for solving quadratic systems of equations*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2016.

[71] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, *Reshaped Wirtinger Flow and Incremental Algorithms for Solving Quadratic Systems of Equations*, preprint, 2017.

[72] X. Zhang, L. Wang, Y. Yu, and Q. Gu, *A primal-dual analysis of global optimality in nonconvex low-rank matrix recovery*, in Proceedings of the International Conference on Machine Learning, 2018, pp. 5857–5866.

[73] Q. Zheng and J. Lafferty, *A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2015, pp. 109–117.