Solving Complex Quadratic Systems with Full-Rank Random Matrices

Shuai Huang∗, Sidharth Gupta∗, Ivan Dokmanić

February 18, 2019

Abstract

We tackle the problem of recovering a complex signal $x \in \mathbb{C}^n$ from quadratic measurements of the form $y_i = x^* A_i x$, where $\{A_i\}_{i=1}^m$ is a set of full-rank, complex random matrices with rotation invariant entries. This non-convex problem is related to the well understood phase retrieval problem where $A_i$ is a rank-1 positive semidefinite matrix. Here we study the general full-rank case which models a number of key applications such as molecular geometry recovery from distance distributions and compound measurements in phaseless diffractive imaging. Most prior work tackles specific measurement models and either addresses the rank-1 case or focuses on real measurements. The several papers that address the full-rank complex case adopt the computationally-demanding semidefinite relaxation approach. In this paper we prove that the general class of rotation invariant measurement models can be efficiently solved with high probability via the standard framework comprising a spectral initialization followed by iterative gradient descent updates. Numerical experiments on simulated data generated from the rotation invariant exponential family corroborate our theoretical analysis and provide insights in designing suitable measurement matrices.

1 Introduction

Systems of quadratic equations model many problems in applied science, including phase retrieval \cite{1,2,3}, unlabeled distance geometry problem (uDGP) \cite{4,5}, and blind channel estimation \cite{6,7}. Phase retrieval, in particular, has motivated considerable recent research on quadratic equations. In phase retrieval, the measurements are $y_i = |a_i^* x|^2 = x^* a_i a_i^* x$, where the measurement matrices $a_i a_i^*$ are rank-1 positive semidefinite matrices. We are interested in a different measurement model with high-rank measurement matrices. Such a measurement model arises in a number of applications including combinatorial optimization problems such as uDGP \cite{5}, the turnpike problem \cite{8}, and in phaseless diffractive imaging where, either because of experimental or algorithmic design, we work with linear combinations of pixel values on the detector: $y_i = \sum_r w_r |a_r^* x|^2$. These problems can be modeled as systems of quadratic equations where the measurement matrices are not necessarily rank-1 or real. Different methods are needed to study such high-rank complex quadratic measurement model.

Recovering the signal $x$ from its complex quadratic measurements is a nonconvex problem, and solving for the the globally optimal solution is in general intractable. Recent works on nonconvex
quadratic problems such as phase retrieval \[9,10\], phase synchronization \[11,12\], and low-rank matrix recovery \[13\] have shown that a globally optimal solution \(z\) can be recovered from sufficient measurements with high probability when iid Gaussian measurement vectors or matrices are used. This motivates the question of what other types of measurement matrices also enjoy such favorable properties. In this paper we show that the class of measurement matrices with rotation invariant entries is one such type, and propose to recover a globally optimal solution via the standard framework comprising a spectral initialization followed by iterative gradient descent updates. We can prove that when the number of measurements exceeds the signal’s length by some constant factor, the complex signal can be recovered up to a certain phase difference with high probability.

1.1 Rotation invariant measurement model

We address the problem of recovering a complex signal \(x \in \mathbb{C}^n\) from its complex quadratic measurements \(y_i \in \mathbb{C}\) of the following form:

\[ y_i = x^* A_i x, \quad i = 1, \ldots, m, \]

where \(A_i \in \mathbb{C}^{n \times n}\) is the \(i\)-th complex measurement matrix. Let \(r_i \in \mathbb{R}^{2n^2}\) denote the real and imaginary coefficients of the entries of \(A_i\):

\[ r_i^T = \begin{bmatrix} \text{Vec} (\text{Re}(A_i))^T & \text{Vec} (\text{Im}(A_i))^T \end{bmatrix}. \]

The \(m\) coefficients vectors \(\{r_i|i=[m]\}\) are independently and identically distributed and follow a multivariate rotation invariant distribution \[14\]—the distribution of \(r_i\) does not change under any unitary transform. In other words, \(r_i\) has a spherically symmetric distribution and so its probability density function, \(p(r_i)\), depends only on the norm \(\|r_i\|_2\). Using \[14, Proposition 4.1.1\], we have:

\[ p(r_{i,k}) = p(-r_{i,k}) \]

\[ p(r_{i,k} \cdot r_{i,l}) = p(-r_{i,k} \cdot r_{i,l}), \quad \forall k \neq l. \]

It is easy to verify that \(E[r_{i,k}] = E[r_{i,l}] = 0\) and \(E[r_{i,k} \cdot r_{i,l}] = 0\). We then have:

\[ E[r_{i,k} \cdot r_{i,l}] = E[r_{i,k}] \cdot E[r_{i,l}] = 0. \]

Hence the entries of \(r_i\) are pairwise uncorrelated. Without loss of generality, we assume the variance \(\mu^2 = \text{Var}(r_{i,k}) = \mathbb{E}[r_{i,k}^2] = 1\), and the following variances

\[ \sigma^2 = \text{Var}(r_{i,k}^2) \]

\[ \rho^2 = \text{Var}(r_{i,k} \cdot r_{i,l}), \quad \forall k \neq l. \]

exist and are bounded. In this paper we shall refer to the model given by \[1\] as the rotation invariant measurement model.

1.2 Problem formulation

We minimize the following loss function \(f(z)\) to obtain the recovered signal \(z\):

\[ f(z) = \frac{1}{m} \sum_{i=1}^{m} |z^* A_i z - y_i|^2. \]
Suppose we are given a good initialization point \( z^{(0)} \) (finding such a point is discussed in Section 2). The solution is then updated iteratively via gradient descent:

\[
    z^{(t+1)} = z^{(t)} - \eta \nabla f(z),
\]

where \( \eta > 0 \) is some suitable step size. The gradient \( \nabla f(z) \) can be computed as follows:

\[
    \nabla f(z) = \frac{1}{m} \sum_{i=1}^{m} (z^* A_i^* z - x^* A_i^* x) A_i z + (z^* A_i z - x^* A_i x) A_i^* z.
\]

(10)

If \( x \) is a global minimum of \( f(z) \), then \( xe^{j\phi} \) is also a global minimum for all \( \phi \in [0, 2\pi) \). Consequently, it is standard to define the squared distance between the recovered solution \( z \) and the true solution \( x \) as

\[
    \text{dist}^2(z, x) = \min_{\phi \in [0, 2\pi]} \| z - xe^{j\phi} \|^2_2 = \min_{\phi \in [0, 2\pi]} \| z \|^2_2 + \| x \|^2_2 - 2 |z^* x| (e^{j\phi(z^* x)} e^{j\phi z} + e^{-j\phi(z^* x)} e^{-j\phi z})
\]

(11)

where \( z^* x = |z^* x| e^{j\phi(z^* x)} \) and \( \phi_z = -\phi(z^* x) \).

### 1.3 Prior art

Similar quadratic equation problems have been studied in other contexts. Candès et al. [10] cast the phase retrieval problem as a system of structured quadratic equations and solve it via vanilla gradient descent with a guaranteed linear convergence rate. As this is a non-convex problem, they use a suitably constructed spectral initializer \( z^{(0)} \) for the Gaussian measurement model: \( z^{(0)} \) can be proven to be close to a globally optimal solution when sufficient measurements are available. The work of Ghods et al. [15] on phase retrieval demonstrates that the signal can also be recovered with guarantees in a deterministic measurement model. Here we complement these results by showing that both the initialization and convergence proofs can be derived using Chebyshev’s inequality for the full-rank rotation invariant measurement model in [1].

Lu and Li [16] study how the spectral initialization can be generalized in the real recovery case for measurements that are independently drawn from \( y_i \sim p(y|a_i^T x) \), where \( a_i^T x \) serves as the parameter of the pdf \( p(y|a_i^T x) \). They focus on the asymptotic behavior of the initializer w.r.t. the sampling ratio \( m/n \) in the high-dimensional limit. To accelerate the phase retrieval process for large-scale problems, [17] computes a optimal step size for the gradient descent method at every iteration, and [18] proposes several coordinate descent algorithms with faster convergence speeds. The works of Wang, Xu, and Huang [19–21] solve a generalized phase retrieval problem where \( A_i \) is a Hermitian matrix. They use algebraic methods [22] to find the number of measurements needed for a successful recovery.

Solving the systems of quadratic equations is closely related to low-rank matrix recovery—it is equivalent to recovering a rank-1 positive semidefinite matrix \( X = xx^* \) with \( y_i = \langle A_i, X \rangle \). Previous works [24, 26] on low-rank matrix recovery focus on relaxing the non-convex low rank constraint to the convex minimum nuclear norm constraint, and then establishing sufficient conditions that warrant such a convex relaxation. However, recovering the relaxed \( X \) is computationally
demanding even for moderate sized problems. To address this issue, [13, 27–29] took the low-rank matrix decomposition approach $\hat{X} = UV^T$, where $U, V \in \mathbb{R}^{n \times r}$ and $\hat{X} \in \mathbb{R}^{n \times n}$ are all real matrices, and the optimization is performed directly over $U$ and $V$. In this work we aim to recover a complex signal $x$ directly up to a certain phase difference.

1.4 Paper outline

The rest of the paper is organized as follows. In Section 2 we extend the derivations from [9] to the general rotation invariant measurement model. We show that the spectral initialization concentrates around a global optimum with high probability and compute the associated concentration bounds. In Section 3 we analyze the regularity condition and derive new results for the rotation invariant measurement model. The two results are then combined to give the main theorem of this paper. In Section 3.2.1 we derive tighter results for the specific case of the complex Gaussian measurement model. In Section 4 we show the phase transition plots for different distributions in the rotation invariant exponential family (defined in Section 3.2) via simulated experiments. The detailed proofs are given in the Appendix.

2 Spectral Initialization

Spectral initialization is widely used to obtain an initialization that is close to a global optimum, $x$. Similar to [9, 10], we show that the spectral initializer $z(0)$ obtained with the rotation invariant measurement model also concentrates around $x$ with high probability and can be used to initialize the gradient descent update in (9). This follows from the observation that in the spectral initializer analysis, Gaussian concentration can be replaced by the weaker Chebyshev inequality.

2.1 Initialization for signals with known norms

We first consider the case where the norm of $x$ is known and fixed. Without loss of generality, we shall assume $\|x\|_2 = 1$. The rationale behind the spectral initialization strategy is that we can get a good estimate $S'$ of $xx^*$ using $\{y_i, A_i| i = [m]\}$ for some type of $A_i$. The spectral initializer $z(0)$ can then be obtained by performing a singular value decomposition of $S'$ and choosing a leading singular vector so that $z(0)$ is highly correlated with $x$.

Unlike in the phase retrieval problem, which uses the leading eigenvector of the Hermitian matrix $\hat{S} = \frac{1}{m} \sum_{i=1}^{m} y_i a_i a_i^*$ as the initializer, we have two possible choices here: the leading left or right singular vector $v_0$ of

$$S = \frac{1}{m} \sum_{i=1}^{m} \overline{y}_i A_i,$$

which can both be chosen as the spectral initializer $z(0) = v_0$, where $\overline{y}_i$ is the complex conjugate of $y_i$. To see why, note that the expectation of $S$ is

$$\mathbb{E}[S] = \mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^{m} \sum_{rc} x_r \overline{x}_c A_{i,rc} \cdot A_i \right] = \sum_{rc} \overline{x_r x_c} \cdot \mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^{m} \overline{A}_{i,rc} \cdot A_i \right].$$

(13)
The real and imaginary coefficients of entries in \( \mathbf{A}_i \) are rotation invariant, pairwise uncorrelated, and have unit variance \( \mu^2 = 1 \). Let \( \mathbf{A}^{(R)}_{i,rc} \) and \( \mathbf{A}^{(I)}_{i,rc} \) denote the real and imaginary parts of \( \mathbf{A}_{i,rc} \). Using (5), it is easy to check that:

\[
\mathbb{E} \left[ \mathbf{A}_{i,rc} \cdot \mathbf{A}_{i,rc} \right] = \mathbb{E} \left[ \mathbf{A}^{(R)}_{i,rc}^2 + \mathbf{A}^{(I)}_{i,rc}^2 \right] = 2
\]

\[
\mathbb{E} \left[ \mathbf{A}_{i,rc} \cdot \mathbf{A}_{i,kl} \right] = \mathbb{E} \left[ \left( \mathbf{A}^{(R)}_{i,rc} \mathbf{A}^{(R)}_{i,kl} + \mathbf{A}^{(I)}_{i,rc} \mathbf{A}^{(I)}_{i,kl} \right) + \left( \mathbf{A}^{(R)}_{i,rc} \mathbf{A}^{(I)}_{i,kl} - \mathbf{A}^{(I)}_{i,rc} \mathbf{A}^{(R)}_{i,kl} \right) \cdot j \right] = 0,
\]

where \( r \neq k \) or \( c \neq l \). Hence \( \mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^{m} \mathbf{A}_{i,rc} \mathbf{A}_i \right] \) is a matrix with the \((r,c)\)-th entry being 2, while the rest of the entries are all 0s. We then have:

\[
\mathbb{E} \left[ \mathbf{S} \right] = 2 \mathbf{x} \mathbf{x}^*.
\]

We prove that for sufficiently large \( m \), the matrix \( \mathbf{S} \) concentrates around \( \mathbb{E} \left[ \mathbf{S} \right] \) in the spectral norm: \( \| \mathbf{S} - \mathbb{E} \left[ \mathbf{S} \right] \| \leq \delta \). As a consequence the leading singular vectors of \( \mathbf{S} \) are both highly correlated with a global optimizer \( \mathbf{x} \) \[16\]. The concentration proof hinges on the rotation invariance of the measurement matrices \( \mathbf{A}_i \) \[14\]: If we define \( \mathbf{B} = \mathbf{R} \mathbf{A} \) with \( \mathbf{R} \in \mathbb{C}^{n \times n} \) being a complex unitary matrix, the real and imaginary coefficients, \( r_i \in \mathbb{R}^{2n^2} \), of entries in \( \mathbf{B} \) still follow the same multivariate distribution as those of \( \mathbf{A} \).

### 2.2 Initialization for signals with unknown norms

When the norm of the signal is unknown, we can estimate it from the quadratic measurements. Using \([16]\), we can compute:

\[
\mathbb{E} \left[ \frac{1}{2m} \sum_{i=1}^{m} \bar{y}_i y_i \right] = \mathbb{E} \left[ \frac{1}{2} \mathbf{x}^* \mathbf{S} \mathbf{x} \right] = \frac{1}{2} \mathbf{x}^* \mathbb{E} \left[ \mathbf{S} \right] \mathbf{x} = \| \mathbf{x} \|_2^4.
\]

When \( m \) is sufficiently large, we prove that \( \frac{1}{2m} \sum_{i=1}^{m} \bar{y}_i y_i \) is close to its expectation \( \| \mathbf{x} \|_2^4 \) with high probability. Based on this result, we can scale one of the leading singular vectors \( \mathbf{v}_0 \) of \( \mathbf{S} \) to get our spectral initializer:

\[
\mathbf{z}^{(0)} = \left( \frac{1}{2m} \sum_{i=1}^{m} \bar{y}_i y_i \right)^{\frac{1}{4}} \cdot \mathbf{v}_0.
\]

We have the following lemma stating that the distance between the spectral initializer \( \mathbf{z}^{(0)} \) and a global optimizer \( \mathbf{x} \) is small with high probability when \( m \) is sufficiently large.

**Lemma 1.** Under the rotation invariant measurement model given by (1), when the number of complex quadratic measurements satisfies \( m > Cn \) for some universal constant \( C \), the distance between the spectral initializer \( \mathbf{z}^{(0)} \) and a global optimizer \( \mathbf{x} \) is upper-bounded by

\[
\text{dist}^2 \left( \mathbf{z}^{(0)}, \mathbf{x} \right) \leq \frac{11}{4} \delta \| \mathbf{x} \|_2^2,
\]

with probability at least \( 1 - \frac{1}{Cn} \cdot \frac{C_1(\sigma,\rho)}{\delta^2} \), where \( \delta > 0 \) is some pre-specified constant and \( C_1(\sigma,\rho,\delta) \) is some constant depending on the variances \( \{\sigma,\rho\} \).
Given the measurement model, the signal length $n$ and the constant $C_1(\sigma, \rho)$ are fixed. To get a good initialization, we would like a small $\delta$ and ensure (19) holds with high probability. Hence $C \leq \frac{m}{n}$ needs to be large so that $\frac{1}{Cn} \cdot C_1(\sigma, \rho) \ll 1$. Experiments in Section 4.1 show that we can achieve this goal by increasing the number of measurements $m$, i.e. by raising the upper bound $\frac{m}{n}$ on $C$.

3 Convergence Analysis

For the objective function $f(z)$ introduced in (8), there is a special neighborhood $E(\epsilon)$ around every global optimizer $x$:

$$E(\epsilon) = \{ z \mid \text{dist}(z, x) \leq \epsilon \|x\|_2 \} .$$

The objective function $f(z)$ is said to satisfy the regularity condition $RC(\alpha, \beta, \epsilon)$ if the following holds for all $z \in E(\epsilon)$ [10]:

$$\text{Re} \left( \left( \nabla f(z), z - xe^{j\phi} \right) \right) \geq \frac{1}{\alpha} \text{dist}^2(z, x) + \frac{1}{\beta} \|\nabla f(z)\|^2, \quad \text{(20)}$$

where $\alpha > 0$, $\beta > 0$, $\epsilon > 0$ are some carefully chosen constants. The regularity condition $RC(\alpha, \beta, \epsilon)$ ensures that the gradient descent updates [9] with a step size $\eta \in \left( 0, \frac{2}{\beta} \right)$ converge linearly to a global optimizer $x$ when initialized within the neighborhood $E(\epsilon)$ [10, Lemma 7.10]:

$$\text{dist}^2 \left( z^{(t)}, x \right) \leq \left( 1 - \frac{2\eta}{\alpha} \right)^t \text{dist}^2 \left( z^{(0)}, x \right). \quad \text{(21)}$$

For the system of complex quadratic equations given by (1), we show that the objective function $f(z)$ satisfies the regularity condition $RC(\alpha, \beta, \epsilon)$ (20), and give a set of $(\alpha, \beta, \epsilon)$-values such that spectral initialization followed by gradient descent succeeds with high probability. We shall make use of the following lemma.

**Lemma 2.** Under the rotation invariant measurement model given by $\text{(1)}$, when the number of complex quadratic measurements satisfies $m > Cn$ for some universal constant $C$, the following two inequalities hold with probability at least $1 - \frac{\nu^2}{C^2(\sigma, \rho)}$:

$$\left| \frac{1}{m} \sum_{i=1}^{m} p^* A_i^* q A_i v - 2u^* q p^* v \right| \leq \nu \|p\|_2 \|q\|_2 \|u\|_2 \|v\|_2 \quad \text{(22)}$$

$$\left| \frac{1}{m} \sum_{i=1}^{m} \text{Re} \left( p^* A_i^* q u^* A_i v \right) - 2\text{Re} \left( u^* q p^* v \right) \right| \leq \nu \|p\|_2 \|q\|_2 \|u\|_2 \|v\|_2, \quad \text{(23)}$$

where $\nu$ is some pre-specified constant, and $C_2(\sigma, \rho)$ is some constant depending on the variances $\{\sigma, \rho\}$.

Following derivations similar to those of [13]-[16], we can see that

$$\mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^{m} p^* A_i^* q A_i \right] = 2qp^*.$$  

(24)
This allows us to interpret Lemma 2. It says that the matrix \( \frac{1}{m} \sum_{i=1}^{m} p^* A_i^* q A_i \) concentrates around its expectation with high probability in the spectral norm.

Let \( h = z e^{-j\phi_z} - x \), we can verify that:

\[
\text{Re} (\langle h, \nabla f(z) \rangle) = \text{Re}(h^* \nabla f(x)) = \text{Re} \left( \left\langle \nabla f(z), z - x e^{j\phi_z} \right\rangle \right). \tag{25}
\]

To prove that our objective function (1.2) satisfies the regularity condition, we have three steps.

1. **Finding a lower bound on** \( \text{Re} (\langle h, \nabla f(z) \rangle) \).

   We first find a lower bound on \( \text{Re} (\langle h, \nabla f(z) \rangle) \) for all \( h \in \mathbb{C}^n \) satisfying \( \|h\|_2 \leq \epsilon \|x\|_2 \). We proceed by showing that \( \mathbb{E} [\text{Re} (\langle h, \nabla f(z) \rangle)] > 0 \), and that \( \text{Re} (\langle h, \nabla f(z) \rangle) \) is close to \( \mathbb{E} [\text{Re} (\langle h, \nabla f(z) \rangle)] \). Thus a positive lower bound can be established with high probability when \( m \) is sufficiently large. We break this problem down by lower-bounding every term in the following expansion:

\[
\text{Re} (\langle h, \nabla f(z) \rangle) = \frac{1}{m} \sum_{i=1}^{m} [2 \text{Re} (h^* A_i^* h \cdot h^* A_i h) + 3 \text{Re} (h^* A_i^* h \cdot h^* A_i x) + 3 \text{Re} (h^* A_i^* x \cdot h^* A_i h) + 2 \text{Re} (h^* A_i^* x \cdot h^* A_i x) + \text{Re} (x^* A_i^* h \cdot h^* A_i x) + \text{Re} (x^* A_i^* h \cdot h^* A_i x)] . \tag{26}
\]

Using (23) and the Cauchy-Schwartz inequality, we have:

\[
\text{Re} (\langle h, \nabla f(z) \rangle) \geq 4 \left[ \left( \|h\|_2^2 + \text{Re} (h^* x) \right)^2 + \|h\|_2^2 \left( \|x\|_2^2 + \text{Re} (x^* h) \right) \right] - 2 \|h\|_2^2 \nu \left( \|h\|_2^2 + 3 \|h\|_2 \|x\|_2^2 + 2 \|x\|_2^2 \right) \geq 4 \left[ \|h\|_2^2 (\|x\|_2^2 + \text{Re} (x^* h)) \right] - 2 \|h\|_2^2 \nu \left( \|h\|_2^2 + 3 \|h\|_2 \|x\|_2^2 + 2 \|x\|_2^2 \right) \geq 4 \left[ \|h\|_2^2 \|x\|_2^2 \left( 1 - \epsilon - \frac{\nu}{2} (2 + 3\epsilon + \epsilon^2) \right) \right], \tag{27}
\]

where \( \epsilon = \frac{\|h\|_2^2}{\|x\|_2^2} \), and \( \nu \) is some pre-specified constant.

2. **Finding an upper bound on** \( \| \nabla f(z) \|_2^2 \).

   We next compute an upper bound on \( \| \nabla f(z) \|_2^2 \). This is equivalent to finding an upper bound of \( |u^* \nabla f(z)|^2 \) over all \( u \in \mathbb{C}^N \) with \( \|u\|_2 = 1 \). We have:

\[
|u^* \nabla f(z)|^2 = \left| \frac{1}{m} \sum_{i=1}^{m} \nabla f(z) \cdot \left[ h^* A_i h \cdot u^* A_i^* h + h^* A_i h \cdot u^* A_i^* x + h^* A_i x \cdot u^* A_i^* h + h^* A_i x \cdot u^* A_i^* x + x^* A_i h \cdot u^* A_i^* h + x^* A_i h \cdot u^* A_i^* x + x^* A_i^* h \cdot u^* A_i h + x^* A_i^* h \cdot u^* A_i x + h^* A_i^* h \cdot u^* A_i h + h^* A_i^* h \cdot u^* A_i x + h^* A_i^* x \cdot u^* A_i h + h^* A_i^* x \cdot u^* A_i x + x^* A_i^* h \cdot u^* A_i h + x^* A_i^* h \cdot u^* A_i x \right] \right|^2 . \tag{28}
\]

We can then use (22) to obtain an upper bound on \( |u^* \nabla f(z)| \) for all \( u \in \mathbb{C}^n \) satisfying \( \|u\|_2 = 1 \):

\[
|u^* \nabla f(z)| \leq 2(2 + \nu) \|h\|_2 \|x\|_2^2 (\epsilon^2 + 3\epsilon + 2) . \tag{29}
\]
We then have:
\[
\|\nabla f(z)\|_2^2 = \sup_{\|u\|_2 = 1} |u^* \nabla f(z)|^2 \leq 4(2 + \nu)^2 \|h\|_2^2 \|x\|_2^4 \left(\epsilon^2 + 3\epsilon + 2\right)^2.
\] (30)

3. Choosing a suitable set of \((\alpha, \beta, \epsilon, \nu)\)-values.

By combining (27) and (30), we can obtain the regularity condition \(RC(\alpha, \beta, \epsilon)\).

Main result Using Lemma 1 to provide an upper bound on the distance between the spectral initializer \(z^{(0)}\) and a global optimizer \(x\), we can finally state our main theorem:

Theorem 1. Under the rotation invariant measurement model given by (1), when the number of complex quadratic measurements \(m > Cn\) for some universal constant \(C\), and the step size \(0 < \eta < \frac{2}{\beta}\), the gradient descent update (9) initialized with the spectral initializer \(z^{(0)}\) obeying (19) converges linearly to a global optimizer \(x\):

\[
\text{dist}_2\left(z^{(t)}, x\right) \leq \left(1 - \frac{2\eta}{\alpha}\right)^t \cdot \text{dist}_2\left(z^{(0)}, x\right) \leq 1 + 2\eta \delta \|x\|_2^2, \tag{31}
\]

with probability at least
\[
1 - \frac{1}{Cn} \left(\frac{C_1(\sigma, \rho)}{\delta^2} + \frac{C_2(\sigma, \rho)}{\nu^2}\right),
\]
where \(\delta > 0, \nu > 0\) are some pre-specified constants to ensure \(\text{Re}(\langle h, \nabla f(z)\rangle) > 0\), \(\{\alpha, \beta\}\) are constants depending on \(\{\delta, \nu\}\) and \(\{C_1(\sigma, \rho), C_2(\sigma, \rho)\}\) are two constants depending on the variances \(\{\sigma, \rho\}\).

As discussed in the proof of Theorem 1, we first start by choosing the \(\{\delta > 0, \nu > 0\}\) so that the lower bound in (27) is positive. The values of \(\{\alpha, \beta\}\) should satisfy \(\frac{1}{\alpha \beta} \leq 1\) so that the gradient descent updates bring the solution \(z^{(t)}\) closer towards \(x\) in each iteration.

3.1 Choosing the \(\{\alpha, \beta, \epsilon\}\)-values for the regularity condition

There are multiple choices for \(\{\alpha, \beta, \epsilon\}\)-values to make sure the initializer \(z^{(0)}\) stays in the neighborhood of \(x\) and the objective function \(f(x)\) satisfies the regularity condition (20). For example, if we choose \(\delta = 0.01\), using Lemma 1 with high probability we can get that:

\[
\epsilon = \frac{\|h\|_2}{\|x\|_2} \leq \sqrt{\frac{11}{400}} \leq 0.2. \tag{32}
\]

Now, if we choose \(\epsilon \leq 0.2\) and \(\nu = 0.01\), it is easy to verify that

\[
\text{Re}(\langle h, \nabla f(z)\rangle) > 3\|h\|_2^2 \|x\|_2^2 \tag{33}
\]

\[
\|\nabla f(z)\|_2^2 < 120\|h\|_2^2 \|x\|_4^4. \tag{34}
\]
We can then choose $\alpha = \frac{2}{3} \cdot \frac{1}{\|x\|_2^2}$ and $\beta = \frac{2}{3} \cdot 120 \|x\|_2^2$ to obtain:

$$\Re \langle \langle h, \nabla f(z) \rangle \rangle > \frac{3}{2} \|h\|_2^2 \|x\|_2^2 + \frac{3}{2} \|h\|_2^2 \|x\|_2^2$$

$$> \frac{1}{\alpha} \|h\|_2^2 + \frac{1}{\beta} \|\nabla f(z)\|_2^2. \quad (35)$$

When $m$ is sufficiently large, the regularity condition holds for all $h$ satisfying $\epsilon = \frac{\|h\|_2^2}{\|x\|_2^2} \leq 0.2$ with high probability. We can show that the following update converges linearly to a global optimizer:

$$z^{(t+1)} = z^{(t)} - \eta \nabla f(z), \quad (9 \text{ revisited})$$

where $0 < \eta < \frac{2}{\beta}$.

### 3.2 Rotation invariant exponential family

We next give examples of rotation invariant distributions and introduce the rotation invariant exponential family $\mathcal{E}$. The multivariate random variable $r \in \mathbb{R}^{2n^2}$ can be generated from a multivariate standard Gaussian distribution $s \in \mathbb{R}^{2n^2} \sim \mathcal{N}(0, I)$ by

$$r = \kappa \cdot \|s\|_2^q \cdot s. \quad (36)$$

- $\kappa > 0$ is a scaling constant chosen so that the variance $\mu^2 = \text{Var}(r_i) = \mathbb{E}[r_i^2] = 1$.
- $q \neq -1$. In fact if $q = -1$, the variable $r$ is uniformly distributed on the sphere $\|r\|_2 = \kappa$.

As derived in Appendix B, the pdf of $r$ is:

$$p(r) = \frac{1}{q + 1} \kappa^{-\frac{q}{q+1}} \|r\|_2^{-\frac{q}{q+1}} (2\pi)^{-\frac{q}{2}} \exp \left( -\frac{1}{2} \kappa^{-\frac{q}{q+1}} \|r\|_2^{\frac{2}{q+1}} \right). \quad (37)$$

Given $q, \kappa$, we can see that the probability distribution of $r$ only depends on the norm $\|r\|_2$, and is therefore invariant under unitary transforms.

- When $q = 0$, the variable $r = s$ and its entries are iid standard Gaussian. We obtain tighter bounds for the complex Gaussian measurement model in Section 3.2.1.
- When $q \neq 0$ and $q \neq -1$, the entries of $r$ are pairwise uncorrelated.

### 3.2.1 Complex Gaussian Measurement Model

The standard multivariate Gaussian distribution is a special case of a rotation invariant distribution: the individual entries are not just pairwise uncorrelated, but also pairwise independent. In this case we have $\mu^2 = \text{Var}(r_{i,k}) = \mathbb{E}[r_{i,k}^2] = 1$, $\sigma^2 = \text{Var} (r_{i,k}^2) = 2$ and $\rho^2 = \text{Var} (r_{i,k} \cdot r_{i,l}) = 1, \forall k \neq l$. We are able to obtain tighter probability bounds for complex iid Gaussian measurement matrices. For this measurement model, we derive lemmas similar to Lemmas 1 and 2.
Lemma 3. Under the complex Gaussian measurement model, when the number of complex quadratic measurements satisfies $m > Cn$ for some universal constant $C$, the distance between the spectral initializer $z^{(0)}$ and a global optimizer $x$ is upper-bounded as

$$\text{dist}^2(z^{(0)}, x) \leq \frac{11}{4} \delta \|x\|_2^2,$$

with probability at least $1 - 6 \exp(-Cn \cdot C'_1(\delta))$, where $\delta > 0$ is a pre-specified constant, and $C'_1(\delta)$ is some constant depending on $\{\delta\}$.

Lemma 4. Under the complex Gaussian measurement model, when the number of complex quadratic measurements satisfies $m > Cn$ for some universal constant $C$, the following two inequalities hold with probability at least $1 - 20 \exp(-Cn \cdot C'_2(\nu))$:

$$\left| \frac{1}{m} \sum_{i=1}^{m} p^* A_i^* q u^* A_i v - 2 u^* q p^* v \right| \leq \nu \|p\|_2 \|q\|_2 \|u\|_2 \|v\|_2,$$

$$\left| \frac{1}{m} \sum_{i=1}^{m} \text{Re}(p^* A_i^* q u^* A_i v) - 2 \text{Re}(u^* q p^* v) \right| \leq \nu \|p\|_2 \|q\|_2 \|u\|_2 \|v\|_2,$$

where $\nu > 0$ is a pre-specified constant, and $C'_2(\nu)$ is some constant depending on $\{\nu\}$.

We can show that the regularity condition holds with high probability for the complex Gaussian measurement model as before. We then have the following corollary:

Corollary 1. Under the complex Gaussian measurement model given by (1), when the number of complex quadratic measurements $m > Cn$ for some universal constant $C$, and the step size $0 < \eta \leq \frac{2}{\beta}$, the gradient descent update (9) initialized with the spectral initializer $z^{(0)}$ obeying (38) converges linearly to a global optimizer $x$:

$$\text{dist}^2(z^{(t)}, x) \leq \left(1 - \frac{2\eta}{\alpha}\right)^t \cdot \text{dist}^2(z^{(0)}, x) \leq \left(1 - \frac{2\eta}{\alpha}\right)^t \cdot \frac{11}{4} \delta \|x\|_2^2,$$

with probability at least

$$1 - 26 \exp(-Cn \cdot C'_3(\delta, \nu)),$$

where $\delta > 0, \nu > 0$ are some pre-specified constants chosen to ensure $\text{Re}(\langle h, \nabla f(z) \rangle) > 0$, $\{\alpha, \beta\}$ are constants depending on $\{\delta, \nu\}$ and $C'_3(\delta, \nu)$ is some constant depending on $\{\delta, \nu\}$.

We obtain these tighter bounds for the Gaussian measurement model by replacing Chebyshev’s inequality with the Bernstein-type inequality [30, Proposition 5.16] in the proofs.

4 Experimental Results

We perform numerical experiments to empirically evaluate the performance of our approach. Theorem 1 and Corollary 1 state that the step size is upper-bounded by $\frac{2}{\beta}$ where $\beta$ is one of the regularity condition parameters in (20). In Section 3.1 $\beta$ is proportional to the squared norm of the signal to recover. Hence, in all experiments the step size is chosen to be $\frac{0.4}{\|x\|_2}$ where the signal norm, $\|x\|_2$, is estimated using (17).\footnote{Code available at https://github.com/swing-research/random_quadratic_equations under the MIT License.}
Figure 1: Closeness of spectral initialization with varying number of measurements where the complex random measurement matrices are from the rotation invariant exponential family [37].

4.1 Closeness of spectral initializer

In this experiment we monitor how the distance between the initialization and the true solution varies with the number of measurements. We fix $n = 100$ and try different values of $m$ with $\frac{m}{n}$ uniformly sampled between 1 and 10. We run 100 random trials for each $\frac{m}{n}$ value and calculate the average distance between the initialization and a global optimizer. In each trial we generate a random signal $x \in \mathbb{C}^n$ and $m$ complex random rotation invariant matrices from the same distribution [37] to produce $m$ complex quadratic measurements. We repeat this experiment with multiple distributions from the rotation invariant exponential family by varying the $q$ parameter in (37). Recall that $q = 0$ gives us the complex Gaussian measurement model (Section 3.2.1).

Distance between complex signals is defined in (11). The iterative gradient descent reconstruction is terminated if the relative distance between successive iterations is less than $10^{-6}$ or if 2500 iterations are completed. We define relative distance as $\frac{\text{dist}(x, \hat{x})}{\|x\|}$ where $\hat{x}$ is the recovered image and $x$ is the original image. As the true signal and its norm is unknown during gradient descent updates, (17) is used to estimate the norm. In Figure 1 we can see that the spectral initializer becomes increasingly closer to a global optimizer as $\frac{m}{n}$ increases. Furthermore, increasing $q$ increases the average initialization distance.

4.2 Phase transition behavior

In this experiment we evaluate how the proposed approach transits from a failure phase to a success phase as we increase the number of measurements. We fix $n = 100$ and try different values of $m$ with $\frac{m}{n}$ sampled uniformly sampled between 1.5 and 5.5. We again run 100 random trials for each $\frac{m}{n}$ value and calculate the success rate. Success is declared if the distance (11) between the recovered
Figure 2: Success transition plot showing the empirical probability of success based on 100 trials with varying number of measurements where the complex random measurement matrices are from the rotation invariant exponential family \( (37) \).

and true signal is less than \( 10^{-5} \). Again, in each trial a random signal \( x \in \mathbb{C}^n \) is reconstructed and multiple distributions from the rotation invariant exponential family \( (37) \) are used to generate the measurement matrices.

As in Section 4.1, the iterative gradient descent reconstruction is terminated if the relative distance between successive iterations is less than \( 10^{-6} \) or if 2500 iterations are completed. Figure 2 shows that approximately \( 4n \) measurements are needed to successfully recover the signal for the complex Gaussian measurement model \( (q = 0) \) and for when the entries of the measurement matrices are uniformly distributed on the sphere \( (q = 1) \). Increasing \( q \) reduces the success rate which is expected as increasing \( q \) increases the initialization distance as in Figure 1.

4.3 Reconstruction of an image

In this experiment we reconstruct an image via its complex quadratic measurements given by (1). For images the iterative gradient descent reconstruction is terminated if the distance (11) between successive iterations is less than \( 10^{-6} \) or if 2500 iterations are completed. The measurement matrices are complex random Gaussian matrices (Section 3.2.1 and \( q = 0 \) in (37)). We reconstruct the three color channels of an image separately. We choose \( \frac{m}{n} = 4 \). Figure 3 shows the absolute value of the spectral initialization and the corresponding reconstruction. We define the relative error as \( \frac{\|\hat{x} - x\|}{\|x\|} \), where \( |\hat{x}| \) is the absolute value of the recovered image and \( x \) is the original image.

\(^2\)We note that changing the minimum distance required for success or the algorithm stopping criteria can shift the curve.
Figure 3: Spectral initialization and reconstruction of the University of Illinois at Urbana-Champaign logo from its complex random quadratic Gaussian measurements.

Figure 4: Reconstruction performance of the image from Figure 3 with varying number of measurements.

The spectral initialization relative error is 0.34. The relative distances between the three channels of the original and their spectral initializations are 0.53, 0.49 and 0.51. The reconstruction relative error is $4.78 \times 10^{-7}$. The relative distances between the three channels of the original and their reconstructions are $5.45 \times 10^{-7}$, $7.73 \times 10^{-7}$ and $8.66 \times 10^{-7}$. Additionally, we also reconstruct the same image with varying number of measurements. Figure 4 shows the relative distances of the recovered images for each channel.

5 Conclusion and future work

We addressed the problem of recovering a signal $x \in \mathbb{C}^n$ from a system of complex random quadratic equations $y_i = x^*A_ix$, where $\{A_i\}_{i=1}^m$ are independent rotation invariant matrices. When the number of complex measurements $m > Cn$ for some universal constant $C$, we can prove that with high probability: 1) the spectral initializer $z^{(0)}$ is close to a global optimizer, 2) the gradient descent update initialized with $z^{(0)}$ converges linearly to a global optimizer. Numerical experiments
corroborate the theoretical analysis and show that a global optimum can be successfully recovered when $m$ is sufficiently large. Additionally, our analysis complements the existing results for real measurements and rank-1 positive semidefinite measurement matrices.

Recent phase retrieval works show that a regularized spectral initialization and gradient descent update can improve the robustness and performance of the recovery algorithm [31, 32]. Chen, et al., [33] further proved that vanilla gradient descent with random initialization enjoys favorable convergence guarantees in solving the phase retrieval problem. Our future work involves analyzing similar algorithmic enhancements for the random complex Gaussian measurement model as well as extending our approach to recovering structured high-rank complex matrices that arise in key applications.

References

[1] J. R. Fienup, “Phase retrieval algorithms: a comparison,” *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, August 1982.

[2] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, “Phase retrieval with application to optical imaging: A contemporary overview,” *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, May 2015.

[3] K. Jaganathan, Y. C. Eldar, and B. Hassibi, “Phase retrieval: An overview of recent developments,” *CoRR*, vol. abs/1510.07713, October 2015.

[4] P. Duxbury, L. Granlund, S. Gujarathi, P. Juhas, and S. Billinge, “The unassigned distance geometry problem,” *Discrete Applied Mathematics*, vol. 204, no. Supplement C, pp. 117 – 132, May 2016.

[5] S. Huang and I. Dokmanić, “Reconstructing point sets from distance distributions,” *CoRR*, vol. abs/1804.02465, April 2018.

[6] A. Ahmed, B. Recht, and J. Romberg, “Blind deconvolution using convex programming,” *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, March 2014.

[7] J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli, “Phase retrieval for sparse signals: uniqueness conditions,” *CoRR*, vol. abs/1308.3058, August 2013.

[8] T. Dakic, *On the turnpike problem*, Ph.D. thesis, Simon Fraser University, August 2000.

[9] P. Netrapalli, P. Jain, and S. Sanghavi, “Phase retrieval using alternating minimization,” *IEEE Transactions on Information Processing*, vol. 63, no. 18, pp. 4814–4826, September 2015.

[10] E. J. Candès, X. Li, and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *IEEE Transactions on Information Theory*, vol. 61, pp. 1985–2007, April 2015.

[11] N. Boumal, “Nonconvex phase synchronization,” *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2355–2377, 2016.

[12] H. Liu, M. Yue, and A. Man-Cho So, “On the estimation performance and convergence rate of the generalized power method for phase synchronization,” *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2426–2446, 2017.
[13] Y. Chen and M. J. Wainwright, “Fast low-rank estimation by projected gradient descent: general statistical and algorithmic guarantees,” CoRR, vol. abs/1509.03025, September 2015.

[14] W. Bryc, Rotation invariant distributions, pp. 51–69, Springer New York, New York, NY, 1995.

[15] R. Ghods, A. S. Lan, T. Goldstein, and C. Studer, “Linear spectral estimators and an application to phase retrieval,” in Proceedings of the 35th International Conference on Machine Learning, July 2018.

[16] Y. M. Lu and G. Li, “Phase transitions of spectral initialization for high-dimensional nonconvex estimation,” CoRR, vol. abs/1702.06435, August 2017.

[17] X. Jiang, S. Rajan, and X. Liu, “Wirtinger flow method with optimal stepsize for phase retrieval,” IEEE Signal Processing Letters, vol. 23, no. 11, pp. 1627–1631, Nov 2016.

[18] W. Zeng and H. So, “Coordinate descent algorithms for phase retrieval,” CoRR, vol. abs/1706.03474, 2017.

[19] Y. Wang and Z. Xu, “Generalized phase retrieval: measurement number, matrix recovery and beyond,” Applied and Computational Harmonic Analysis, September 2017.

[20] Z. Xu, “The minimal measurement number for low-rank matrix recovery,” Applied and Computational Harmonic Analysis, vol. 44, no. 2, pp. 497 – 508, March 2018.

[21] M. Huang and Z. Xu, “Solving Systems of Quadratic Equations via Exponential-type Gradient Descent Algorithm,” CoRR, vol. abs/1806.00904, June 2018.

[22] A. Conca, D. Edidin, M. Hering, and C. Vinzant, “An algebraic characterization of injectivity in phase retrieval,” Applied and Computational Harmonic Analysis, vol. 38, no. 2, pp. 346 – 356, March 2015.

[23] E. J. Candès, X. Li, and M. Soltanolkotabi, “Phase retrieval from coded diffraction patterns,” Applied and Computational Harmonic Analysis, vol. 39, no. 2, pp. 277 – 299, September 2015.

[24] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” SIAM Review, vol. 52, no. 3, pp. 471–501, August 2010.

[25] Y. Eldar, D. Needell, and Y. Plan, “Uniqueness conditions for low-rank matrix recovery,” Applied and Computational Harmonic Analysis, vol. 33, no. 2, pp. 309 – 314, September 2012.

[26] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege, “Stable low-rank matrix recovery via null space properties,” Information and Inference: A Journal of the IMA, vol. 5, no. 4, pp. 405–441, August 2016.

[27] T. Zhao, Z. Wang, and H. Liu, “A nonconvex optimization framework for low rank matrix estimation,” in Advances in Neural Information Processing Systems 28, pp. 559–567. Curran Associates, Inc., 2015.
[28] 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 28, pp. 109–117. Curran Associates, Inc., 2015.

[29] S. Tu, R. Boczar, M. Simchowitz, M. Soltanolkotabi, and B. Recht, “Low-rank solutions of linear matrix equations via procrustes flow,” Proceedings of the 33rd International Conference on International Conference on Machine Learning, vol. 48, pp. 964–973, June 2016.

[30] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, p. 210268, Cambridge University Press, 2012.

[31] Y. Chen and E. Candes, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” Advances in Neural Information Processing Systems, pp. 739–747, December 2015.

[32] G. Wang, G. B. Giannakis, and Y. C. Eldar, “Solving systems of random quadratic equations via truncated amplitude flow,” IEEE Transactions on Information Theory, vol. 64, no. 2, pp. 773–794, February 2018.

[33] Y. Chen, Y. Chi, J. Fan, and C. Ma, “Gradient descent with random initialization: fast global convergence for nonconvex phase retrieval,” CoRR, vol. 1803.07726, March 2018.
A Proofs for Rotation Invariant Measurement Model

A.1 Proof of Lemma 1

Proof. The spectral initializer $z^{(0)}$ can be the left or right leading singular vector of the matrix $S$ in (12) scaled by $\|x\|_2$. By the rotational invariance of the random matrix $A_i$, we can simply choose $x = |x|_2 \cdot e_1$. For the complex entries of $A$, we have:

$$E[A_{rc}^{(R)}] = E[A_{rc}^{(L)}] = 1$$

$$Var(A_{rc}^{(R)}) = Var(A_{rc}^{(L)}) = \sigma^2$$

$$E[A_{rc}^{(R)} A_{kl}^{(R)}] = E[A_{rc}^{(L)} A_{kl}^{(L)}] = E[A_{rc}^{(R)} A_{kl}^{(L)}] = 0$$

$$Var(A_{rc}^{(R)} A_{kl}^{(R)}) = Var(A_{rc}^{(L)} A_{kl}^{(L)}) = Var(A_{rc}^{(R)} A_{kl}^{(L)}) = \rho^2,$$

where $(r, c) \neq (k, l)$.

We first show that $S$ concentrates around $E[S] = 2xx^*$ in terms of the spectral norm: $\|S - E[S]\| \leq \delta$ holds with high probability. Specifically, we have:

$$\|S - 2xx^*\| = \left\| \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \cdot e_1^T A_i^* e_1 \cdot A_i - 2\|x\|_2^2 \cdot e_1 e_1^T \right\|$$

$$\leq \left\| \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 (\bar{A}_{i,11} A_{i,11} - 2) e_1 e_1^T \right\| + \left\| \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \bar{A}_{i,11} \tilde{A}_i \right\|,$$

where $\tilde{A}_i$ is the matrix $A_i$ with its $(1, 1)$-th entry replaced by 0.

- For the first term of (46), we have:

$$\left\| \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 (\bar{A}_{i,11} A_{i,11} - 2) e_1 e_1^T \right\| = \left\| \frac{1}{m} \sum_{i=1}^{m} (\bar{A}_{i,11} A_{i,11} - 2) \right\| \leq \left\| \frac{m}{m} \sum_{i=1}^{m} (\bar{A}_{i,11} A_{i,11} - 2) \right\|.$$  

Using Chebyshev’s inequality, if $m \geq Cn$ for some constant $C$, we have:

$$Pr \left( \left\| \frac{m}{m} \sum_{i=1}^{m} (A_{i,11}^{(R)} - 1) \right\| \geq \frac{\delta}{4} \|x\|_2^2 \right) \leq \frac{16\sigma^2}{m\delta^2} \leq \frac{1}{Cn} \frac{16\sigma^2}{\delta^2}$$

$$Pr \left( \left\| \frac{m}{m} \sum_{i=1}^{m} (A_{i,11}^{(L)} - 1) \right\| \geq \frac{\delta}{4} \|x\|_2^2 \right) \leq \frac{16\sigma^2}{m\delta^2} \leq \frac{1}{Cn} \frac{16\sigma^2}{\delta^2}.$$  

- Finding an upper bound on the second term in (46) is equivalent to finding an upper bound on the following (50) for all $u, v \in \mathbb{C}^n$ satisfying $\|u\|_2 = 1, \|v\|_2 = 1$:

$$\left\| u^* \cdot \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \bar{A}_{i,11} \tilde{A}_i \cdot v \right\| \leq \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \bar{A}_{i,11} \cdot \sum_{rc} \bar{u}_r v_c \cdot \tilde{A}_{i,rc} \right\| \leq \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \bar{A}_{i,11} \cdot G_i,$$  

(50)
where $G_i = \sum_{rc} \pi_r v_c \cdot \bar{A}_{i,rc}$. Next, let's examine the variable $\bar{A}_{i,11} \cdot G_i$:

$$\bar{A}_{i,11} \cdot G_i = \left( A_{i,11}^{(R)} - A_{i,11}^{(I)} j \right) \cdot \left( G_i^{(R)} + G_i^{(I)} j \right) = \left( A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right) + \left( A_{i,11}^{(R)} G_i^{(I)} - A_{i,11}^{(I)} G_i^{(R)} \right) j .$$

We then have:

$$\left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} \bar{A}_{i,11} G_i \right| = \left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right) + \frac{\|x\|^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_i^{(I)} - A_{i,11}^{(I)} G_i^{(R)} \right) j \right|$$

$$\leq \left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right) \right| + \left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_i^{(I)} - A_{i,11}^{(I)} G_i^{(R)} \right) \right| .$$

Take the first term of (52) for example, let $B_{rc} = \pi_r v_c$ we have:

$$A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} = \sum_{(r,c) \neq (1,1)} B_{rc}^{(R)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} - B_{rc}^{(I)} \cdot A_{i,11}^{(I)} A_{i,rc}^{(R)} + B_{rc}^{(R)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} + B_{rc}^{(I)} \cdot A_{i,11}^{(I)} A_{i,rc}^{(R)} .$$

Since the coefficients of $A_i$ follow a spherically symmetric distribution, we can get

$$\Pr \left( A_{i,11}^{(R)} A_{i,rc}^{(I)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} \right) = \Pr \left( -A_{i,11}^{(R)} A_{i,rc}^{(I)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} \right)$$

$$\mathbb{E} \left[ A_{i,11}^{(R)} A_{i,rc}^{(I)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} \right] = 0 = \mathbb{E} \left[ A_{i,11}^{(R)} A_{i,rc}^{(I)} \cdot A_{i,11}^{(R)} A_{i,rc}^{(I)} \right] .$$

Hence $A_{i,11}^{(R)} A_{i,rc}^{(I)}$ and $A_{i,11}^{(R)} A_{i,rc}^{(I)}$ are uncorrelated. Similarly, we can see that (53) is a summation of pairwise uncorrelated terms.

For the first term of (52), we can compute that:

$$\mathbb{E} \left[ \frac{\|x\|^2}{m} \sum_{i=1}^{m} A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right] = 0$$

$$\text{Var} \left( \frac{\|x\|^2}{m} \sum_{i=1}^{m} A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right) = \frac{\rho^2 \|x\|^4}{m} \sum_{(r,c) \neq (1,1)} 2|\pi_r v_c|^2 \leq \frac{2\rho^2 \|x\|^4}{m} .$$

Using Chebyshev's inequality, if $m \geq Cn$, we have:

$$\Pr \left( \left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} A_{i,11}^{(R)} G_i^{(R)} + A_{i,11}^{(I)} G_i^{(I)} \right| \geq \frac{\delta}{4} \|x\|^2 \right) \leq \frac{1}{Cn} \frac{32\rho^2}{\delta^2} .$$

Similar result can be obtained for the second term of (52):

$$\Pr \left( \left| \frac{\|x\|^2}{m} \sum_{i=1}^{m} A_{i,11}^{(R)} G_i^{(I)} - A_{i,11}^{(I)} G_i^{(R)} \right| \geq \frac{\delta}{4} \|x\|^2 \right) \leq \frac{1}{Cn} \frac{32\rho^2}{\delta^2} .$$
Combining (48), (49), (58), (59), if $m \geq Cn$, we have:

$$\Pr \left( \| S - 2xx^* \| \leq \delta \| x \|_2^2 \right) \geq \Pr \left( \frac{\| x \|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} - 1 \right) \leq \frac{\delta}{4} \| x \|_2^2 \right) \cdot \Pr \left( \frac{\| x \|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(I)} - 1 \right) \leq \frac{\delta}{4} \| x \|_2^2 \right) \cdot \Pr \left( \frac{\| x \|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_{i}^{(R)} + A_{i,11}^{(I)} G_{i}^{(I)} \right) \leq \delta \| x \|_2^2 \right) \cdot \Pr \left( \frac{\| x \|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_{i}^{(I)} - A_{i,11}^{(I)} G_{i}^{(R)} \right) \leq \delta \| x \|_2^2 \right)$$

(60)

$$\geq \left( 1 - \frac{1}{Cn} \frac{16\sigma^2}{\delta^2} \right)^2 \cdot \left( 1 - \frac{1}{Cn} \frac{32\rho^2}{\delta^2} \right)^2$$

$$> 1 - \frac{1}{Cn} \cdot \frac{32\sigma^2 + 64\rho^2}{\delta^2}$$

$$= 1 - \frac{1}{Cn} \cdot C_1(\sigma, \rho) \frac{\delta^2}{\delta^2}.$$  

So far we have shown that $\| S - 2xx^* \| \leq \delta \| x \|_2^2$ holds with high probability when $m \geq Cn$. Let $\{u_0, v_0\}$ be the leading left and right singular vectors of $S$ and $\tau_0$ be the largest singular value of $S$. With high probability we have:

$$|\tau_0 - u_0^* \cdot 2xx^* \cdot v_0| = |u_0^* (S - 2xx^*) v_0|$$

$$\leq \| S - 2xx^* \|$$

$$\leq \delta \| x \|_2^2.$$  

(61)

Hence $u_0^* \cdot 2xx^* \cdot v_0 \geq \tau_0 - \delta \| x \|_2^2$. With high probability we also have:

$$\tau_0 \geq \frac{1}{\| x \|_2^2} x^* S x$$

$$= \frac{1}{\| x \|_2^2} x^* (S - 2xx^*) x + 2 \| x \|_2^2$$

$$\geq (2 - \delta) \| x \|_2^2.$$  

(62)

1) When the signal norm $\| x \|_2$ is known, using Cauchy-Schwartz inequality, we have $|u_0^* x| \leq \| x \|_2$. If $\| x \|_2 v_0$ is chosen as the spectral initializer, we can get:

$$2 |x^* v_0 \cdot \| x \|_2 \| \geq 2 |u_0^* x \cdot x^* v_0|$$

$$\geq u_0^* \cdot 2xx^* \cdot v_0$$

$$\geq (2 - \delta) \| x \|_2^2.$$  

(63)

Using (11), the squared distance between the spectral initializer $z^{(0)} = \| x \|_2 v_0$ and $x$ is then
upper-bounded by $2\delta \|x\|_2^2$:

$$\text{dist}^2(\|x\|_2v_0, x) = \min_{\phi_0 \in [0, 2\pi]} \|x\|_2v_0 - xe^{j\phi_0}\|_2^2$$

$$= \|\|x\|_2v_0\|_2^2 + \|x\|_2^2 - 2\|x\|_2v_0^*x\|_2$$

$$\leq 2\|x\|_2^2 - (2 - 2\delta)\|x\|_2^2$$

$$= 2\delta \|x\|_2^2,$$

with probability at least $1 - \frac{1}{C_n} \cdot \frac{C_1(\sigma, \rho)}{\delta^2}$.  

2) When the signal norm $\|x\|_2$ is unknown, we need to estimate it using $R = \frac{1}{2m} \sum_{i=1}^{m} y_i^*y_i$. By rotational invariance property of the random Gaussian matrix $A_i$, we can simply assume $x = \|x\|_2e_1$. We have:

$$R = \|x\|_2^2 \sum_{i=1}^{m} A_{i,11} \cdot A_{i,11}. \tag{65}$$

Using (48) and (49), when $m \geq Cn$ we know that the following equality

$$\left(2 - \frac{\delta}{2}\right) m \leq \sum_{i=1}^{m} \|A_{i,11}\|^2 \leq \left(2 + \frac{\delta}{2}\right) m, \tag{66}$$

holds with probability at least $1 - \frac{1}{C_n} \cdot \frac{32\sigma^2}{\delta^2}$. The following inequalities then hold with high probability:

$$\left(1 - \frac{\delta}{4}\right) \|x\|_2^4 \leq R \leq \left(1 + \frac{\delta}{4}\right) \|x\|_2^4. \tag{67}$$

The spectral initializer is chosen to be $z^{(0)} = \sqrt{R}v_0$. Assuming $\delta < 4$, we have the following

$$\text{dist}^2(\sqrt{R}v_0, x) = \|\sqrt{R}v_0 - x\|_2^2$$

$$\leq \sqrt{R} + \|x\|_2^2 - (2 - 2\delta) \sqrt{R}\|x\|_2$$

$$\leq \sqrt{1 + \frac{\delta}{4}} \|x\|_2^2 + \|x\|_2^2 - 2(1 - \delta) \sqrt{1 - \frac{\delta}{4}}\|x\|_2^2$$

$$\leq \left(1 + \frac{\delta}{4}\right) \|x\|_2^2 + \|x\|_2^2 - 2(1 - \delta) \left(1 - \frac{\delta}{4}\right)\|x\|_2^2$$

$$= \left(\frac{11}{4} \delta - \frac{1}{2}\delta^2\right) \|x\|_2^2$$

$$\leq \frac{11}{4} \delta\|x\|_2^2,$$

holds with probability at least $1 - \frac{1}{C_n} \cdot \frac{C_1(\sigma, \rho)}{\delta^2}$. \qed
A.2 Proof of Lemma 2

Proof. Let $\mathbf{p} = \|\mathbf{p}\|\mathbf{\hat{p}}$, $\mathbf{q} = \|\mathbf{q}\|\mathbf{\hat{q}}$, $\mathbf{u} = \|\mathbf{u}\|\mathbf{\hat{u}}$, and $\mathbf{v} = \|\mathbf{v}\|\mathbf{\hat{v}}$, we first prove the inequality (22). We need to show the following:

$$\left\| \frac{1}{m} \sum_{i=1}^{m} \mathbf{p}^* A_i^* \mathbf{\hat{q}} A_i \mathbf{\hat{v}} - 2 \mathbf{\hat{u}}^* \mathbf{\hat{q}}^* \mathbf{\hat{v}} \right\| = \left\| \mathbf{\hat{u}}^* \left( \frac{1}{m} \sum_{i=1}^{m} \mathbf{p}^* A_i^* \mathbf{\hat{q}} A_i - 2 \mathbf{\hat{q}}^* \right) \mathbf{\hat{v}} \right\| \leq \frac{1}{m} \sum_{i=1}^{m} \mathbf{p}^* A_i^* \mathbf{\hat{q}} A_i - 2 \mathbf{\hat{q}}^*$$

(69)

By rotational invariance, let $\mathbf{\hat{p}} = \mathbf{e}_1$, $\mathbf{\hat{q}} = r_1 e^{\phi_1} \mathbf{e}_1 + r_2 e^{\phi_2} \mathbf{e}_2$. Let $b_i = r_1 e^{\phi_1} A_{i,11} + r_2 e^{\phi_2} A_{i,21}$, and $\hat{A}_i$ denote the matrix $A_i$ with the $(1,1)$-th and $(2,1)$-th entries replaced by 0s, we have:

$$\left\| \frac{1}{m} \sum_{i=1}^{m} \mathbf{p}^* A_i^* \mathbf{\hat{q}} A_i - 2 \mathbf{\hat{q}}^* \right\| \leq \left\| \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2 r_1 e^{\phi_1} \right\| + \left\| \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,21} - 2 r_2 e^{\phi_2} \right\| + \left\| \frac{1}{m} \sum_{i=1}^{m} b_i \hat{A}_i \right\|$$

(70)

- For the first term of (70), we have:

$$\left| \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2 r_1 e^{\phi_1} \right|$$

$$= \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 \left( \|A_{i,11}\|^2 - 2 \right) + r_2 \cos \phi_2 A_{i,21}^\text{(R)} A_{i,11}^\text{(R)} + r_2 \cos \phi_2 A_{i,21}^\text{(I)} A_{i,11}^\text{(I)}$$

$$- r_2 \sin \phi_2 A_{i,21}^\text{(R)} A_{i,11}^\text{(I)} + r_2 \sin \phi_2 A_{i,21}^\text{(I)} A_{i,11}^\text{(R)}$$

$$\leq \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 \left( A_{i,11}^\text{(R)} - 1 \right) + \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 \left( A_{i,11}^\text{(I)} - 1 \right)$$

$$+ \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 A_{i,21}^\text{(R)} A_{i,11}^\text{(R)} + r_2 \cos \phi_2 A_{i,21}^\text{(I)} A_{i,11}^\text{(I)} - r_2 \sin \phi_2 A_{i,21}^\text{(R)} A_{i,11}^\text{(I)} + r_2 \sin \phi_2 A_{i,21}^\text{(I)} A_{i,11}^\text{(R)}$$

(71)

It is also easy to verify that $A_{i,21}^\text{(R)} A_{i,11}^\text{(R)}$, $A_{i,21}^\text{(I)} A_{i,11}^\text{(I)}$, $A_{i,21}^\text{(R)} A_{i,11}^\text{(I)}$ and $A_{i,21}^\text{(I)} A_{i,11}^\text{(R)}$ are pairwise uncorrelated.
Using Chebyshev’s inequality, if \( m \geq Cn \), we can get:

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 \left( A_{i,11}^{(R)} - 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576r_1^2 \cos^2 \phi_1 \sigma^2}{\nu^2} \tag{72}
\]

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 \left( A_{i,11}^{(I)} - 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576r_1^2 \cos^2 \phi_1 \sigma^2}{\nu^2} \tag{73}
\]

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 A_{i,21}^{(R)} A_{i,11}^{(R)} + r_2 \cos \phi_2 A_{i,21}^{(I)} A_{i,11}^{(I)} - r_2 \sin \phi_2 A_{i,21}^{(R)} A_{i,11}^{(I)} + r_2 \sin \phi_2 A_{i,21}^{(I)} A_{i,11}^{(R)} \geq \frac{\nu}{12} \right) \leq \frac{1}{Cn} \frac{288r_1^2 \rho^2}{\nu^2}. \tag{74}
\]

- For the second term of (70), we have:

\[
\left| \text{Re} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,21} - 2r_2 e^{j\phi_2} \right) \right| = \left| \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( \|A_{i,21}\|_2^2 - 2 \right) + r_1 \cos \phi_1 A_{i,21}^{(R)} A_{i,11}^{(R)} + r_1 \cos \phi_1 A_{i,21}^{(I)} A_{i,11}^{(I)} - r_1 \sin \phi_1 A_{i,21}^{(I)} A_{i,11}^{(R)} + r_1 \sin \phi_1 A_{i,21}^{(R)} A_{i,11}^{(I)} \right| \\
\leq \left| \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( A_{i,21}^{(R)} - 1 \right) \right| + \left| \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( A_{i,21}^{(I)} - 1 \right) \right| \\
+ \left| \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 A_{i,21}^{(R)} A_{i,11}^{(R)} + r_1 \cos \phi_1 A_{i,21}^{(I)} A_{i,11}^{(I)} - r_1 \sin \phi_1 A_{i,21}^{(I)} A_{i,11}^{(R)} + r_1 \sin \phi_1 A_{i,21}^{(R)} A_{i,11}^{(I)} \right|. \tag{75}
\]

We have the following probabilities:

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( A_{i,21}^{(R)} - 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576r_2^2 \cos^2 \phi_2 \sigma^2}{\nu^2} \tag{76}
\]

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( A_{i,21}^{(I)} - 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576r_2^2 \cos^2 \phi_2 \sigma^2}{\nu^2} \tag{77}
\]

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 A_{i,21}^{(R)} A_{i,11}^{(R)} + r_1 \cos \phi_1 A_{i,21}^{(I)} A_{i,11}^{(I)} - r_1 \sin \phi_1 A_{i,21}^{(I)} A_{i,11}^{(R)} + r_1 \sin \phi_1 A_{i,21}^{(R)} A_{i,11}^{(I)} \geq \frac{\nu}{12} \right) \leq \frac{1}{Cn} \frac{288r_1^2 \rho^2}{\nu^2}. \tag{78}
\]
• For the third term of (70), we have:

\[
\left| \text{Im} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2r_1 e^{j \phi_1} \right) \right|
\]

\[
= \frac{1}{m} \sum_{i=1}^{m} r_1 \sin \phi_1 \left( \| A_{i,11} \|^2 \right) - 2r_1 \cos \phi_2 A_{i,21} A_{i,11} + r_2 \cos \phi_2 A_{i,21} A_{i,11}
\]

\[
+ r_2 \sin \phi_2 A_{i,21} A_{i,11} + r_2 \sin \phi_2 A_{i,21} A_{i,11} \right|
\]

\[
\leq \frac{1}{m} \sum_{i=1}^{m} r_1 \sin \phi_1 \left( A_{i,11}^{(R)} + 1 \right) + \frac{1}{m} \sum_{i=1}^{m} r_1 \sin \phi_1 \left( A_{i,11}^{(I)} - 1 \right)
\]

\[
+ \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 A_{i,21} A_{i,11} - r_2 \cos \phi_2 A_{i,21} A_{i,11} + r_2 \sin \phi_2 A_{i,21} A_{i,11} + r_2 \sin \phi_2 A_{i,21} A_{i,11} \right|
\]

(79)

We have the following probabilities:

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \sin \phi_1 \left( A_{i,11}^{(R)} + 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576 r_1^2 \sin^2 \phi_1 \sigma^2}{\nu^2}
\]

(80)

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \sin \phi_1 \left( A_{i,11}^{(I)} - 1 \right) \geq \frac{\nu}{24} \right) \leq \frac{1}{Cn} \frac{576 r_1^2 \sin^2 \phi_1 \sigma^2}{\nu^2}
\]

(81)

\[
\text{Pr} \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 A_{i,21} A_{i,11} - r_2 \cos \phi_2 A_{i,21} A_{i,11} + r_2 \sin \phi_2 A_{i,21} A_{i,11} + r_2 \sin \phi_2 A_{i,21} A_{i,11} \right| \geq \frac{\nu}{12} \right)
\]

\[
\leq \frac{1}{Cn} \frac{288 r_2^2 \rho^2}{\nu^2}
\]

(82)

• For the fourth term of (70), we have:

\[
\left| \text{Im} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,21} - 2r_2 e^{j \phi_2} \right) \right|
\]

\[
= \frac{1}{m} \sum_{i=1}^{m} r_2 \sin \phi_2 \left( \| A_{i,21} \|^2 \right) - 2r_2 \cos \phi_1 A_{i,21} A_{i,11} + r_1 \cos \phi_1 A_{i,21} A_{i,11}
\]

\[
+ r_1 \sin \phi_1 A_{i,21} A_{i,11} + r_1 \sin \phi_1 A_{i,21} A_{i,11} \right|
\]

\[
\leq \frac{1}{m} \sum_{i=1}^{m} r_2 \sin \phi_2 \left( A_{i,21}^{(R)} + 1 \right) + \frac{1}{m} \sum_{i=1}^{m} r_2 \sin \phi_2 \left( A_{i,21}^{(I)} - 1 \right)
\]

\[
+ \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_1 A_{i,21} A_{i,11} - r_2 \cos \phi_1 A_{i,21} A_{i,11} + r_2 \sin \phi_1 A_{i,21} A_{i,11} + r_2 \sin \phi_1 A_{i,21} A_{i,11} \right|
\]

(83)
We have the following probabilities:

\[
\Pr \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \sin \phi_2 \left( A_{i,21}^{(R)} \right)^2 - 1 \geq \frac{\nu}{24} \right) \leq \frac{1}{C_n} \frac{576 r_2^2 \sin^2 \phi_2 \sigma^2}{\nu^2} \tag{84}
\]

\[
\Pr \left( \frac{1}{m} \sum_{i=1}^{m} r_2 \sin \phi_2 \left( A_{i,21}^{(I)} \right)^2 - 1 \geq \frac{\nu}{24} \right) \leq \frac{1}{C_n} \frac{576 r_2^2 \sin^2 \phi_2 \sigma^2}{\nu^2} \tag{85}
\]

\[
\Pr \left( \frac{1}{m} \sum_{i=1}^{m} r_1 \cos \phi_1 A_{i,21}^{(R)} A_{i,11}^{(R)} - r_1 \cos \phi_1 A_{i,21}^{(R)} A_{i,11}^{(I)} + r_1 \sin \phi_1 A_{i,21}^{(R)} A_{i,11}^{(R)} + r_1 \sin \phi_1 A_{i,21}^{(I)} A_{i,11}^{(I)} \geq \frac{\nu}{12} \right) \leq \frac{1}{C_n} \frac{288 \rho^2}{\nu^2} . \tag{86}
\]

To make sure the fifth term of (70) \( \| \frac{1}{m} \sum_{i=1}^{m} b_i \tilde{A}_i \| \leq \frac{\nu}{3} \), it suffices to prove the term \( \| \frac{1}{m} \sum_{i=1}^{m} b_i g^* \tilde{A}_i w \| \) is bounded for all \( g \in \mathbb{C}^n, \ w \in \mathbb{C}^n \) satisfying \( \| g \|_2 = 1, \| w \|_2 = 1 \). Specifically we have:

\[
\left| \frac{1}{m} \sum_{i=1}^{m} b_i g^* \tilde{A}_i w \right| = \left| \frac{1}{m} \sum_{i=1}^{m} b_i \sum_{k,l} \bar{g}_k w_l \cdot \tilde{A}_{i,kl} \right| = \left| \frac{1}{m} \sum_{i=1}^{m} b_i \cdot G_i \right| , \tag{87}
\]

where \( G_i = \sum_{k,l} \bar{g}_k w_l \cdot \tilde{A}_{i,kl} \). Let \( B_{k,l} = \bar{g}_k w_l \), we have:

\[
G_i = \sum_{(k,l) \neq (1,1)} (B_{k,l} A_{i,kl}^{(R)} - B_{k,l} A_{i,kl}^{(I)}) + (B_{k,l} A_{i,kl}^{(I)} + B_{k,l} A_{i,kl}^{(R)}) j . \tag{88}
\]

The variable \( b_i \) can be expressed as:

\[
b_i = r_1 \left( \cos \phi_1 A_{i,11}^{(R)} A_{i,11}^{(R)} + \sin \phi_1 A_{i,11}^{(I)} A_{i,11}^{(I)} \right) + r_2 \left( \cos \phi_2 A_{i,21}^{(R)} A_{i,21}^{(R)} + \sin \phi_2 A_{i,21}^{(I)} A_{i,21}^{(I)} \right) + \left[ r_1 \left( \sin \phi_1 A_{i,11}^{(R)} A_{i,11}^{(R)} - \cos \phi_1 A_{i,11}^{(I)} A_{i,11}^{(I)} \right) + r_2 \left( \sin \phi_2 A_{i,21}^{(R)} A_{i,21}^{(R)} - \cos \phi_2 A_{i,21}^{(I)} A_{i,21}^{(I)} \right) \right] j . \tag{89}
\]

We then have:

\[
\left| \frac{1}{m} \sum_{i=1}^{m} b_i \cdot G_i \right| = \left| \frac{1}{m} \sum_{i=1}^{m} \left( b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right) + \frac{1}{m} \sum_{i=1}^{m} \left( b_i^{(R)} G_i^{(I)} + b_i^{(I)} G_i^{(R)} \right) \right| \leq \left| \frac{1}{m} \sum_{i=1}^{m} \left( b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right) \right| + \left| \frac{1}{m} \sum_{i=1}^{m} \left( b_i^{(R)} G_i^{(I)} + b_i^{(I)} G_i^{(R)} \right) \right| . \tag{90}
\]
• For the first term of (90), we can compute that:

\[
\begin{align*}
&b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \\
&= \sum_{(k,l) \neq (1,1)} r_1 \left( B_{kl}^{(R)} \cos \phi_1 - B_{kl}^{(I)} \sin \phi_1 \right) A_{i,k}^{(R)} A_{i,l}^{(R)} + r_1 \left( B_{kl}^{(R)} \sin \phi_1 + B_{kl}^{(I)} \cos \phi_1 \right) A_{i,k}^{(R)} A_{i,l}^{(R)} \\
&\hspace{1cm} + r_2 \left( B_{kl}^{(R)} \cos \phi_2 - B_{kl}^{(I)} \sin \phi_2 \right) A_{i,k}^{(R)} A_{i,21}^{(R)} + r_2 \left( B_{kl}^{(R)} \sin \phi_2 + B_{kl}^{(I)} \cos \phi_2 \right) A_{i,k}^{(R)} A_{i,21}^{(R)} \\
&\hspace{1cm} - r_1 \left( B_{kl}^{(R)} \cos \phi_1 + B_{kl}^{(I)} \sin \phi_1 \right) A_{i,k}^{(R)} A_{i,11}^{(R)} - r_1 \left( B_{kl}^{(I)} \sin \phi_1 - B_{kl}^{(R)} \cos \phi_1 \right) A_{i,k}^{(I)} A_{i,11}^{(I)} \\
&\hspace{1cm} - r_2 \left( B_{kl}^{(I)} \cos \phi_2 + B_{kl}^{(R)} \sin \phi_2 \right) A_{i,k}^{(I)} A_{i,21}^{(I)} - r_2 \left( B_{kl}^{(I)} \sin \phi_2 - B_{kl}^{(R)} \cos \phi_2 \right) A_{i,k}^{(I)} A_{i,21}^{(I)}.
\end{align*}
\]  

(91)

It is easy to verify that (91) is a summation of pairwise uncorrelated terms. We can compute:

\[
\begin{align*}
&\mathbb{E} \left[ \frac{1}{m} \sum_{i=1}^{m} b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right] = 0 \quad \text{(92)} \\
&\text{Var} \left( \frac{1}{m} \sum_{i=1}^{m} b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right) = \frac{2\rho^2}{m} \sum_{(k,l) \neq (1,1)} |\mathcal{G}_k w_l|^2 \leq \frac{2\rho^2}{m}. \quad \text{(93)}
\end{align*}
\]

Using the Chebyshev’s inequality, if \( m \geq Cn \), we can get:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right| \geq \nu \right) \leq \frac{1}{Cn} \frac{72\rho^2}{\nu^2}. \quad \text{(94)}
\]

• We can derive similar result for the second term of (90).

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} b_i^{(R)} G_i^{(I)} + b_i^{(I)} G_i^{(R)} \right| \geq \nu \right) \leq \frac{1}{Cn} \frac{72\rho^2}{\nu^2}. \quad \text{(95)}
\]

• If \( m \geq Cn \), combining all the probability bounds, we can get the following probability bound:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} \mathbf{p}_* \mathbf{A}_i^* \hat{\mathbf{q}} \mathbf{u}^* \mathbf{A}_i \hat{\mathbf{v}} - 2\mathbf{\hat{u}}^* \mathbf{q} \mathbf{p}^* \hat{\mathbf{v}} \right| \leq \nu \right) \geq 1 - \frac{1}{Cn} \frac{1152\sigma^2 + 720\rho^2}{\nu^2} \\
\hspace{3cm} = 1 - \frac{1}{Cn} \cdot \frac{C_2(\sigma, \rho)}{\nu^2}. \quad \text{(96)}
\]

Hence (22) also holds with high probability.

For any \( x \in \mathbb{C} \), we have \(|\text{Re}(x)| \leq |x|\), hence (23) holds with high probability. \(\square\)

### A.3 Proof of Theorem 1

**Proof.** According to Lemma 1, starting from the spectral initializer \( \mathbf{z}^{(0)} \), the following

\[
\epsilon^2 = \frac{\|h\|^2_2}{\|x\|^2_2} \leq \frac{11}{4} \delta, \quad \text{(97)}
\]
holds with probability at least $1 - \frac{1}{C_n} \cdot \frac{C_1(\sigma, \rho)}{\delta^2}$.

Plugging (97) into (27), we have:

$$\text{Re} (\langle h, \nabla f(z) \rangle) \geq 4 \left( 1 - \sqrt{\frac{11}{4} \delta - \frac{\nu}{2}} \left( 2 + 3 \sqrt{\frac{11}{4} \delta + \frac{11}{4} \delta} \right) \right) \cdot \|h\|_2^2 \|x\|_2^2$$

$$= c_1(\delta, \nu) \cdot \|h\|_2^2 \|x\|_2^2,$$

where $\delta > 0$, $\nu > 0$ should be chosen so that $c_1(\delta, \nu) > 0$. Plugging (97) into (30), we have:

$$\|\nabla f(z)\|_2^2 \leq 4(2 + \nu)^2 \left( \frac{11}{4} \delta + 3 \sqrt{\frac{11}{4} \delta + 2} \right)^2 \cdot \|h\|_2^2 \|x\|_4^4$$

$$= c_2(\delta, \nu) \cdot \|h\|_2^2 \|x\|_4^4.$$  

Both (98) and (99) hold with probability at least $1 - \frac{1}{C_n} \cdot \frac{C_1(\sigma, \rho)}{\nu^2}$ according to Lemma 2.

We next discuss how to select the $\{\alpha, \beta\}$-values. In order to achieve a linear convergence rate for (21), the $\alpha$ and $\beta$ values of (20) should satisfy $4 \alpha \beta \leq 1$. Let $\alpha = a \cdot \frac{1}{\|x\|_2^2}$ and $\beta = b \cdot c_2(\delta, \nu) \|x\|_2^2$, we then require:

$$\frac{4}{ab} \leq c_2(\delta, \nu).$$  

In order to satisfy the regularity condition (20), we require:

$$\text{Re} (\langle h, \nabla f(z) \rangle) \geq c_1(\delta, \nu) \cdot \|h\|_2^2 \|x\|_2^2$$

$$\geq \frac{1}{\alpha} \|h\|_2^2 + \frac{1}{\beta} \cdot c_2(\delta, \nu) \|h\|_2^2 \|x\|_4^4$$

$$= \left( \frac{1}{a} + \frac{1}{b} \right) \|h\|_2^2 \|x\|_2^2.$$  

In other words, we require:

$$\frac{1}{a} + \frac{1}{b} \leq c_1(\delta, \nu).$$  

Given some $\delta > 0$, $\nu > 0$ satisfying $c_1(\delta, \nu) > 0$, there are multiple choices of $\alpha, \beta$ that satisfy the regularity condition (20). In order to speed up the convergence towards $x$, we would like to find the $a$, $b$ that maximize $\frac{4}{ab}$ subject to (100) and (102).

- If $c_2(\delta, \nu) \geq c_1(\delta, \nu)^2$, we can choose

$$a = \frac{2}{c_1(\delta, \nu)},$$

$$b = \frac{2}{c_1(\delta, \nu)}.$$

- If $c_2(\delta, \nu) < c_1(\delta, \nu)^2$, any $a$, $b$ that satisfy the following constraints would work:

$$ab = \frac{4}{c_2(\delta, \nu)},$$

$$a + b \leq \frac{4c_1(\delta, \nu)}{c_2(\delta, \nu)}.$$
Using Lemma 7.10 and combining (97)-(99), we can get that (31) holds with probability at least 
\[ 1 - \frac{1}{C_n} \cdot \left( \frac{C_1(\sigma, \rho) \delta^2}{\sigma^2} + \frac{C_2(\sigma, \rho) \nu^2}{\nu^2} \right). \]

### B Rotation invariant exponential family

Since the random variable \( r = \kappa \cdot \|s\|^q_2 \cdot s \), where \( q \neq -1 \), and \( s \in \mathbb{R}^d \sim N(0, I) \), we then have:

\[ s = \kappa^{-\frac{1}{q+1}} \cdot \|r\|^{-\frac{q}{q+1}}_2 \cdot r. \]  

(107)

where \( \kappa = \kappa(q) \) ensures that \( \mathbb{E}[r_i^2] = 1 \).

Let \( J \) be the Jacobian matrix of \( s \) w.r.t. \( r \), its entries are given by:

\[ \frac{\partial s_i}{r_i} = \kappa^{-\frac{1}{q+1}} \cdot \|r\|^{-\frac{q}{q+1}}_2 \left( 1 - \frac{q}{q+1} \cdot \|r\|^{-2}_2 \cdot r_i^2 \right) \]  

(108)

\[ \frac{\partial s_i}{r_j} = \kappa^{-\frac{1}{q+1}} \cdot \|r\|^{-\frac{q}{q+1}}_2 \left( -\frac{q}{q+1} \cdot \|r\|^{-2}_2 \cdot r_i r_j \right), \quad i \neq j. \]  

(109)

The Jacobian matrix \( J \) is:

\[ J = \kappa^{-\frac{1}{q+1}} \cdot \|r\|^{-\frac{q}{q+1}}_2 \left( I - \frac{q}{q+1} \cdot \|r\|^{-2}_2 \cdot rr^T \right), \]  

(110)

with the determinant given by

\[ \det(J) = \kappa^{-\frac{d}{q+1}} \cdot \|r\|^{-\frac{qd}{q+1}}_2 \cdot \det \left( I - \frac{q}{q+1} \cdot \|r\|^{-2}_2 \cdot rr^T \right) \]

\[ = \kappa^{-\frac{d}{q+1}} \cdot \|r\|^{-\frac{qd}{q+1}}_2 \left( 1 - \frac{q}{q+1} \cdot \|r\|^{-2}_2 \cdot rr^T \right) \]  

(111)

\[ = \frac{1}{q+1} \cdot \kappa^{-\frac{d}{q+1}} \cdot \|r\|^{-\frac{qd}{q+1}}_2. \]

We obtain the expression for the pdf as

\[ p(r) = \frac{1}{q+1} \cdot \kappa^{-\frac{d}{q+1}} \cdot \|r\|^{-\frac{qd}{q+1}}_2 \cdot (2\pi)^{-\frac{d}{2}} \cdot \exp \left( -\frac{1}{2} \kappa^{-\frac{2}{q+1}} \cdot \|r\|^{-\frac{2}{q+1}}_2 \right). \]  

(112)

### C Proofs for Complex Gaussian Measurement Model

#### C.1 Proof of Lemma 3

**Proof.** As done in Section A.1, we first establish an upper bound on \( \|S - 2xx^*\| \). We have:

\[ \|S - 2xx^*\| \leq \left| \frac{\|x\|_2^2}{m} \sum_{i=1}^{m} (\bar{A}_{i,11} A_{i,11} - 2) \right| + \left| \frac{1}{m} \sum_{i=1}^{m} \|x\|_2^2 \bar{A}_{i,11} \bar{A}_i \right|. \]  

(113)
The real and imaginary coefficients of $A_i$ are iid standard Gaussian. The first term of (113) is a summation of $2m$ iid sub-exponential random variables. Using the Bernstein-type inequality [30, Proposition 5.16], if $m > Cn$, we have:

$$
\Pr \left( \left\| x \right\|_2^2 \sum_{i=1}^m (A_{i,11} A_{i,11} - 2) \geq \frac{\delta}{2} \left\| x \right\|_2^2 \right) \leq 2 \exp \left( -m \cdot \min \left\{ \frac{c_\delta^2}{8K^2}, \frac{c_\delta}{2K} \right\} \right) \leq 2 \exp \left( -Cn \cdot \min \left\{ \frac{c_\delta^2}{8K^2}, \frac{c_\delta}{2K} \right\} \right),
$$

(114)

where $c > 0$ is some constant that is different from $C$, and $K$ is the sub-exponential norm of $A_{i,11}^2 - 1$ and/or $A_{i,11}^2 - 1$ [30, Definition 5.13]. Here $A_{i,11}^2 - 1$ and $A_{i,11}^2 - 1$ are both the centered Chi-square random variable with 1 degree of freedom.

As done in Section A.1, finding an upper bound on the second term of (113) is equivalent to finding an upper bound on the following (115) for all $u, v \in \mathbb{C}^n$ satisfying $\|u\|_2 = 1$, $\|v\|_2 = 1$:

$$
\left\| u^* \cdot \frac{1}{m} \sum_{i=1}^m \left\| x \right\|_2^2 A_{i,11} \cdot v \right\| \leq \frac{1}{m} \sum_{i=1}^m \left\| x \right\|_2^2 A_{i,11} \cdot G_i, \tag{115}
$$

(115)

where $G_i = \sum_{(r,c)\neq(1,1)} \pi_{r,v_c} \cdot A_{i,rc}$. Let $B_{rc} = \pi_{r,v_c}$, we have:

$$
G_i = \sum_{(r,c)\neq(1,1)} B_{rc}(R) A_{i,rc}^{(R)} - B_{rc}(I) A_{i,rc}^{(I)} + \left( B_{rc}(R) A_{i,rc}^{(R)} + B_{rc}(I) A_{i,rc}^{(I)} \right) \cdot j. \tag{116}
$$

Given $u, v$, the real coefficient of $G_i$ is the summation of $n^2 - 1$ iid Gaussian $\mathcal{N}(0, |\pi_{r,v_c}|^2)$. Hence $\text{Re}(G_i)$ follows $\mathcal{N}(0, 1 - \text{Im}(\pi_{r,v_c})^2)$, its imaginary coefficient $\text{Im}(G_i)$ also independently follows the same distribution. We then have:

$$
\left\| \frac{x}{m} \sum_{i=1}^m A_{i,11} G_i \right\| \leq \left\| \frac{x}{m} \sum_{i=1}^m \left( A_{i,11} G_i^{(R)} + A_{i,11} G_i^{(I)} \right) \right\| + \left\| \frac{x}{m} \sum_{i=1}^m \left( A_{i,11} G_i^{(R)} - A_{i,11} G_i^{(I)} \right) \right\|. \tag{117}
$$

Take the first term of (117) for example, we have:

$$
A_{i,11} G_i^{(R)} + A_{i,11} G_i^{(I)} = \sum_{(r,c)\neq(1,1)} B_{rc}^{(R)} A_{i,rc}^{(R)} A_{i,rc}^{(R)} - B_{rc}^{(I)} A_{i,rc}^{(I)} A_{i,rc}^{(I)} + B_{rc}^{(R)} A_{i,rc}^{(R)} A_{i,rc}^{(I)} + B_{rc}^{(I)} A_{i,rc}^{(I)} A_{i,rc}^{(R)}. \tag{53 revisited}
$$

We further have:

$$
A_{i,11}^{(R)} A_{i,rc}^{(R)} = \frac{1}{2} \left\{ \left( \frac{A_{i,11}^{(R)} + A_{i,rc}^{(R)}}{\sqrt{2}} \right)^2 - 1 \right\} - \frac{1}{2} \left\{ \left( \frac{A_{i,11}^{(R)} - A_{i,rc}^{(R)}}{\sqrt{2}} \right)^2 - 1 \right\}. \tag{118}
$$

It is easy to verify that $\frac{A_{i,11}^{(R)} + A_{i,rc}^{(R)}}{\sqrt{2}}$ and $\frac{A_{i,11}^{(R)} - A_{i,rc}^{(R)}}{\sqrt{2}}$ are both standard Gaussian variables, and they are also uncorrelated. Since uncorrelated standard Gaussian variables are independent, $A_{i,11}^{(R)} A_{i,rc}^{(R)}$
is a linear combination of two independent centered sub-exponential random variables. We can repeat this process for every term in (53).

Using the Bernstein inequality, if \( m > Cn \), we have:

\[
\Pr \left( \frac{\|x\|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_{i}^{(R)} + A_{i,11}^{(I)} G_{i}^{(I)} \right) \geq \frac{\delta}{4} \|x\|_2^2 \right) \leq 2 \exp \left( -Cn \cdot \min \left\{ \frac{c\delta^2}{16K^2}, \frac{c\delta}{2K} \right\} \right)
\]

(119)

\[
\Pr \left( \frac{\|x\|_2^2}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} G_{i}^{(I)} - A_{i,11}^{(I)} G_{i}^{(R)} \right) \geq \frac{\delta}{4} \|x\|_2^2 \right) \leq 2 \exp \left( -Cn \cdot \min \left\{ \frac{c\delta^2}{16K^2}, \frac{c\delta}{2K} \right\} \right).
\]

(120)

- Combining (114), (119), (120), we can get:

\[
\Pr (\|S - 2x^*x\| \leq \delta) \geq 1 - 6 \exp \left( -Cn \cdot \min \left\{ \frac{c\delta^2}{16K^2}, \frac{c\delta}{2K} \right\} \right)
\]

(121)

- Following the same procedure as in Section [A.1], we can get that the following

\[
dist^2 \left( \sqrt{K}v_0, x \right) \leq \frac{11}{4} \delta \|x\|_2^2.
\]

(122)

holds with probability at least \( 1 - 6 \exp \left( -Cn \cdot C_1'(\delta) \right) \).

\[\square\]

C.2 Proof of Lemma 4

Proof. As done in Section [A.2], we need to find an upper bound on the following:

\[
\left\| \frac{1}{m} \sum_{i=1}^{m} \tilde{p}^* A_i q A_i - 2\tilde{q} \tilde{p}^* \right\| \leq \left| \text{Re} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2r_1 e^{j\phi_1} \right) \right| + \left| \text{Re} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,21} - 2r_2 e^{j\phi_2} \right) \right|
\]

\[
+ \left| \text{Im} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2r_1 e^{j\phi_1} \right) \right| + \left| \text{Im} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,21} - 2r_2 e^{j\phi_2} \right) \right| + \left\| \frac{1}{m} \sum_{i=1}^{m} b_i A_i \right\| .
\]

(70) revisited

- For the first term of (70), we have:

\[
\left| \text{Re} \left( \frac{1}{m} \sum_{i=1}^{m} b_i A_{i,11} - 2r_1 e^{j\phi_1} \right) \right| \leq \frac{r_1 \cos \phi_1}{m} \sum_{i=1}^{m} \left( A_{i,11}^{(R)} - 1 \right) + \left( A_{i,11}^{(I)} - 1 \right)
\]

\[
+ \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 A_{i,21} A_{i,11}^{(R)} + r_2 \cos \phi_2 A_{i,21} A_{i,11}^{(I)} - r_2 \sin \phi_2 A_{i,21} A_{i,11}^{(I)} + r_2 \sin \phi_2 A_{i,21} A_{i,11}^{(R)} .
\]

(123)
From [118], we can see that the two terms of [123] are both summations of iid centered sub-exponential random variables. Using the Bernstein inequality [30, Proposition 5.16], if \( m > Cn \), we have:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} (|A_{i,11}|^2 - 2) \right| \geq \frac{\nu}{12} \right) \leq 2 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{288K^2}, \frac{cv}{12K} \} \right)
\]

\[
\leq 2 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{144K^2}, \frac{cv}{6K} \} \right).
\]

Similarly we can get that:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} r_2 \cos \phi_2 \left( A_{i,21}^{(R)} A_{i,11}^{(R)} + A_{i,21}^{(I)} A_{i,11}^{(I)} \right) - r_2 \sin \phi_2 \left( A_{i,21}^{(R)} A_{i,11}^{(I)} - A_{i,21}^{(I)} A_{i,11}^{(R)} \right) \right| \geq \frac{\nu}{12} \right) \leq 2 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{288K^2}, \frac{cv}{12K} \} \right).
\]

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} \left( \left( b_i^{(R)} G_i^{(R)} - b_i^{(I)} G_i^{(I)} \right) \right) \right| \geq \frac{\nu}{6} \right) \leq 2 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{36K^2}, \frac{cv}{3K} \} \right).
\]

Similarly we can get that:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} \left( \left( b_i^{(R)} G_i^{(I)} + b_i^{(I)} G_i^{(R)} \right) \right) \right| \geq \frac{\nu}{6} \right) \leq 2 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{36K^2}, \frac{cv}{3K} \} \right).
\]

Combining all the probability bounds, we can get:

\[
\Pr \left( \left| \frac{1}{m} \sum_{i=1}^{m} \tilde{p}^* \tilde{A}_i \tilde{q} \tilde{A}_i \tilde{v} - 2 \hat{u}^* \hat{q} \hat{p} \hat{v} \right| \leq \nu \right) \geq 1 - 20 \exp \left( -Cn \cdot \min \{ \frac{cv^2}{288K^2}, \frac{cv}{12K} \} \right).
\]

\[
= 1 - 20 \exp \left( -Cn \cdot C_2'(\nu) \right).
\]

\[\square\]
C.3 Proof of Corollary 1

Proof. The proofs are basically the same as done in Appendix A.3. Using Lemmas 3 and 4, we can show that \((41)\) holds with probability at least

\[
1 - 6 \exp \left( -Cn \cdot \min \left\{ \frac{c\delta^2}{16K^2}, \frac{c\delta}{2K} \right\} \right) - 20 \exp \left( -Cn \cdot \min \left\{ \frac{cv^2}{288K^2}, \frac{cv}{12K} \right\} \right) \\
\geq 1 - 26 \exp \left( -Cn \cdot \min \left\{ \frac{c\delta^2}{16K^2}, \frac{c\delta}{2K}, \frac{cv^2}{288K^2}, \frac{cv}{12K} \right\} \right) \\
= 1 - 26 \exp \left( -Cn \cdot C_3' (\delta, \nu) \right) .
\]