Stable optimizationless recovery from phaseless linear measurements

Laurent Demanet and Paul Hand

Massachusetts Institute of Technology, Department of Mathematics,
77 Massachusetts Avenue, Cambridge, MA 02139

August 2012, Revised May 2013

Abstract

We address the problem of recovering an n-vector from m linear measurements lacking sign or phase information. We show that lifting and semidefinite relaxation suffice by themselves for stable recovery in the setting of $m = O(n \log n)$ random sensing vectors, with high probability. The recovery method is optimizationless in the sense that trace minimization in the PhaseLift procedure is unnecessary. That is, PhaseLift reduces to a feasibility problem. The optimizationless perspective allows for a Douglas-Rachford numerical algorithm that is unavailable for PhaseLift. This method exhibits linear convergence with a favorable convergence rate and without any parameter tuning.

1 Introduction

We study the recovery of a vector $\mathbf{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ from the set of phaseless linear measurements

$$|\langle \mathbf{x}_0, \mathbf{z}_i \rangle| \text{ for } i = 1, \ldots, m,$$

where $\mathbf{z}_i \in \mathbb{R}^n$ or $\mathbb{C}^n$ are known random sensing vectors. Such amplitude-only measurements arise in a variety of imaging applications, such as X-ray crystallography [15, 17, 5], optics [23], and microscopy [16]. We seek stable and efficient methods for finding $\mathbf{x}_0$ using as few measurements as possible.

This recovery problem is difficult because the set of real or complex numbers with a given magnitude is nonconvex. In the real case, there are 2^m possible assignments of sign to the m phaseless measurements. Hence, exhaustive searching is infeasible. In the complex case, the situation is even worse, as there are a continuum of phase assignments to consider. A method based of alternated projections avoids an exhaustive search but does not always converge toward a solution [11, 12, 14].

In [5, 7, 8], the authors convexify the problem by lifting it to the space of $n \times n$ matrices, where $\mathbf{x}\mathbf{x}^*$ is a proxy for the vector $\mathbf{x}$. A key motivation for this lifting is that the nonconvex measurements on vectors become linear measurements on matrices [1]. The rank-1 constraint is then relaxed to a trace minimization over the cone of positive semi-definite matrices, as is now standard in matrix

completion [19]. This convex program is called PhaseLift in [7], where it is shown that $\mathbf{x}_0$ can be found robustly in the case of random $\mathbf{z}_i$, if $m = O(n \log n)$. The matrix minimizer is unique, which in turn determines $\mathbf{x}_0$ up to a global phase.

The contribution of the present paper is to show that trace minimization is unnecessary in this lifting framework for the phaseless recovery problem. The vector $\mathbf{x}_0$ can be recovered robustly by an optimizationless convex problem: one of finding a positive semi-definite matrix that is consistent with linear measurements. We prove there is only one such matrix, provided that there are $O(n \log n)$ measurements. In other words, the phase recovery problem can be solved by intersecting two convex sets, without minimizing an objective. We show empirically that two algorithms converge linearly (exponentially fast) toward the solution. We remark that these methods are simpler than methods for PhaseLift because they require less or no parameter tuning. A result subsequent to the posting of this paper has improved the number of required measurements to $O(n)$ by considering an alternative construction of the dual certificate that allows tighter probabilistic bounds [6].

In [2], the authors show that the complex phaseless recovery problem from random measurements is *determined* if $m \geq 4n - 2$ (with probability one). This means that the $\mathbf{x}$ satisfying $|\langle \mathbf{x}, \mathbf{z}_i \rangle| = |\langle \mathbf{x}_0, \mathbf{z}_i \rangle|$ is unique and equal to $\mathbf{x}_0$, regardless of the method used to find it. A corollary of the analysis in [7], and of the present paper, is that this property is stable under perturbations of the data, provided $m = O(n \log n)$. This determinacy is in contrast to compressed sensing and matrix completion, where a prior (sparsity, low-rank) is used to select a solution of an otherwise underdetermined system of equations. The relaxation of this prior (ℓ_1 norm, nuclear norm) is then typically shown to determine the same solution. No such prior is needed here; the semi-definite relaxation helps find the solution, not determine it.

The determinacy of the recovery problem over $n \times n$ matrices may be unexpected because there are n^2 unknowns and only $O(n \log n)$ measurements. What compensates for the apparent lack of data is the fact that the matrix we seek has rank one and is thus on the edge of the cone of positive semi-definite matrices. Most perturbed matrices that are consistent with the measurements cease to remain positive semi-definite. In other words, the positive semi-definite cone $\mathbf{X} \succeq 0$ is "spiky" around a rank-1 matrix $\mathbf{X}_0$. That is, with high probability, particular random hyperplanes that contain $\mathbf{X}_0$ and have large enough codimension will have no other intersection with the cone.

The present paper does not advocate for fully abandoning trace minimization in the context of phase retrieval. The structure of the sensing matrices appears to affect the number of measurements required for recovery. Consider measurements of the form $\mathbf{x}_0^* \Phi \mathbf{x}_0$, for some Φ. Numerical simulations (not shown) suggest that $O(n^2)$ measurements are needed if Φ is a matrix with Gaussian i.i.d. entries. On the other hand, it was shown in [19] that minimization of the nuclear norm constrained by $\mathrm{Tr}(\mathbf{X}\Phi) = \mathbf{x}_0^* \Phi \mathbf{x}_0$ recovers $\mathbf{x}_0 \mathbf{x}_0^*$ with high probability as soon as $m = O(n \log n)$. Other numerical observations (not shown) suggest that it is the symmetric, positive semi-definite character of Φ that allows for optimizationless recovery.

The present paper owes much to [7], as our analysis is very similar to theirs. We wish to also reference the papers [20, 22], where phase recovery is cast as synchronization problem and solved via a semi-definite relaxation of max-cut type over the complex torus (i.e., the magnitude information is first factored out.) The idea of lifting and semi-definite relaxation was introduced very successfully for the max-cut problem in [13]. The paper [20] also introduces a fast and efficient method based on eigenvectors of the graph connection Laplacian for solving the angular synchronization problem. The performance of this latter method was further studied in [3].

1.1 Problem Statement and Main Result

Let $x_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ be a vector for which we have the m measurements $|\langle \mathbf{x}_0, \mathbf{z}_i \rangle| = \sqrt{b_i}$, for independent sensing vectors $\mathbf{z}_i$ distributed uniformly on the unit sphere. We write the phaseless recovery problem for $\mathbf{x}_0$ as

$$\text{Find } \mathbf{x} \text{ such that } A(\mathbf{x}) = \mathbf{b}, \tag{1}$$

where $A : \mathbb{R}^n \to \mathbb{R}^m$ is given by $A(\mathbf{x})_i = |\langle \mathbf{x}, \mathbf{z}_i \rangle|^2$, and $A(\mathbf{x}_0) = \mathbf{b}$.

Problem (1) can be convexified by lifting it to a matrix recovery problem. Let $\mathcal{A}$ and its adjoint be the linear operators

$$\mathcal{A} : \quad \mathcal{H}^{n \times n} \to \mathbb{R}^m \qquad\qquad \mathcal{A}^* : \quad \mathbb{R}^m \to \mathcal{H}^{n \times n}$$

$$\mathbf{X} \quad \mapsto \{\mathbf{z}_i^* \mathbf{X} \mathbf{z}_i\}_{i=1,\dots,m}, \qquad\qquad \lambda \quad \mapsto \sum_i \lambda_i \mathbf{z}_i \mathbf{z}_i^*,$$

where $\mathcal{H}^{n \times n}$ is the space of $n \times n$ Hermitian matrices. Observe that $\mathcal{A}(\mathbf{x}\mathbf{x}^*) = A(\mathbf{x})$ for all vectors $\mathbf{x}$. Letting $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$, we note that $\mathcal{A}(\mathbf{X}_0) = b$. We emphasize that $\mathcal{A}$ is linear in $\mathbf{X}$ whereas A is nonlinear in $\mathbf{x}$.

The matrix recovery problem we consider is

$$\text{Find } \mathbf{X} \succeq 0 \text{ such that } \mathcal{A}(\mathbf{X}) = \mathbf{b}. \tag{2}$$

Without the positivity constraint, there would be multiple solutions whenever $m < \frac{(n+1)n}{2}$. We include the constraint in order to allow for recovery in this classically underdetermined regime.

Our main result is that the matrix recovery problem (2) has a unique solution when there are $O(n \log n)$ measurements.

Theorem 1. Let $\mathbf{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ and $\mathbf{X}_0 = \mathbf{x}_0 \mathbf{x}_0^*$. Let $m \geq cn \log n$ for a sufficiently large c. With high probability, $\mathbf{X} = \mathbf{X}_0$ is the unique solution to $\mathbf{X} \succeq 0$ and $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. This probability is at least $1 - e^{-\gamma \frac{m}{n}}$, for some $\gamma > 0$.

As a result, the phaseless recovery problem has a unique solution, up to a global phase, with $O(n \log n)$ measurements. In the real-valued case, the problem is determined up to a minus sign.

Corollary 2. Let $\mathbf{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$. Let $m \geq cn \log n$ for a sufficiently large c. With high probability, $\{e^{i\phi} \mathbf{x}_0\}$ are the only solutions to $A(\mathbf{x}) = \mathbf{b}$. This probability is at least $1 - e^{-\gamma \frac{m}{n}}$, for some $\gamma > 0$.

Theorem 1 suggests ways of recovering $\mathbf{x}_0$. If an $\mathbf{X} \in \{\mathbf{X} \succeq 0\} \cap \{\mathbf{X} \mid \mathcal{A}(\mathbf{X}) = \mathbf{b}\}$ can be found, $\mathbf{x}_0$ is given by the leading eigenvector of $\mathbf{X}$. See Section 6 for more details on how to find $\mathbf{X}$.

1.2 Stability result

In practical applications, measurements are contaminated by noise. To show stability of optimizationless recovery, we consider the model

$$A(\mathbf{x}) + \nu = \mathbf{b},$$

where ν corresponds to a noise term with bounded ℓ_2 norm, $\|\nu\|_2 \leq \varepsilon$. The corresponding noisy variant of (1) is

$$\text{Find } \mathbf{x} \text{ such that } \|A(\mathbf{x}) - \mathbf{b}\|_2 \leq \varepsilon \|\mathbf{x}_0\|_2^2. \tag{3}$$

We note that all three terms in (3) scale quadratically in $\mathbf{x}$ or $\mathbf{x}_0$.

Problem (3) can be convexified by lifting it to the space of matrices. The noisy matrix recovery problem is

$$\text{Find } \mathbf{X} \succeq 0 \text{ such that } \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \varepsilon\|\mathbf{X}_0\|_2. \tag{4}$$

We show that all feasible $\mathbf{X}$ are within an $O(\varepsilon)$ ball of $\mathbf{X}_0$ provided there are $O(n \log n)$ measurements.

Theorem 3. *Let $\mathbf{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ and $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$. Let $m \geq cn \log n$ for a sufficiently large c. With high probability,*

$$\mathbf{X} \succeq 0 \text{ and } \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \varepsilon\|\mathbf{X}_0\|_2 \implies \|\mathbf{X} - \mathbf{X}_0\|_2 \leq C\varepsilon\|\mathbf{X}_0\|_2,$$

for some $C > 0$. This probability is at least $1 - e^{-\gamma\frac{m}{n}}$, for some $\gamma > 0$.

As a result, the phaseless recovery problem is stable with $O(n \log n)$ measurements.

Corollary 4. *Let $\mathbf{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$. Let $m \geq cn \log n$ for a sufficiently large c. With high probability,*

$$\|A(\mathbf{x}) - \mathbf{b}\|_2 \leq \varepsilon\|x_0\|_2^2 \implies \left\|\mathbf{x} - e^{i\phi}\mathbf{x}_0\right\|_2 \leq C\varepsilon\|\mathbf{x}_0\|_2,$$

for some $\phi \in [0, 2\pi)$, and for some $C > 0$. This probability is at least $1 - e^{-\gamma\frac{m}{n}}$, for some $\gamma > 0$.

Theorem 3 ensures that numerical methods can be used to find $\mathbf{X}$. See Section 6 for ways of finding $\mathbf{X} \in \{\mathbf{X} \succeq 0\} \cap \{\mathcal{A}(\mathbf{X}) \approx \mathbf{b}\}$. As the recovered matrix may have large rank, we approximate $\mathbf{x}_0$ with the leading eigenvector of $\mathbf{X}$.

1.3   Organization of this paper

In Section 2, we prove a lemma containing the central argument for the proof of Theorem 1. Its assumptions involve ℓ_1-isometry properties and the existence of an inexact dual certificate. Section 2.3 provides the proof of Theorem 1 in the real-valued case. It cites [7] for the ℓ_1-isometry properties and Section 3 for existence of an inexact dual certificate. In Section 3 we construct an inexact dual certificate and show that it satisfies the required properties in the real-valued case. In section 4 we prove Theorem 3 on stability in the real-valued case. In Section 5, we discuss the modifications in the complex-valued case. In Section 6, we present computational methods for the optimizationless problem with comparisons to PhaseLift. We also simulate them to establish stability empirically.

1.4   Notation

We use boldface for variables representing vectors or matrices. We use normal typeface for scalar quantities. Let $z_{i,k}$ denote the kth entry of the vector $\mathbf{z}_i$. For two matrices, let $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Tr}(\mathbf{Y}^*\mathbf{X})$ be the Hilbert-Schmidt inner product. Let σ_i be the singular values of the matrix $\mathbf{X}$. We define the norms

$$\|\mathbf{X}\|_p = \left(\sum_i \sigma_i^p\right)^{1/p}.$$

In particular, we write the Frobenius norm of $\mathbf{X}$ as $\|\mathbf{X}\|_2$. We write the spectral norm of $\mathbf{X}$ as $\|\mathbf{X}\|$.

An n-vector $\mathbf{x}$ generates a decomposition of $\mathbb{R}^n$ or $\mathbb{C}^n$ into two subspaces. These subspaces are the span of $\mathbf{x}$ and the span of all vectors orthogonal to $\mathbf{x}$. Abusing notation, we write these

subspaces as $\mathbf{x}$ and $\mathbf{x}^\perp$. The space of n-by-n matrices is correspondingly partitioned into the four subspaces $\mathbf{x} \otimes \mathbf{x}$, $\mathbf{x} \otimes \mathbf{x}^\perp$, $\mathbf{x}^\perp \otimes \mathbf{x}$, and $\mathbf{x}^\perp \otimes \mathbf{x}^\perp$, where $\otimes$ denotes the outer product. We write $T_{\mathbf{x}}$ for the set of symmetric matrices which lie in the direct sum of the first three subspaces, namely $T_{\mathbf{x}} = \{\mathbf{xy}^* + \mathbf{yx}^* \mid \mathbf{y} \in \mathbb{R}^n \text{ or } \mathbb{C}^n\}$. Correspondingly, we write $T_{\mathbf{x}}^\perp$ for the set of symmetric matrices in the fourth subspace. We note that $T_{\mathbf{x}}^\perp$ is the orthogonal complement of $T_{\mathbf{x}}$ with respect to the Hilbert-Schmidt inner product. Let $\mathbf{e}_1$ be the first coordinate vector. For short, let $T = T_{\mathbf{e}_1}$ and $T^\perp = T_{\mathbf{e}_1}^\perp$. We denote the projection of $\mathbf{X}$ onto T as either $\mathcal{P}_T \mathbf{X}$ or $\mathbf{X}_T$. We denote projections onto $T^\perp$ similarly.

We let $\mathbf{I}$ be the $n \times n$ identity matrix. We denote the range of $\mathcal{A}^*$ by $\mathcal{R}(\mathcal{A}^*)$.

2   Proof of Main Result

Because of scaling and the property that the measurement vectors $\mathbf{z}_i$ come from a rotationally invariant distribution, we take $\mathbf{x}_0 = \mathbf{e}_1$ without loss of generality. Because all measurements scale with the length $\|\mathbf{z}_i\|_2$, it is equivalent to establish the result for independent unit normal sensing vectors $\mathbf{z}_i$. To prove Theorem 1, we use an argument based on inexact dual certificates and ℓ_1-isometry properties of $\mathcal{A}$. This argument parallels that of [7]. We directly use the ℓ_1-isometry properties they establish, but we require different properties on the inexact dual certificate.

2.1   About Dual Certificates

As motivation for the introduction of an inexact dual certificate in the next section, observe that if $\mathcal{A}$ is injective on T, and if there exists a (exact) dual certificate $\mathbf{Y} \in \mathcal{R}(\mathcal{A}^*)$ such that

$$\mathbf{Y}_T = 0 \quad \text{and} \quad \mathbf{Y}_{T^\perp} \succ 0,$$

then $\mathbf{X}_0$ is the only solution to $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. This is because

$$0 = \langle \mathbf{X} - \mathbf{X}_0, \mathbf{Y} \rangle = \langle \mathbf{X}_{T^\perp}, \mathbf{Y}_{T^\perp} \rangle \Rightarrow \mathbf{X}_{T^\perp} = 0 \Rightarrow \mathbf{X} = \mathbf{X}_0,$$

where the first equality is because $\mathbf{Y} \in \mathcal{R}(\mathcal{A}^*)$ and $\mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{X}_0)$. The last implication follows from injectivity on T.

Conceptually, $\mathbf{Y}$ arises as a Lagrange multiplier, dual to the constraint $\mathbf{X} \succeq 0$ in the feasibility problem

$$\min 0 \quad \text{such that} \quad \mathcal{A}(\mathbf{X}) = \mathbf{b}, \qquad \mathbf{X} \succeq 0.$$

Dual feasibility requires $\mathbf{Y} \succeq 0$. As visualized in Figure 1a, $\mathbf{Y}$ acts as a vector normal to a codimension-1 hyperplane that separates the lower-dimensional space of solutions $\{\mathcal{A}(\mathbf{X}) = b\}$ from the positive matrices not in T. The condition $\mathbf{Y}_{T^\perp} \succ 0$ is further needed to ensure that this hyperplane only intersects the cone along T, ensuring uniqueness of the solution.

The nullspace condition $\mathbf{Y}_T = 0$ is what makes the certificate exact. As $\mathbf{Y} \in \mathcal{R}(\mathcal{A}^*)$, $\mathbf{Y}$ must be of the form $\sum_i \lambda_i \mathbf{z}_i \mathbf{z}_i^*$. The strict requirement that $\mathbf{Y}_T = 0$ would force the λ_i to be complicated (at best algebraic) functions of all the $\mathbf{z}_j$, $j = 1, \ldots, m$. We follow [7] in constructing instead an inexact dual certificate, such that $\mathbf{Y}_T$ is close to but not equal to 0, and for which the λ_i are more tractable (quadratic) polynomials in the $\mathbf{z}_i$. A careful inspection of the injectivity properties of $\mathcal{A}$, in the form of the RIP-like condition in [7], is what allows the relaxation of the nullspace condition on $\mathbf{Y}$.

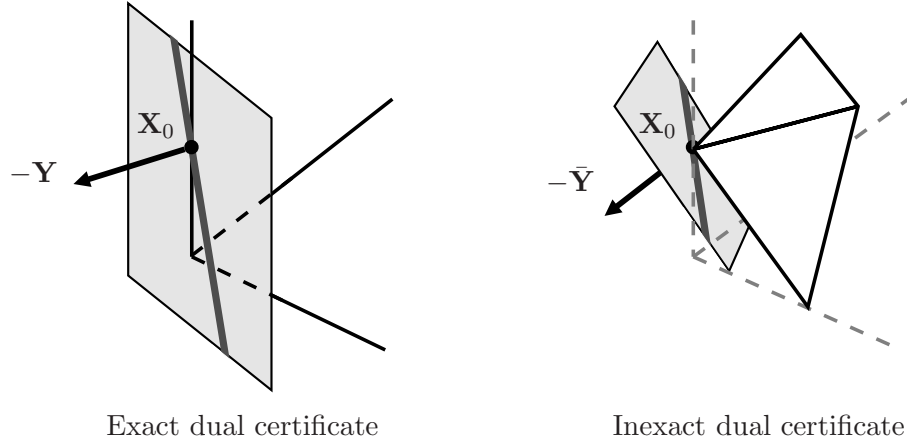Exact dual certificate   Inexact dual certificate

Figure 1: The graphical interpretation of the exact and inexact dual certificates. The positive axes represent the cone of positive matrices. The thick gray line represents the solutions to $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. The exact dual certificate $\mathbf{Y}$ is a normal vector to a hyperplane that separates the space of solutions from positive matrices. When the dual certificate is inexact, we use the fact that ℓ_1-isometry properties imply $\mathbf{X}$ is restricted to the cone (8). The inexact dual certificate $\bar{\mathbf{Y}}$ is normal to a hyperplane that separates $\mathbf{X}_0$ from the rest of this restricted cone. As shown, the hyperplane normal to $\bar{\mathbf{Y}}$ does not separate $\mathbf{X}_0$ from positive matrices.

2.2 Central Lemma on Inexact Dual Certificates

With further information about feasible $\mathbf{X}$, we can relax the property that $\mathbf{Y}_T$ is exactly zero. In [7], the authors show that all feasible $\mathbf{X}$ lie in a cone that is approximately $\{\|\mathbf{X}_{T^\perp}\|_1 \geq \|\mathbf{X}_T - \mathbf{X}_0\|\}$, provided there are $O(n)$ measurements. As visualized in Figure 1b, $\bar{\mathbf{Y}}$ acts as a vector normal to a hyperplane that separates $\mathbf{X}_0$ from the rest of this cone. The proof of Theorem 1 hinges on the existence of such an inexact dual certificate, along with ℓ_1-isometry properties that establish $\mathbf{X}$ is in this cone with high probability.

Lemma 1. *Suppose that $\mathcal{A}$ satisfies*

$$m^{-1}\|\mathcal{A}(\mathbf{X})\|_1 \leq (1+\delta)\|\mathbf{X}\|_1 \qquad \text{for all } \mathbf{X} \succeq 0, \tag{5}$$

$$m^{-1}\|\mathcal{A}(\mathbf{X})\|_1 \geq 0.94(1-\delta)\|\mathbf{X}\| \quad \text{for all } \mathbf{X} \in T, \tag{6}$$

for some $\delta \leq 1/9$. Suppose that there exists $\bar{\mathbf{Y}} \in \mathcal{R}(\mathcal{A}^*)$ satisfying

$$\|\bar{\mathbf{Y}}_T\|_1 \leq 1/2 \quad \text{and} \quad \bar{\mathbf{Y}}_{T^\perp} \succeq \mathbf{I}_{T^\perp}. \tag{7}$$

Then, $\mathbf{X}_0$ is the unique solution to (2).

Proof of Lemma 1. Let $\mathbf{X}$ solve (2), and let $\mathbf{H} = \mathbf{X} - \mathbf{X}_0$. We start by showing, as in [7], that the ℓ_1-isometry conditions (5)–(6) guarantee solutions lie on the cone

$$\|\mathbf{H}_{T^\perp}\|_1 \geq \frac{0.94(1-\delta)}{1+\delta}\|\mathbf{H}_T\|. \tag{8}$$

This is because

$$0.94(1-\delta)\|\mathbf{H}_T\| \leq m^{-1}\|\mathcal{A}(\mathbf{H}_T)\|_1 = m^{-1}\|\mathcal{A}(\mathbf{H}_{T^\perp})\|_1 \leq (1+\delta)\|\mathbf{H}_{T^\perp}\|_1,$$

6

where the equality comes from $0 = \mathcal{A}(\mathbf{H}) = \mathcal{A}(\mathbf{H}_T) + \mathcal{A}(\mathbf{H}_{T^\perp})$, and the two inequalities come from the ℓ_1-isometry properties (5)–(6) and the fact that $\mathbf{H}_{T^\perp} \succeq 0$.

Because $\mathcal{A}(\mathbf{H}) = 0$ and $\bar{\mathbf{Y}} \in \mathcal{R}(\mathcal{A}^*)$,

$$
\begin{aligned}
0 &= \langle \mathbf{H}, \bar{\mathbf{Y}} \rangle \\
&= \langle \mathbf{H}_T, \bar{\mathbf{Y}}_T \rangle + \langle \mathbf{H}_{T^\perp}, \bar{\mathbf{Y}}_{T^\perp} \rangle \\
&\geq \|\mathbf{H}_{T^\perp}\|_1 - \frac{1}{2}\|\mathbf{H}_T\| & (9) \\
&\geq \left(\frac{0.94(1-\delta)}{1+\delta} - \frac{1}{2} \right) \|\mathbf{H}_T\|, & (10)
\end{aligned}
$$

where (9) and (10) follow from (7) and (8), respectively. Because the constant in (10) is positive, we conclude $\mathbf{H}_T = 0$. Then, (9) establishes $\mathbf{H}_{T^\perp} = 0$. $\qquad\square$

2.3   Proof of Theorem 1 and Corollary 2

We use Lemma 1 to prove Theorem 1 for real-valued signals.

Proof of Theorem 1. We need to show that (5)–(7) hold with high probability if $m > cn \log n$ for some c. Lemmas 3.1 and 3.2 in [7] show that (5) and (6) both hold with probability of at least $1 - 3e^{-\gamma_1 m}$ provided $m > c_1 n$ for some c_1. In section 3, we construct $\bar{\mathbf{Y}} \in \mathcal{R}(\mathcal{A}^*)$. As per Lemma 2, $\|\bar{\mathbf{Y}}_T\|_1 \leq 1/2$ with probability at least $1 - e^{-\gamma_2 m/n}$ if $m > c_2 n$. As per Lemma 3, $\|\bar{\mathbf{Y}}_{T^\perp} - 2\mathbf{I}_{T^\perp}\| \leq 1$ with probability at least $1 - 2e^{-\gamma_2 m/\log n}$ if $m > c_3 n \log n$. Hence, $\bar{\mathbf{Y}}_{T^\perp} \succeq \mathbf{I}_{T^\perp}$ with at least the same probability. Hence, all of the conditions of Lemma 1 hold with probability at least $1 - e^{-\gamma m/n}$ if $m > cn \log n$ for some c and γ. $\qquad\square$

The proof of Corollary 2 is immediate because, with high probability, Theorem 1 implies

$$
A(\mathbf{x}_1) = A(\mathbf{x}_0) \Rightarrow \mathbf{x}_1 \mathbf{x}_1^* = \mathbf{x}_0 \mathbf{x}_0^* \Rightarrow \mathbf{x}_1 = e^{i\phi} \mathbf{x}_0.
$$

3   Existence of Inexact Dual Certificate

To use Lemma 1 in the proof of Theorem 1, we need to show that there exists an inexact dual certificate satisfying (7) with high probability. Our inexact dual certificate vector is different from that in [7], but we use identical tools for its construction and analysis. We also adopt similar notation.

We note that $\mathcal{A}^*\mathcal{A}(\mathbf{X}) = \sum_i \langle \mathbf{X}, \mathbf{z}_i \mathbf{z}_i^* \rangle \mathbf{z}_i \mathbf{z}_i^*$, which can alternatively be written as

$$
\mathcal{A}^*\mathcal{A} = \sum_{i=1}^m \mathbf{z}_i \mathbf{z}_i^* \otimes \mathbf{z}_i \mathbf{z}_i^*.
$$

We let $\mathcal{S} = \mathbb{E}[\mathbf{z}_i \mathbf{z}_i^* \otimes \mathbf{z}_i \mathbf{z}_i^*]$. The operator $\mathcal{S}$ is invertible. It and its inverse are given by

$$
\mathcal{S}(\mathbf{X}) = 2\mathbf{X} + \mathrm{Tr}(\mathbf{X})\mathbf{I},
$$

$$
\mathcal{S}^{-1}(\mathbf{X}) = \frac{1}{2}\left(\mathbf{X} - \frac{1}{n+2}\mathrm{Tr}(\mathbf{X})\mathbf{I} \right). \tag{11}
$$

We define the inexact dual certificate

$$
\bar{\mathbf{Y}} = \frac{1}{m}\sum_{i=1}^m \mathbf{1}_{E_i} \mathbf{Y}_i, \tag{12}
$$

where

$$\mathbf{Y}_i = \left[\frac{3}{n+2} \|\mathbf{z}_i\|_2^2 - z_{i,1}^2 \right] \mathbf{z}_i \mathbf{z}_i^*, \tag{13}$$

$$E_i = \{ |z_{i,1}| \le \sqrt{2\beta \log n} \} \cap \{ \|\mathbf{z}_i\|_2 \le \sqrt{3n} \}. \tag{14}$$

Alternatively, we can write the inexact dual certificate vector as

$$\bar{\mathbf{Y}} = \frac{1}{m} \mathcal{A}^* \left(\mathbf{1}_E \circ \mathcal{A}\mathcal{S}^{-1} 2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*) \right), \tag{15}$$

where $(\mathbf{1}_E)_i = 1_{E_i}$ and $\circ$ is the elementwise product of vectors. In our notation, truncated quantities have overbars. We subsequently omit the subscript i in $\mathbf{z}_i$ when it is implied by context.

3.1   Motivation for the Dual Certificate

For ease of understanding, we first consider a candidate dual certificate given by

$$\widetilde{\mathbf{Y}} = \frac{1}{m} \mathcal{A}^* \mathcal{A}\mathcal{S}^{-1} 2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*).$$

The motivation for this candidate is twofold: $\widetilde{\mathbf{Y}} \in \mathcal{R}(\mathcal{A}^*)$, and $\widetilde{\mathbf{Y}} \approx 2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*)$ as $m \to \infty$ because $\mathbb{E}[\mathcal{A}^*\mathcal{A}] = m\mathcal{S}$. In this limit, $\widetilde{Y}$ becomes an exact dual certificate. For finite m, it should be close but inexact. We can write

$$\widetilde{\mathbf{Y}} = \frac{1}{m} \sum_i \mathbf{Y}_i,$$

where $\mathbf{Y}_i$ is an independent sample of the random matrix

$$\left[\frac{3}{n+2} \|\mathbf{z}\|_2^2 - z_1^2 \right] \mathbf{z}\mathbf{z}^*,$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$. Because the vector Bernstein inequality requires bounded vectors, we truncate the dual certificate in the same manner as [7]. That is, we consider $1_{E_i} \mathbf{Y}_i$, completing the derivation of (12).

3.2   Bounds on $\bar{\mathbf{Y}}$

We define $\pi(\beta) = \mathbb{P}(E^c)$, where E is the event given by

$$E = \{ |z_1| \le \sqrt{2\beta \log n} \} \cap \{ \|\mathbf{z}\|_2 \le \sqrt{3n} \}, \tag{16}$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$. In [7], the authors provide the bound $\pi(\beta) \le \mathbb{P}(|z_1| > \sqrt{2\beta \log n}) + \mathbb{P}(\|\mathbf{z}\|_2^2 > 3n) \le n^{-\beta} + e^{-n/3}$, which holds if $2\beta \log n \ge 1$.

We now present two lemmas that establish that $\bar{\mathbf{Y}}$ is approximately $2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*)$, and is thus an inexact dual certificate satisfying (7).

Lemma 2. *Let $\bar{\mathbf{Y}}$ be given by (12). There exists positive γ and c such that for sufficiently large n*

$$\mathbb{P}\left(\left\| \bar{\mathbf{Y}}_T \right\|_1 \ge \frac{1}{2} \right) \le \exp\left(-\gamma \frac{m}{n} \right)$$

if $m \ge cn$.

Lemma 3. *Let $\bar{\mathbf{Y}}$ be given by (12). There exists positive γ and c such that for sufficiently large n*

$$\mathbb{P}\left(\left\| \bar{\mathbf{Y}}_{T^\perp} - 2\mathbf{I}_{T^\perp} \right\| \ge 1 \right) \le 2\exp\left(-\gamma \frac{m}{\log n} \right)$$

if $m \ge cn \log n$.

3.3 Proof of Lemma 2: $\bar{\mathbf{Y}}$ on T

We prove Lemma 2 in a way that parallels the corresponding proof in [7]. Observe that

$$\|\bar{\mathbf{Y}}_T\|_1 \leq \sqrt{2}\|\bar{\mathbf{Y}}_T\|_2 \leq 2\|\bar{\mathbf{Y}}_T\mathbf{e}_1\|_2,$$

where the first inequality follows because $\bar{\mathbf{Y}}_T$ has rank at most 2, and the second inequality follows because $\bar{\mathbf{Y}}_T$ can be nonzero only in its first row and column. We can write

$$\bar{\mathbf{Y}}_T\mathbf{e}_1 = \frac{1}{m}\sum_{i=1}^{m}\bar{\mathbf{y}}_i,$$

where $\bar{\mathbf{y}}_i = \mathbf{y}_i 1_{E_i}$, and $\mathbf{y}_i$ are independent samples of

$$\mathbf{y} = \left[\frac{3}{n+2}\|\mathbf{z}\|_2^2 - z_1^2\right]z_1\mathbf{z} =: \xi z_1\mathbf{z}.$$

To bound the ℓ_2 norm of $\bar{\mathbf{Y}}_T\mathbf{e}_1$, we use the Vector Bernstein inequality on $\bar{\mathbf{y}}_i$.

Theorem 5 (Vector Bernstein inequality). *Let $\mathbf{x}_i$ be a sequence of independent random vectors and set $V \geq \sum_i \mathbb{E}\|\mathbf{x}_i\|_2^2$. Then for all $t \leq V/\max\|\mathbf{x}_i\|_2$, we have*

$$\mathbb{P}\left(\left\|\sum_i(\mathbf{x}_i - \mathbb{E}\mathbf{x}_i)\right\|_2 \geq \sqrt{V} + t\right) \leq e^{-t^2/4V}.$$

In order to apply this inequality, we need to compute $\max\|\bar{\mathbf{y}}\|_2$, $\mathbb{E}\bar{\mathbf{y}}$, and $\mathbb{E}\|\bar{\mathbf{y}}\|_2$, where $\bar{\mathbf{y}} = \mathbf{y}1_E$.

First, we compute $\max\|\bar{\mathbf{y}}\|_2$. On the event E, $|z_1| \leq \sqrt{2\beta\log n}$ and $\|\mathbf{z}\|_2 \leq \sqrt{3n}$. If n is large enough that $2\beta\log n \geq 9$, then $|\xi| \leq 2\beta\log n$. Thus,

$$\|\bar{\mathbf{y}}\|_2 \leq \sqrt{24n}(\beta\log n)^{3/2}$$

for sufficiently large n.

Second, we find an upper bound for $\mathbb{E}\bar{\mathbf{y}}$. Note that $\mathbb{E}y_1 = 0$ because

$$\mathbb{E}[z_1^4] = 3,$$
$$\mathbb{E}[z_1^2\|\mathbf{z}\|_2^2] = n + 2.$$

By symmetry, every entry of $\bar{\mathbf{y}}$ has zero mean except the first. Hence,

$$\|\mathbb{E}\bar{\mathbf{y}}\|_2 = |\mathbb{E}\bar{y}_1| = |\mathbb{E}(y_1 - y_1 1_{E^c})| = |\mathbb{E}y_1 1_{E^c}| \leq \sqrt{\mathbb{P}(E^c)}\sqrt{\mathbb{E}y_1^2} = \sqrt{\pi(\beta)}\sqrt{\mathbb{E}y_1^2}.$$

Computing,

$$y_1^2 = (\xi z_1^2)^2 = z_1^8 - \frac{6}{n+2}z_1^6\|z\|_2^2 + \frac{9}{(n+2)^2}z_1^4\|z\|_2^4,$$

we find

$$\mathbb{E}y_1^2 \leq 44,$$

where we have used

$$\mathbb{E}[z_1^8] \quad = 105, \tag{17}$$
$$\mathbb{E}[z_1^6\|\mathbf{z}\|_2^2] = 15n + 90, \tag{18}$$
$$\mathbb{E}[z_1^4\|\mathbf{z}\|_2^4] = 3n^2 + 30n + 72. \tag{19}$$

Thus,

$$\|\mathbb{E}\bar{\mathbf{y}}\|_2 \le \sqrt{44(n^{-\beta} + e^{-n/3})}. \tag{20}$$

Third, we find an upper bound for $\mathbb{E}\|\bar{\mathbf{y}}\|_2^2$. Because $\|\bar{\mathbf{y}}\|_2^2 \le \|\mathbf{y}\|_2^2$, we write out

$$\|\mathbf{y}\|_2^2 = \xi^2 z_1^2 \|\mathbf{z}\|_2^2 = z_1^6 \|\mathbf{z}\|_2^2 - \frac{6}{n+2} z_1^4 \|\mathbf{z}\|_2^4 + \frac{9}{(n+2)^2} z_1^2 \|\mathbf{z}\|_2^6.$$

Hence,

$$\mathbb{E}[\|\mathbf{y}\|_2^2] = (15n + 90) - \frac{6}{n+2}(3n^2 + 30n + 72) + \frac{9}{(n+2)^2}(n+2)(n+4)(n+6) \tag{21}$$

$$\le 8n + 16, \tag{22}$$

where we have used (18), (19), and

$$\mathbb{E}[z_1^2 \|\mathbf{z}\|_2^6] = (n+2)(n+4)(n+6). \tag{23}$$

Applying the vector Bernstein inequality with $V = m(8n + 16)$, we have that for all $t \le (8n + 16)/[\sqrt{24n}(\beta \log n)^{3/2}]$,

$$\mathbb{P}\left(\frac{1}{m}\left\|\sum_i \bar{\mathbf{y}}_i - \mathbb{E}\bar{\mathbf{y}}_i\right\|_2 \ge \sqrt{\frac{8n+16}{m}} + t\right) \le \exp\left(-\frac{mt^2}{4(8n+16)}\right).$$

Using the triangle inequality and (20), we get

$$\mathbb{P}\left(\frac{1}{m}\left\|\sum_i \bar{\mathbf{y}}_i\right\|_2 \ge \sqrt{44(n^{-\beta} + e^{-n/3})} + \sqrt{\frac{8n+16}{m}} + t\right) \le \exp\left(-\frac{mt^2}{4(8n+16)}\right).$$

Lemma 2 follows by choosing t, β, and $m \ge cn$ where n and c are large enough that

$$\sqrt{44(n^{-\beta} + e^{-n/3})} + \sqrt{\frac{8n+16}{m}} + t \le \frac{1}{4}.$$

3.4 Proof of Lemma 3: $\bar{\mathbf{Y}}$ on $T^\perp$

We prove Lemma 3 in a way that parallels the corresponding proof in [7]. We write

$$\bar{\mathbf{Y}}_{T^\perp} - 2\mathbf{I}_{T^\perp} = \frac{1}{m}\sum_i (\mathbf{W}_i 1_{E_i} - 2\mathbf{I}_{T^\perp} 1_{E_i^c}),$$

where $\mathbf{W}_i$ are independent samples of

$$\mathbf{W} = \left[\frac{3}{n+2}\|\mathbf{z}\|_2^2 - z_1^2\right]\mathcal{P}_{T^\perp}(\mathbf{z}\mathbf{z}^*) - 2\mathbf{I}_{T^\perp}. \tag{24}$$

We decompose $\mathbf{W}$ into the three terms

$$\mathbf{W} = -\left[z_1^2 - 1\right]\mathcal{P}_{T^\perp}(\mathbf{z}\mathbf{z}^*) + 3\left[\frac{1}{n+2}\|\mathbf{z}\|_2^2 - 1\right]\mathcal{P}_{T^\perp}(\mathbf{z}\mathbf{z}^*) + 2(\mathcal{P}_{T^\perp}\mathbf{z}\mathbf{z}^* - \mathbf{I}_{T^\perp}) \tag{25}$$

$$:= \mathbf{W}^{(0)} + \mathbf{W}^{(1)} + \mathbf{W}^{(2)}. \tag{26}$$

Letting $\bar{\mathbf{W}}_i^{(k)} = \mathbf{W}_i^{(k)} 1_{E_i}$, it suffices to show that with high probability

$$\frac{1}{m}\left\|\sum_i 2\mathbf{I}_{T^\perp} 1_{E_i^c}\right\| \le \frac{1}{4} \text{ and, } \frac{1}{m}\left\|\sum_i \bar{\mathbf{W}}_i^{(k)}\right\| \le \frac{1}{4} \text{ for } k = 0, 1, 2. \tag{27}$$

10

3.4.1 Bound on $\mathbf{I}_{T^\perp} 1_{E_i^c}$

We show that $m^{-1} \| \sum_i \mathbf{I}_{T^\perp} 1_{E_i^c} \| = m^{-1} \sum_i 1_{E_i^c}$ is small with probability at least $1 - 2e^{-\gamma m}$ for some constant $\gamma > 0$. To do this, we use the scalar Bernstein inequality.

Theorem 6 (Bernstein inequality). *Let $\{X_i\}$ be a finite sequence of independent random variables. Suppose that there exists V and c such that for all X_i and all $k \geq 3$,*

$$\sum_i \mathbb{E}|X_i|^k \leq \frac{1}{2} k! V c_0^{k-2}.$$

Then for all $t \geq 0$,

$$\mathbb{P}\left(\left| \sum_i X_i - \mathbb{E} X_i \right| \geq t \right) \leq 2 \exp\left(-\frac{t^2}{2V + 2c_0 t} \right). \tag{28}$$

Observing that $\mathbb{E}|1_{E_i^c}|^k = \mathbb{E} 1_{E_i^c} = \pi(\beta)$, we apply the Bernstein inequality with $V = \pi(\beta) m$ and $c_0 = 1/3$. Thus,

$$\mathbb{P}\left(\left| \frac{1}{m} \sum_i 1_{E_i^c} - \pi(\beta) \right| \geq t \right) \leq 2 \exp\left(-\frac{mt^2}{2\pi(\beta) + 2t/3} \right).$$

Using the triangle inequality and taking t and β such that $\pi(\beta) + t \leq 1/8$ for sufficiently large n, we get

$$\mathbb{P}\left(\left| \frac{1}{m} \sum_i 1_{E_i^c} \right| \geq \frac{1}{8} \right) \leq 2 \exp\left(-\gamma m \right)$$

for a $\gamma > 0$.

3.4.2 Bound on $\bar{\mathbf{W}}^{(0)}$

We show $m^{-1} \| \sum_i \bar{\mathbf{X}}^{(0)} \|$ is small with probability at least $1 - 2\exp(-\gamma/\log n)$. We write this norm as a supremum over all unit vector perpendicular to $\mathbf{e}_1$:

$$\left\| \sum_i \bar{\mathbf{W}}^{(0)} \right\| = \sup_{\mathbf{u} \perp \mathbf{e}_1, \|\mathbf{u}\|=1} \left| \sum_i \langle \mathbf{u}, \bar{\mathbf{W}}_i^{(0)} \mathbf{u} \rangle \right|, \tag{29}$$

To control the supremum, we follow the same reasoning as in [7]. We bound $\sum_i \langle \mathbf{u}, \bar{\mathbf{W}}_i^{(0)} \mathbf{u} \rangle$ for fixed $\mathbf{u}$ and apply a covering argument over the sphere of $\mathbf{u}$'s. We write

$$\sum_i \langle \mathbf{u}, \bar{\mathbf{X}}_i^{(0)} \mathbf{u} \rangle = \sum_i \eta_i 1_{E_i},$$

where η_i are independent samples of

$$\eta = -\left[z_1^2 - 1 \right] \langle \mathbf{z}, \mathbf{u} \rangle^2.$$

To apply the scalar Bernstein inequality, we compute $\mathbb{E}|\eta 1_E|^k$. Because $\mathbf{u} \perp \mathbf{e}_1$, z_1 and $\langle \mathbf{z}, \mathbf{u} \rangle$ are independent. Hence,

$$\mathbb{E}|\eta 1_E|^k \leq \mathbb{E}|(z_1^2 - 1) 1_E|^k \mathbb{E}|\langle \mathbf{z}, \mathbf{u} \rangle|^{2k}.$$

Bounding the first factor, we get

$$\mathbb{E}|(z_1^2 - 1) 1_E|^k = \mathbb{E}|(z_1^2 - 1)^{k-2} 1_E (z_1^2 - 1)^2| \leq (2\beta \log n)^{k-2} \mathbb{E}(z_1^2 - 1)^2 = 2(2\beta \log n)^{k-2}.$$

Observing that $\langle \mathbf{z}, \mathbf{u} \rangle$ is a chi-squared variable with one degree of freedom, we have

$$\mathbb{E}|\langle \mathbf{z}, \mathbf{u} \rangle|^{2k} = 1 \times 3 \times \ldots \times (2k-1) \leq 2^k k!$$

Applying the scalar Bernstein inequality with $V = 16m$ and $c_0 = 4\beta \log n$, we get

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i} - \mathbb{E}[\eta_i 1_{E_i}]\right| \geq t\right) \leq 2\exp\left(-\frac{mt^2}{2(16 + 4\beta t \log n)}\right).$$

Because $\mathbb{E}\eta_i = 0$, we get

$$|\mathbb{E}\eta_i 1_{E_i}| = |\mathbb{E}\eta_i 1_{E_i^c}| \leq \sqrt{\mathbb{P}(E_i^c)}\sqrt{\mathbb{E}\eta_i^2} = 2\sqrt{\pi(\beta)},$$

where we have used $\mathbb{E}(1 - z_1^2)^2 = 2$, and $\mathbb{E}|\langle \mathbf{z}, \mathbf{u} \rangle|^4 = 3$. Hence,

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i}\right| \geq t + 2\sqrt{\pi(\beta)}\right) \leq 2\exp\left(-\frac{mt^2}{2(16 + 4\beta t \log n)}\right).$$

Taking $t, \beta, m \geq c_1 n$ with n large enough so that $t + 2\sqrt{\pi(\beta)} \leq 1/8$, we have

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i}\right| \geq 1/8\right) \leq 2\exp\left(-\gamma'\frac{m}{\log n}\right),$$

for some $\gamma' > 0$. To complete the bound on (29), we use Lemma 4 in [21]:

$$\sup_{\mathbf{u}}\left|\langle \mathbf{u}, \bar{\mathbf{W}}^{(0)}\mathbf{u} \rangle\right| \leq 2\sup_{\mathbf{u} \in \mathcal{N}_{1/4}}\left|\langle \mathbf{u}, \bar{\mathbf{W}}^{(0)}\mathbf{u} \rangle\right|,$$

where $\mathcal{N}_{1/4}$ is a 1/4-net of the unit sphere of vectors $\mathbf{u} \perp \mathbf{e}_1$. As $|\mathcal{N}_{1/4}| \leq 9^n$, a union bound gives

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i}\right| \geq 1/8\right) \leq 9^n \cdot 2\exp\left(-\gamma'\frac{m}{\log n}\right).$$

Hence,

$$\mathbb{P}\left(\frac{1}{m}\left\|\sum_i \bar{\mathbf{W}}^{(0)}\right\| \geq \frac{1}{4}\right) \leq 2\exp\left(-\gamma m/\log n\right)$$

for some $\gamma > 0$.

3.4.3 Bounds on $\bar{\mathbf{W}}^{(1)}$ and $\bar{\mathbf{W}}^{(2)}$

The bound for the $\|\sum_i \bar{\mathbf{W}}^{(1)}\|$ term is similar. We write

$$\sum_i \langle \mathbf{u}, \bar{\mathbf{W}}_i^{(1)}\mathbf{u} \rangle = \sum_i \eta_i 1_{E_i},$$

where η_i are independent samples of

$$\eta = 3\left[\frac{\|\mathbf{z}\|_2^2}{n+2} - 1\right]\langle \mathbf{z}, \mathbf{u} \rangle^2.$$

We can bound $\mathbb{E}|\eta_i 1_E|^k \leq 12^k k!$ because $\|\mathbf{z}\|_2^2 \leq 3n$ on E. Applying the scalar Bernstein inequality with $c_0 = 12$ and $V = 288m$ gives

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i} - \mathbb{E}[\eta_i 1_{E_i}]\right| \geq t\right) \leq 2\exp\left(-\frac{mt^2}{2(288 + 12t)}\right).$$

The rest of the bound is similar to that of $\|\sum_i \bar{\mathbf{X}}^{(0)}\|$ above.

Finally, we also bound $\|\sum_i \bar{\mathbf{W}}^{(2)}\|$ similarly. We write

$$\sum_i \langle \mathbf{u}, \bar{\mathbf{W}}_i^{(2)} \mathbf{u}\rangle = \sum_i \eta_i 1_{E_i},$$

where η_i are independent samples of

$$\eta = 2\langle \mathbf{z}, \mathbf{u}\rangle^2 - 2.$$

Observing that

$$\mathbb{E}|\eta_i 1_E|^k \leq 4^k k!,$$

we apply the scalar Bernstein inequality with $c_0 = 4$ and $V = 32m$, giving

$$\mathbb{P}\left(\frac{1}{m}\left|\sum_i \eta_i 1_{E_i} - \mathbb{E}[\eta_i 1_{E_i}]\right| \geq t\right) \leq 2\exp\left(-\frac{mt^2}{2(32 + 4t)}\right).$$

The rest of the bound is as above.

4 Stability

We now prove Theorem 3, establishing the stability of the matrix recovery problem (4). We also prove Corollary 4, establishing the stability of the vector recovery problem (3). As in the exact case, the proof of Theorem 3 hinges on the ℓ_1-isometry properties (5)–(6) and the existence of an inexact dual certificate satisfying (7). For stability, we use the additional property that $\mathbf{Y} = \mathcal{A}^* \lambda$ for a λ controlled in ℓ_2. It suffices to establish an analogue of Lemma 1 along with a bound on $\|\lambda\|_2$.

Lemma 4. *Suppose that $\mathcal{A}$ satisfies (5) – (6) and there exists $\mathbf{Y} = \mathcal{A}^* \lambda$ satisfying (7) and $\|\lambda\|_1 \leq 5$. Then,*
$$\mathbf{X} \succeq 0 \text{ and } \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \varepsilon \|\mathbf{X}_0\|_2 \implies \|\mathbf{X} - \mathbf{X}_0\|_2 \leq C\varepsilon \|\mathbf{X}_0\|_2,$$
for some $C > 0$.

Proof of Lemma 4. As before, we take $\mathbf{x}_0 = \mathbf{e}_1$ and $\mathbf{X}_0 = \mathbf{e}_1 \mathbf{e}_1^*$ without loss of generality. Consider any $\mathbf{X} \succeq 0$ such that $\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \varepsilon$, and let $\mathbf{H} = \mathbf{X} - \mathbf{X}_0$. Whereas $\mathcal{A}(\mathbf{H}) = 0$ in the noiseless case, it is now of order ε because

$$\|\mathcal{A}(\mathbf{H})\|_2 \leq \|\mathcal{A}(\mathbf{X} - \mathbf{b})\|_2 + \|\mathcal{A}(\mathbf{X}_0 - \mathbf{b})\|_2 \leq 2\varepsilon. \tag{30}$$

Similarly, $|\langle \mathbf{H}, \mathbf{Y}\rangle|$ is also of order ε because

$$|\langle \mathbf{H}, \mathbf{Y}\rangle| = |\langle \mathcal{A}(\mathbf{H}), \lambda\rangle| \leq \|\mathcal{A}(\mathbf{H})\|_\infty \|\lambda\|_1 \leq \|\mathcal{A}(\mathbf{H})\|_2 \|\lambda\|_1 \leq 10\varepsilon.$$

13

Analogous to the proof of Lemma 1, we use (7) to compute that

$$10\varepsilon \geq \langle \mathbf{H}, \mathbf{Y} \rangle \geq \|\mathbf{H}_{T^\perp}\|_1 - \frac{1}{2}\|\mathbf{H}_T\|. \tag{31}$$

Using the ℓ_1-isometry properties (5) – (6), we have

$$\begin{aligned}
0.94(1-\delta)\|\mathbf{H}_T\| \leq m^{-1}\|\mathcal{A}(\mathbf{H}_T)\|_1 &\leq m^{-1}\|\mathcal{A}(\mathbf{H})\|_1 + m^{-1}\|\mathcal{A}(\mathbf{H}_{T^\perp})\|_1 \\
&\leq m^{-1/2}\|\mathcal{A}(\mathbf{H})\|_2 + (1+\delta)\|\mathbf{H}_{T^\perp}\|_1 \\
&\leq 2\varepsilon m^{-1/2} + (1+\delta)\|\mathbf{H}_{T^\perp}\|_1. \tag{32}
\end{aligned}$$

Thus (31) becomes

$$\left(10 + \frac{m^{-1/2}}{0.94(1-\delta)}\right)\varepsilon \geq \left(1 - \frac{1+\delta}{2 \cdot 0.94(1-\delta)}\right)\|\mathbf{H}_{T^\perp}\|_1, \tag{33}$$

which, along with (32), implies

$$\|\mathbf{H}_{T^\perp}\|_1 \leq C_0\varepsilon \text{ and } \|\mathbf{H}_T\| \leq C_1\varepsilon \tag{34}$$

for some $C_0, C_1 > 0$. Recalling that $\mathbf{H}_T$ has rank at most 2,

$$\|\mathbf{H}\|_2 \leq \|\mathbf{H}_T\|_2 + \|\mathbf{H}_{T^\perp}\|_2 \leq \sqrt{2}\|\mathbf{H}_T\| + \|\mathbf{H}_{T^\perp}\|_1 \leq (\sqrt{2}C_1 + C_0)\varepsilon \leq C\varepsilon.$$

$\square$

4.1 Dual Certificate Property

It remains to show $\|\lambda\|_1 \leq 5$ for $\bar{\mathbf{Y}} = \mathcal{A}^*\lambda$. From (15), we identify $\lambda = m^{-1}(\mathbf{1}_E \circ \mathcal{A}\mathcal{S}^{-1}2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*))$. Computing,

$$\begin{aligned}
\|\lambda\|_1 &= m^{-1}\|\mathbf{1}_E \circ \mathcal{A}\mathcal{S}^{-1}2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*)\|_1 \\
&\leq m^{-1}\|\mathcal{A}\mathcal{S}^{-1}2(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^*)\|_1 \\
&\leq m^{-1}\left\|\mathcal{A}\left(\frac{3}{n+2}\mathbf{I}\right) - \mathcal{A}\left(\mathbf{e}_1\mathbf{e}_1^*\right)\right\|_1 \tag{35} \\
&\leq (1+\delta)\left(\left\|\frac{3}{n+2}\mathbf{I}\right\|_1 + \|\mathbf{e}_1\mathbf{e}_1^*\|_1\right) \tag{36} \\
&\leq 4(1+\delta),
\end{aligned}$$

where (35) follows from (11), and (36) follows from the triangle inequality and the ℓ_1-isometry property (5). Hence $\|\lambda\|_1 \leq 5$.

4.2 Proof of Corollary 4

Now we prove Corollary 4, showing that stability of the lifted problem (4) implies stability of the unlifted problem (3). As before, we take $\mathbf{x}_0 = \mathbf{e}_1$ without loss of generality. Hence $\|\mathbf{X}_0\|_2 = 1$. Lemma 4 establishes that $\|\mathbf{X} - \mathbf{X}_0\| \leq C_0\varepsilon$. Recall that $\mathbf{X}_0 = \mathbf{x}_0\mathbf{x}_0^*$. Decompose $X = \sum_j \lambda_j \mathbf{v}_j \mathbf{v}_j^t$ with unit-normalized eigenvectors $\mathbf{v}_j$ sorted by decreasing eigenvalue. By Weyl's perturbation theorem,

$$\max\{|1 - \lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\} \leq C_0\varepsilon. \tag{37}$$

14

Writing

$$\mathbf{X}_0 - \mathbf{v}_1\mathbf{v}_1^* = (\mathbf{X}_0 - \mathbf{X}) + \left((\lambda_1 - 1)\mathbf{v}_1\mathbf{v}_1^* + \sum_{j=2}^m \lambda_j\mathbf{v}_j\mathbf{v}_j^* \right), \tag{38}$$

we use the triangle inequality to form the spectral bound

$$\|\mathbf{X}_0 - \mathbf{v}_1\mathbf{v}_1^*\| \le 2C_0\varepsilon.$$

Noting that

$$1 - |\langle \mathbf{x}_0, \mathbf{v}\rangle|^2 = \frac{1}{2}\|\mathbf{X}_0 - \mathbf{v}_1\mathbf{v}_1^*\|_2^2 \le \|\mathbf{X}_0 - \mathbf{v}_1\mathbf{v}_1^*\|^2 \le 4C_0^2\varepsilon^2,$$

we conclude

$$\|\mathbf{x}_0 - \mathbf{v}\|_2^2 = 2 - 2\langle \mathbf{x}_0, \mathbf{v}\rangle \le 8C_0^2\varepsilon^2.$$

5 Complex Case

The proof of Theorems 1 and 3 are analogous to the complex-valued cases. There are a few minor differences, as outlined and proved in [7]. The sensing vectors are assumed to be of the form $\Re\mathbf{z}_i \sim \mathcal{N}(0, \mathbf{I})$ and $\Im\mathbf{z}_i \sim \mathcal{N}(0, \mathbf{I})$. The ℓ_1-isometry conditions for complex $\mathcal{A}$ have weaker constants. Lemma 1 becomes

Lemma 5. *Suppose that $\mathcal{A}$ satisfies*

$$m^{-1}\|\mathcal{A}(\mathbf{X})\|_1 \le (1 + \delta)\|\mathbf{X}\|_1 \qquad \text{for all } \mathbf{X} \succeq 0,$$
$$m^{-1}\|\mathcal{A}(\mathbf{X})\|_1 \ge 0.828(1 - \delta)\|\mathbf{X}\| \quad \text{for all } \mathbf{X} \in T,$$

for some $\delta \le 3/13$. Suppose that there exists $\bar{\mathbf{Y}} \in \mathcal{R}(\mathcal{A}^*)$ satisfying

$$\|\bar{\mathbf{Y}}_T\|_1 \le 1/2 \quad \text{and} \quad \bar{\mathbf{Y}}_{T^\perp} \succeq \mathbf{I}_{T^\perp}.$$

Then, $\mathbf{X}_0$ is the unique solution to (2).

The proof of this lemma is identical to the real-valued case. The conditions of the lemma are satisfied with high probability, as before.

The construction of the inexact dual certificate is slightly different because $\mathcal{S}(\mathbf{X}) = \mathbf{X} + \mathrm{Tr}(\mathbf{X})\mathbf{I}$ and $\mathcal{S}^{-1}(\mathbf{X}) = \mathbf{X} - \frac{1}{n+1}\mathrm{Tr}(\mathbf{X})\mathbf{I}$. As a result

$$\mathbf{Y}_i = \left[\frac{4}{n+1}\|\mathbf{z}_i\|_2^2 - 2|z_{i,1}|^2 \right]\mathbf{z}_i\mathbf{z}_i^*.$$

The remaining modifications are identical to those in [7], and we refer interested readers there for details.

6 Numerical Simulations

In this section, we show that the optimizationless perspective allows for additional numerical algorithms that are unavailable for PhaseLift directly. These methods give rise to simpler algorithms with less or no parameter tuning. We demonstrate successful recovery under Douglas-Rachford and Nesterov algorithms, and we empirically show that the convergence of these algorithms is linear.

6.1   Optimization Framework

From the perspective of nonsmooth optimization, PhaseLift and the optimizationless feasibility problem can be viewed as a two-term minimization problem

$$\min_{\mathbf{X}} F(\mathbf{X}) + G(\mathbf{X}). \tag{39}$$

See, for example, the introduction to [18]. Numerical methods based on this splitting include Forward-Backward, ISTA, FISTA, and Douglas-Rachford [18, 9, 4, 10]. If F is smooth, it enables a forward step based on a gradient descent. Nonsmooth terms admit backward steps involving proximal operators. We recall that the proximal operator for a function G is given by

$$\mathrm{prox}_G(\mathbf{X}) = \mathrm{argmin}_{\mathbf{Y}} \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|^2 + G(\mathbf{Y}), \tag{40}$$

and we note that the proximal operator for a convex indicator function is the projector onto the indicated set.

PhaseLift can be put in this two-term form by softly enforcing the data fit. That gives the minimization problem

$$\min_{\mathbf{X}} \underbrace{\frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|^2 + \lambda\,\mathrm{tr}(\mathbf{X})}_{F} + \underbrace{\iota_{\mathbf{X}\succeq 0}(\mathbf{X})}_{G} \tag{41}$$

where $\iota_{\mathbf{X}\succeq 0}$ is the indicator function that is zero on the positive semidefinite cone and infinite otherwise, and where λ is small and positive. If $\lambda = 0$, (41) reduces to the optimizationless feasibility problem. The smoothness of F enables methods that are forward on F and backward on G. As a representative of this class of methods, we will consider a Nesterov iteration for our simulations below.

The optimizationless view suggests the splitting

$$\min_{\mathbf{X}} \underbrace{\iota_{\mathcal{A}(\mathbf{X})=\mathbf{b}}(\mathbf{X})}_{F} + \underbrace{\iota_{\mathbf{X}\succeq 0}(\mathbf{X})}_{G}. \tag{42}$$

where the data fit term is enforced in a hard manner by the indicator function $\iota_{\mathcal{A}(\mathbf{X})=\mathbf{b}}$. Because of the lack of smoothness, we can only use the proximal operators for F and G. These operators are projectors on to the affine space $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ and $X \succeq 0$, which we denote by $\mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}$ and $\mathcal{P}_{\mathrm{psd}}$, respectively.

The simplest method for (42) is Projection onto Convex Sets (POCS), which is given by the backward-backward iteration $\mathbf{X}_{n+1} = \mathcal{P}_{\mathrm{psd}}\mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}\mathbf{X}_n$. Douglas-Rachford iteration often gives superior performance than POCS, so we consider it as a representative of this class of backward-backward methods.

A strength of the optimizationless perspective is that it does not require as much parameter tuning as PhaseLift. For example, formulation (41) requires a numerical choice for λ. Nonzero λ will generally change the minimizer. It is possible to consider a sequence of problems with varying λ, or perhaps to create a schedule of λ within a problem, but these considerations are unnecessary because the optimizationless perspective says we can take $\lambda = 0$. In particular, formulation (42) has the further strength of requiring no parameters at all.

We note that PhaseLift could alternatively give rise to the two-term splitting

$$\min \underbrace{\mathrm{tr}(\mathbf{X})}_{F} + \underbrace{\iota_{\mathbf{X}\succeq 0}(\mathbf{X}) + \iota_{\mathcal{A}(\mathbf{X})=\mathbf{b}}(\mathbf{X})}_{G}, \tag{43}$$

16

where the data fit term is enforced in a hard manner. An iterative approach with this splitting would have an inner loop which approximates the proximal operator of G. This inner iteration is equivalent to solving the optimizationless problem.

6.2 Numerical Results

First, we present a Douglas-Rachford [9] approach for finding $\mathbf{X} \in \{\mathbf{X} \succeq 0\} \cap \{\mathcal{A}(\mathbf{X}) \approx \mathbf{b}\}$ by the splitting (42). It is given by the iteration

$$\mathbf{X}_0 = \mathbf{Y}_0 = 0 \tag{44}$$

$$\mathbf{Y}_n = \mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}(2\mathbf{X}_{n-1} - \mathbf{Y}_{n-1}) - \mathbf{X}_{n-1} + \mathbf{Y}_{n-1} \tag{45}$$

$$\mathbf{X}_n = \mathcal{P}_{\mathrm{psd}}(\mathbf{Y}_n) \tag{46}$$

where $\mathcal{P}_{\mathrm{psd}}$ is the projector onto the positive semi-definite cone of matrices, and $\mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}$ is the projector onto the affine space of solutions to $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. In the classically underdetermined case, $m < \frac{(n+1)n}{2}$, we can write

$$\mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}\mathbf{X} = \mathbf{X} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(\mathbf{X}) + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathbf{b}.$$

In the case that $m \geq \frac{(n+1)n}{2}$, we interpret $\mathcal{P}_{\mathcal{A}(\mathbf{X})=\mathbf{b}}$ as the least squares solution to $\mathcal{A}(\mathbf{X}) = b$.

Second, we present a Nesterov gradient-based method for solving the problem (41). Letting $g(\mathbf{X}) = \frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|^2 + \lambda \operatorname{tr}(\mathbf{X})$, we consider the following Nesterov iteration [5] with constant step size α:

$$\mathbf{X}_0 = \mathbf{Y}_0 = 0 \tag{47}$$

$$\mathbf{X}_n = \mathcal{P}_{\mathrm{psd}}(\mathbf{Y}_{n-1} - \alpha\nabla g(\mathbf{Y}_n - 1)) \tag{48}$$

$$\theta_n = 2\left(1 + \sqrt{1 + 4/\theta_{n-1}^2}\right)^{-1} \tag{49}$$

$$\beta_n = \theta_n(\theta_{n-1}^{-1} - 1) \tag{50}$$

$$\mathbf{Y}_n = \mathbf{X}_n + \beta_n(\mathbf{X}_n - \mathbf{X}_{n-1}) \tag{51}$$

For our simulations, we consider $\mathbf{x}_0 \in \mathbb{R}^n$ sampled uniformly at random from the unit sphere. We take independent, real-valued $\mathbf{z}_i \sim \mathcal{N}(0, \mathbf{I})$, and let the measurements $\mathbf{b}$ be subject to additive Gaussian noise corresponding to $\varepsilon = 1/10$. We let n vary from 5 to 50 and let m vary from 10 to 250. We define the recovery error as $\|\mathbf{X} - \mathbf{X}_0\|_2/\|\mathbf{X}_0\|_2$.

Figure 2 shows the average recovery error for the optimizationless problem under the Douglas-Rachford method and the Nesterov method over a range of values of n and m. For the Nesterov method, we consider the optimizationless case of $\lambda = 0$, and we let the step size parameter $\alpha = 2 \cdot 10^{-4}$. Each pair of values was independently sampled 10 times, and both methods were run for 1000 iterations. The plot shows that the number of measurements needed for recovery is approximately linear in n, significantly lower than the amount for which there are an equal number of measurements as unknowns. The artifacts around the curve $m = \frac{n(n+1)}{2}$ appear because the problem is critically determined, and the only solution to the noisy $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ is not necessarily positive in that case.

Figure 3 shows recovery error versus iteration number under the Douglas-Rachford method, the Nesterov method for $\lambda = 0$ and the Nesterov method for $\lambda = 10^{-5}$. For the Nesterov methods, we let the step size parameter be $\alpha = 10^{-4}$. For noisy data, convergence is initially linear until it tapers off around the noise level. For noiseless data, convergence for feasibility problem is
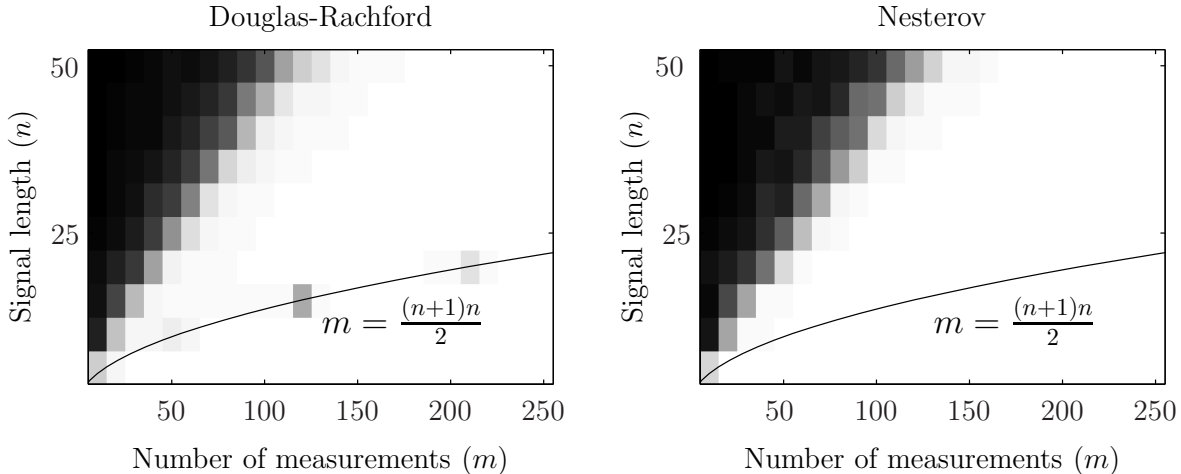
17

Figure 2: Recovery error for the Douglas-Rachford (DR) and Nesterov methods for the noisy optimizationless matrix recovery problem (4) as a function of n and m. In these plots, $\varepsilon = 10^{-1}$. For the Nesterov method, $\lambda = 0$. Black represents an average recovery error of 100%. White represents zero average recovery error. Each block corresponds to the average of 10 independent samples. The solid curve depicts when there are the same number of measurements as degrees of freedom. The number of measurements required for recovery appears to be roughly linear, as opposed to quadratic, in n. The DR algorithm has large recovery errors near the curve where the number of measurements equals the number of degrees of freedom.

linear under both the Douglas-Rachford and Nesterov methods. The Nesterov implementation of PhaseLift shows initial linear convergence until it tapers off. Because any nonzero λ allows for some data misfit in exchange for a smaller trace, the computed minimum is not $\mathbf{X}_0$ and the procedure converges to some nearby matrix. The convergence rates of the Nesterov method could probably be improved by tuning the step-sizes in a more complicated way. Nonetheless, we observe that the Douglas-Rachford method exhibits a favorable convergence rate while requiring no parameter tuning.

We would like to remark that work subsequent to this paper shows that the number of measurements needed by the optimizationless feasibility problem is about the same as the number needed by PhaseLift [22]. That is, the phase transition in Figure 2 occurs in about the same place for both problems.

References

[1] R. Balan, B. Bodmann, P.G. Casazza, D. Edidin. Painless reconstruction from magnitudes of frame vectors, *J. Fourier Anal. Appl.*, 15(4), 488-501, 2009.

[2] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Appl. Comput. Harmon. Anal.*, 20(3), 345-356, 2006

[3] A. S. Bandeira, A. Singer, D. A. Spielman, A Cheeger Inequality for the Graph Connection Laplacian. To appear in *SIAM Journal on Matrix Analysis and Applications.*

[4] A. Beck, M. Teboulle, A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems *SIAM J. Imaging Sci.* 2(1):183-202, 2009.
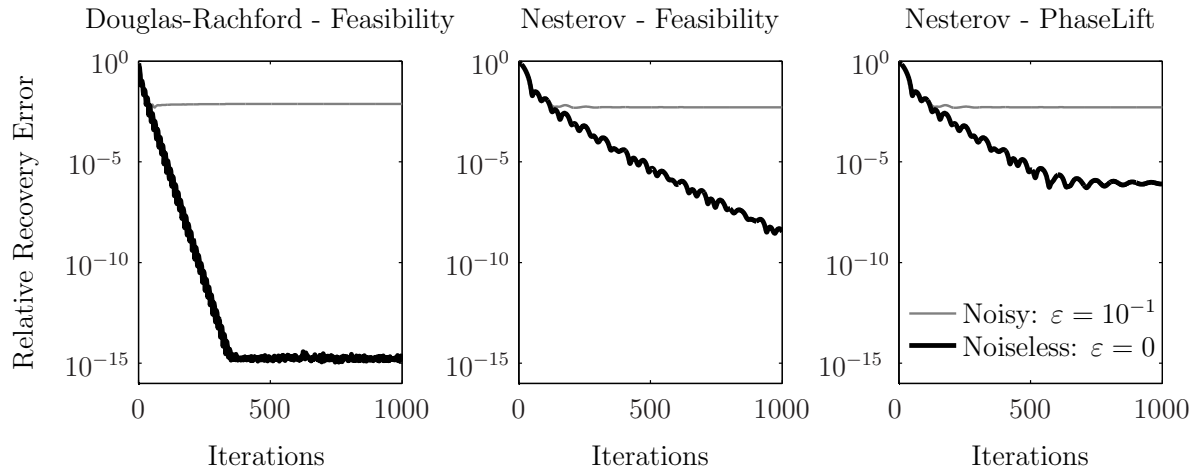
Figure 3: The relative error versus iteration number for the noiseless and noisy matrix recovery problems, (2) and (4), under the Douglas-Rachford and Nesterov methods. The left panel corresponds to the iteration (44)–(46). The middle panel corresponds to the iteration (47)–(51) in the optimizationless case, where $\lambda = 0$. The right panel corresponds to the iteration (47)–(51) for $\lambda = 10^{-5}$. As expected, convergence is linear until it saturates due to noise. In the Nesterov implementation of PhaseLift, convergence is linear until it saturates because the nonzero λ makes the solution to the problem different than $\mathbf{X}_0$.

[5] E. J. Candes, Y. C. Eldar, T. Strohmer, V. Voroninski. Phase Retrieval via Matrix Completion. *SIAM J. on Imaging Sciences* 6(1), 199–225, 2011.

[6] E. J. Candes, X. Li. Solving Quadratic Equations via PhaseLift when There Are About As Many Equations As Unknowns. To appear in *Found. Comput. Math.*, 2012.

[7] E. J. Candes, T. Strohmer, V. Voroninski. PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. *Commun. Pure Appl. Math.* 66(8), 1241–1274, 2013.

[8] A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. *Inverse Probl.*, 27(1), 015005, 2011.

[9] P. Combettes, J. Pesquet. Proximal Splitting Methods in Signal Processing. In: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering.* New York: Springer-Verlag, 2010.

[10] J. Douglas, H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Trans. AMS*, 82(2), 421-439, 1956.

[11] J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15), 27582769, 1982.

[12] R. Gerchberg, W. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35, 237246, 1972.

[13] M.X. Goemans, D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semi-definite programming. *J. ACM*, 42, 11151145, 1995

[14] D. Griffin, J. Lim. Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2), 236243, 1984.

[15] R.W. Harrison. Phase problem in crystallography. *J. Opt. Soc. Am. A*, 10(5), 10451055, 1993.

[16] J. Miao, T. Ishikawa, Q. Shen, T. Earnest. Extending X-Ray crystallography to allow the imaging of non- crystalline materials, cells and single protein complexes. *Annu. Rev. Phys. Chem.*, 59, 387410, 2008.

[17] R.P. Millane. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am. A*, 7:394411, 1990.

[18] H. Raguet, J. M. Fadili, G. Peyre. A Generalized Forward-Backward Splitting. *SIAM J. Imaging Sci*, 6(3), 1199-1226, 2013.

[19] B. Recht, M. Fazel. P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization *SIAM Rev.*, 52(3), 471501, 2010.

[20] A. Singer. Angular synchronization by eigenvectors and semidefinite programming *Appl. Comput. Harmon. Anal.* 30(1), 20–36, 2011.

[21] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Camb. Univ Press 2010.

[22] I. Waldspurger, A. d'Aspremont, S. Mallat, Phase recovery, Maxcut, and complex semi-definite programming *arXiv Preprint* 1206.0102, 2012

[23] A. Walther. The question of phase retrieval in optics. *Opt. Acta, 10:4149, 1963.*