On the Sample Complexity and Optimization Landscape for Quadratic Feasibility Problems

Parth K. Thaker, Gautam Dasarathy, and Angelia Nedich
Electrical, Computer and Energy Engineering
Arizona State University, Tempe, USA
Email: {pkthaker, gautamd, angelia.nedich}@asu.edu

***Abstract*—We consider the problem of recovering a complex vector $\mathbf{x} \in \mathbb{C}^n$ from m quadratic measurements $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}_{i=1}^m$. This problem, known as quadratic feasibility, encompasses the well known phase retrieval problem, along with being applicable in a range of areas including power system state estimation and x-ray crystallography. In general, not only is the the quadratic feasibility problem NP-hard to solve, but it may in fact be unidentifiable. In this paper, we establish conditions under which this problem becomes identifiable, and further prove stronger isometry properties in the case when the matrices $\{A_i\}_{i=1}^m$ are Hermitian matrices sampled from a complex Gaussian distribution. Moreover, we explore a nonconvex formulation of this problem, and establish salient features of the associated optimization landscape that enables gradient algorithms with an arbitrary initialization to converge to a *globally optimal* point with high probability. Our results also reveal sample complexity requirements for successfully identifying the feasible solution in these contexts.**

I. Introduction

Finding solutions to systems of quadratic equations is an important problem with a wide range of applications. This arises in areas such as power system state estimation [1], phase retrieval [2]–[5], x-ray crystallography [6], the turnpike problem [7], and unlabeled distance geometry problems [8], [9] among others. Such problems are often reduced to a *quadratic feasibility problem*, where one is concerned with finding a *feasible* vector $\mathbf{x}$ that conforms to a set of quadratic observations of the form $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}_{i=1}^m$ with respect to a set of measurement matrices $\{A_i\}_{i=1}^m$. In other words, this can be stated as

$$\text{find } \mathbf{x} \quad \text{such that } \langle A_i\mathbf{x}, \mathbf{x}\rangle = c_i, \quad \forall i = 1, 2, \ldots, m. \qquad \text{(P1)}$$

The quadratic feasibility problem is an instance of quadratically constrained quadratic programs (QCQPs) [10], which has enjoyed a long and rich research history dating back to 1941 [11]. Given their broad applicability to critical problems, research in QCQPs continues to be of active interest [10], [12]–[14]. Unfortunately, it is known that solving QCQPs is an NP-hard problem [15]. This combined with the lack of tractable duality properties [16] has made it hard to establish a sound theoretical framework for understanding the solutions and computing them. However, an extremely productive line of research has instead considered subclasses of QCQPs that are both practically relevant and can be analyzed. In this paper, we take a similar approach and identify an important subclass of QCQPs that have a broad range of applications. In particular, we analyze the quadratic feasibility problem, and establish conditions under which such problems are identifiable, and then show that these conditions are in-fact sufficient for the efficient computation of their solutions.

We start by considering quadratic functions $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}_{i=1}^m$, where $\mathbf{x} \in \mathbb{C}^n$ and $A_i \in \mathbb{C}^{n\times n}$ are Hermitian. We focus on their ability to generate injective maps up-to a phase factor (note that the quadratic functions $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}_{i=1}^m$ are invariant to phase shifts). We establish a relationship between injectivity and isometry and show that, in real world scenarios, it is not difficult for a set of quadratic measurements $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}_{i=1}^m$ to possess such an isometry property by establishing that this holds with very high probability when the matrices $\{A_i\}_{i=1}^m$ are complex Gaussian random matrices.

After establishing injectivity (and hence the identifiability) of the problem, we consider the question of computationally tractable approaches to actually find a feasible solution. Toward this end, a natural approach is to optimize the appropriate ℓ_2-loss. Unfortunately, this turns out to be a nonconvex problem, that is NP-hard to solve in general. However, we show that under the same conditions required to establish injectivity, the landscape of this optimization problem is well-behaved. This allows us to establish that any gradient based algorithm with an arbitrary initialization can recover the a globally optimal solution almost surely.

The rest of the paper is organized as follows [gd:please complete]. In section II highlights the main results of this work. We list out related literature out there in section III. In section IV we establish and analyze isometric properties of the mapping $\{\langle A_i\mathbf{x}, \mathbf{x}\rangle\}$ when the matrices are complex Gaussian random matrices. Section V casts the problem into a loss minimization framework (suitable for algorithmic approaches) and establishes some landscape properties of this loss function.

II. Main Results

A. Notation

Before we state the main results of the paper, we will introduce some notation that will be used in the sequel. For any $r \in \mathbb{N}$, we write $[r]$ to denote the set $\{1, 2, \ldots, r\}$. We let $\mathbb{C}^n$ and $\mathbb{R}^n$ denote the n-dimensional complex and real vector spaces respectively. Unless otherwise stated, bold letters such as $\mathbf{x}$ indicate vectors in $\mathbb{C}^n$; $\mathbf{x}_\mathbb{R}$ and $\mathbf{x}_\mathbb{C}$ denote the real and the imaginary part of the vector $\mathbf{x}$ respectively. We denote complex conjugate of $\mathbf{x}$ by $\bar{\mathbf{x}}$. Capital letters such as X

denote matrices in $\mathbb{C}^{n\times n}$. The use of i (without serif) indicates the complex square root of -1. We will use i to indicate an indexing variable. We let $\mathcal{S}^{a,b}(\mathbb{R}^{n\times n})$ denote the set of all matrices $X \in \mathbb{R}^{n\times n}$ having a non-negative eigenvalues and b negative eigenvalues, where $a+b=n$. The set $\mathbf{H}_n(\mathbb{C})$ denotes the set of all $n\times n$ Hermitian matrices. We write $A^\top$ and A^* to denote, respectively, the transpose and the Hermitian transpose (transpose conjugate) of a matrix A. We use $\langle \cdot, \cdot \rangle$ and to denote the inner vector product in the complex space. The symmetric outer product, denoted by $[[\cdot,\cdot]]$, is defined as

$$[[\mathbf{u},\mathbf{v}]] = \mathbf{u}\mathbf{v}^* + \mathbf{v}\mathbf{u}^*$$

Finally, we will let $\sim$ denote the following equivalence relation on $\mathbb{C}^n$: $\mathbf{x} \sim \mathbf{y}$ if and only if $\mathbf{x} = c\mathbf{y}$ for some $c \in \mathbb{C}$ with $|c| = 1$. We will write $\widehat{\mathbb{C}}^n \triangleq \mathbb{C}^n/\sim$ to denote the associated quotient space. Given a set of matrices $\mathcal{A} = \{A_i\}_{i=1}^m \subset \mathbf{H}_n(\mathbb{C})$, we will let $\mathcal{M}_\mathcal{A}$ denote the following mapping from $\widehat{\mathbb{C}}^n \to \mathbb{C}^m$:

$$\mathcal{M}_\mathcal{A}(\mathbf{x}) = (\langle A_1\mathbf{x},\mathbf{x}\rangle, \langle A_2\mathbf{x},\mathbf{x}\rangle, \ldots, \langle A_m\mathbf{x},\mathbf{x}\rangle). \tag{1}$$

While $\mathcal{M}_\mathcal{A}$ technically operates on the equivalence classes in $\widehat{\mathbb{C}}^n$, we will abuse the notation slightly and think of $\mathcal{M}_\mathcal{A}$ as operating on the elements of $\mathbb{C}^n$.

B. Main Results

We consider the quadratic feasibility problem (P1) with the complex decision vector, i.e., $\mathbf{x} \in \mathbb{C}^n$, Hermitian matrices $A_i \in \mathbb{C}^{n\times n}$ and real numbers $c_i \in \mathbb{R}$ for all $i \in [m]$. In order to understand the properties of this problem, we need a coherent distance metric that is analytically tractable. Toward this, we will use the following:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_1 \quad \text{for any } \mathbf{x},\mathbf{y} \in \mathbb{C}^n. \tag{2}$$

Notice that this distance metric is invariant under phase shifts, and has been used to prove crucial robustness results for the phase retrieval problem (in $\mathbb{R}^n$); see e.g., [3], [17], [18] .

Our first main result (Theorem 1) is that when the matrices A_i are chosen from a complex Gaussian distribution, then the mapping $\mathcal{M}_\mathcal{A}$ is a near-isometry. We will provide a sketch of the statement here

Theorem 1 (sketch). *Let $\mathcal{A} = \{A_i\}_{i=1}^m$ be a set of complex Gaussian random matrices. Then, with high probability, we have the following for $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$*

$$\beta d(\mathbf{x},\mathbf{y}) \geq \|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2 \geq \alpha d(\mathbf{x},\mathbf{y}),$$

provided the number m of measurements satisfies $m > Cn$, for some constants $\alpha, \beta, C > 0$.

In other words, $\mathcal{M}_\mathcal{A}$ nearly preserves distances with respect to the distance measure defined in (2). This of course implies that if $\mathbf{x}$ and $\mathbf{y}$ are distinct, then the corresponding measurements are also necessarily distinct – that is, the measurement operation defined by $\mathcal{M}_\mathcal{A}$ is *identifiable*. The formal statement is presented in Theorem 1, and the full proof is available in [gd:cite relevant portion of full manuscript].

Given that we have established that (P1) has a uniquely identifiable solution (upto a phase ambiguity), we next turn our attention to finding a feasible solution in a computationally efficient manner. Toward this, one may naturally consider recasting the quadratic feasibility problem as a quadratic optimization problem and optimizing the following ℓ_2 loss:

$$\min_{\mathbf{x}\in\mathbb{C}^n} \frac{1}{m}\sum_{i=1}^m |\langle A_i\mathbf{x},\mathbf{x}\rangle - c_i|^2 . \tag{P2}$$

Unfortunately, this optimization problem is non-convex, and in general one may not expect an iterative procedure like gradient descent to converge to a global optimum. However, our next main result shows that with high probability, the optimization landscape of (P2) is in fact amenable to iterative gradient based optimization techniques! Moreover, we are able to demonstrate this result under the same conditions required for the problem to even be identifiable – namely, the measurement matrices are drawn from the complex Gaussian distribution. We now provide a sketch of our second main result.

Theorem 2 (sketch). *Let $\mathcal{A} = \{A_i\}_{i=1}^m$ be a set of complex Gaussian random matrices and let $\mathbf{x}^*$ be a global optimizer of* (P2). *Then, with high probability, the following holds:*

1) *$\mathcal{M}_\mathcal{A}(\mathbf{w}) = \mathcal{M}_\mathcal{A}(\mathbf{x}^*)$ for all local minima $\mathbf{w}$ of* (P2).
2) *The objective function in* (P2) *has the strict saddle property (see Definition 2 for a formal definition)*

The formal statement of this theorem appears as Theorem 2 in Section V. Theorem 2 states that provided the measurement matrices are Gaussian, with high probability, the optimization problem (P2) has no spurious local minima and any saddle point of the function is *strict*, i.e., has a strict negative curvature.

Finally, we use the above established properties of the loss landscape to conclude that one can readily optimize (P2) by applying a gradient based algorithm since such an algorithm is unlikely to converge to a saddle point.

III. Related work

QCQPs have enjoyed a lot of attention over the last century. However, due to the limitation of the duality properties of QCQPs [19], a significant fraction of research has focused predominantly on heuristic approaches to their solution [20]–[22]. Recently, an ADMM-based method has been proposed in [23] with an asymptotic convergence result based on the duality properties QCQPs. Our results in this paper bring new insights to this area by analyzing a subset of QCQPs, namely, the quadratic feasibility problems.

The quadratic feasibility problem (P1) arises in many applications, including phase retrieval [2] and power system state estimation [1]. Phase retrieval in and of itself finds applications in a wide variety of fields such as imaging, optics, quantum tomography, audio signal processing with a wide literature, including [2], [4], [5], [24]. In [3], an approximate ℓ_1 isometry property was established for the phase retrieval problem, but the bounds therein are not strong enough to provide RIP-like guarantees. In this paper, we improve these bounds to establish isometry results for a large class of problems and provide RIP-type bounds for the case when $\{A_i\}_{i=1}^m$ are complex Gaussian random matrices.

A feasibility problem is often cast as a minimization problem with a suitably chosen loss function. Even with a nonconvex objective, gradient based methods have proven to work for phase retrieval [2], [5], [24], matrix factorization [25], [26] and robust linear regression [27]. The work in [28] has established landscape properties for the phase retrieval problem, which sheds light on the success of gradient based methods in solving the problem. In this work, we extend these results to a wider class of problems along with additional insights into the problem properties. In [29] it was shown that many nonconvex loss functions have specific landscape properties, which allow gradient based algorithm to recover a globally optimal solution without any additional information. One unfortunately cannot readily transport those results to our setting, not in small part due to the significant differences between the real and complex vector spaces. For instance, a quadratic feasibility problem in $\mathbb{R}^n$ has only two isolated local minima, but the same problem has a continuum of minima in $\mathbb{C}^n$.

[30] provided lower bounds on the minimum number of independent measurements required for successful recovery for the quadratic feasibility problem. More recently [31] showed that the quadratic feasibility problem can be solved, with high probability, by gradient descent provided a good initialization is used. In contrast, the current work takes a parallel track by analyzing the landscape of the associated ℓ_2-loss function. In particular, we prove that, for the ℓ_2-loss function, all local minima are global minima, and all saddle points are strict saddle points. Thus, our results enable a large spectrum of gradient based algorithms with (essentially) arbitrary initialization to recover the solution for the quadratic feasibility problem.

IV. Properties of the Quadratic Mapping

As a first step towards establishing our main results, we start by characterizing when the quadratic mapping $\mathcal{M}_{\mathcal{A}}$ defined in (1) is in fact an injective mapping. Notice that the injectivity of the mapping is equivalent to the problem being identifiable (and hence solvable). It is also worth noting that in the context of the phase retrieval problem, when $\mathcal{M}_{\mathcal{A}}$ is injective, it is said to possess the phase retrievability property [30].

Lemma 1. *The following statements are equivalent:*

1) *The nonlinear map $\mathcal{M}_{\mathcal{A}} : \widehat{\mathbb{C}}^n \to \mathbb{C}^m$ is injective.*
2) *There exist constants $\alpha, \beta > 0$ such that $\forall \mathbf{u}, \mathbf{v} \in \mathbb{C}^n$,*

$$\beta \|[[\mathbf{u},\mathbf{v}]]\|_1^2 \geq \sum_{i=1}^m |\langle A_i, [[\mathbf{u},\mathbf{v}]]\rangle|^2 \geq \alpha \|[[\mathbf{u},\mathbf{v}]]\|_1^2. \quad (3)$$

We refer the reader to Appendix VI-A for a detailed proof. Lemma 1 characterizes the injectivity of $\mathcal{M}_{\mathcal{A}}$ in terms of the action of the A_i matrices on the outer product $[[\mathbf{u},\mathbf{v}]]$. In particular, this says that the mapping $\mathcal{M}_{\mathcal{A}}$ is injective if and only if the matrices A_i do not distort the outer-product $[[\mathbf{u},\mathbf{v}]]$ too much. We use this characterization to obtain a more tractable condition that allows us to establish the injectivity of $\mathcal{M}_{\mathcal{A}}$ in Lemma 2 in the case of Gaussian measurement matrices.

Our tractable condition is based on what we call the *stability* of the mapping. Intuitively speaking, we say that $\mathcal{M}_{\mathcal{A}}$ is stable if the following holds: given an $\epsilon > 0$, there is a $\delta > 0$ such that whenever two complex vectors $\mathbf{x}, \mathbf{y} \in \widehat{\mathbb{C}}^n$ satisfy $d(\mathbf{x},\mathbf{y}) \geq \epsilon$, we should have $\|\mathcal{M}_{\mathcal{A}}(\mathbf{x}) - \mathcal{M}_{\mathcal{A}}(\mathbf{y})\| \geq \delta$. Such stability properties have been considered in the specific context of phase retrieval; see e.g., [3], [4], [32], [33]. We now formally define our notion of stability.

Definition 1 ((α,β)-stability)**.** *Given constants $0 < \alpha \leq \beta$, the mapping $\mathcal{M}_{\mathcal{A}}$ (see (1)) based on the set of matrices $\mathcal{A} = \{A_i\}_{i=1}^m$ is said to be (α,β)-stable if it satisfies the following for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$:*

$$\alpha d(\mathbf{x}_1,\mathbf{x}_2) \leq \|\mathcal{M}_{\mathcal{A}}(\mathbf{x}_1) - \mathcal{M}_{\mathcal{A}}(\mathbf{x}_2)\|_2 \leq \beta d(\mathbf{x}_1,\mathbf{x}_2), \quad (4)$$

where d is a metric on $\mathbb{C}^n$.

The value of the constants α, β crucially depend on the choice of the distance metric $d(\cdot,\cdot)$; we will take the metric to be as in (2) in the sequel. The notion of (α,β)-stability is stronger than the one considered for phase retrieval in [4], [32], since it additionally establishes an upper bound to the difference of the measurements. The constants α, β can also be thought of as a "condition number", thereby allowing one to assess the quality of the map; higher the ratio between α and β, the better the ability of the mapping to distinguish two different inputs $\mathbf{x}_1 \nsim \mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^n$.

AN HERE

Armed with this definition of stability, we now show that $\mathcal{M}_{\mathcal{A}}$ is injective if and only if the map is stable.

Lemma 2. *The mapping $\mathcal{M}_{\mathcal{A}}$ is injective iff it is (α,β)-stable for some constants $0 < \alpha \leq \beta$.*

Please refer to Appendix VI-B for a proof of the lemma. This result demonstrates the usefulness of both our choice of the metric $d(\cdot,\cdot)$ from (2) and of our definition of stability (Definition 1). Lemma 2, as we will see in what follows, allows one to assess the conditions under which the measurement model implied by the mapping $\mathcal{M}_{\mathcal{A}}$ is *identifiable*.

Next, we turn our attention to the question of when one can establish the above stability condition, and thereby establish the identifiability of the underlying measurement model. In this, we take our cues from the compressive sensing and phase retrieval literature, and suppose that the set of Hermitian matrices $\mathcal{A} = \{A_i\}_{i=1}^m$ are sampled from complex Gaussian distribution. To be more precise, we suppose that each entry in the upper triangle (including the diagonal) of A_i is drawn independently from $\mathcal{N}(0,1)$. The remaining entries are determined by the fact that the A_i's are hermitian.

We first make the observation that when $\mathcal{A}$ is chosen as above, then for any $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, we have that for all $i \in [m]$, $\mathbb{E}\left[|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2\right] = d(\mathbf{x},\mathbf{y})^2$ (see [gd:cite full paper] for more details). Our next result shows that these quantities are actually concentrated about their expected values.

Lemma 3. *Let $\mathcal{A} = \{A_i\}_{i=1}^m$ be a set of complex hermitian gaussian random matrices for the measurement model given by* (P1). *Then, given $\epsilon > 0$, there are constants $c, d > 0$ such that*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \frac{1}{m}|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2 - d(\mathbf{x},\mathbf{y})^2\right| \geq \epsilon d(\mathbf{x},\mathbf{y})^2\right) \leq de^{-cm\epsilon} \tag{5}$$

Please see [gd:cite full paper] for a proof of this result. Notice that this concentration result only holds for a fixed pair of vectors $\mathbf{x}, \mathbf{y}$. We however need our stability result to be uniform over all pairs of vectors in $\mathbb{C}^n$. For this, we will next adapt a standard covering argument as follows. Let S^{n-1} denote the $n-$dimensional unit sphere, and let $\|A\|$ denote the spectral norm of the matrix A, i.e., $\|A\| = \sup_{\mathbf{x}\in S^{n-1}}\langle A\mathbf{x}, \mathbf{x}\rangle$. We have the following "covering result"

Lemma 4. *Fix $\delta > 0$ and let $\mathcal{N}_\delta \subset S^{n-1}$ represent the smallest collection of balls of radius δ whose union covers S^{n-1}. Then, for any matrix $A \in \mathbb{C}^{n\times n}$, the following holds,*

$$\begin{aligned}(1-2\delta) &\sup_{\mathbf{x}_1,\mathbf{x}_2\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \\ &\leq \sup_{\mathbf{x}_1,\mathbf{x}_2\in \mathcal{N}_\delta} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \\ &\leq (1+2\delta) \sup_{\mathbf{x}_1,\mathbf{x}_2\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle|\end{aligned}$$

We refer the reader to [gd:cite appropriate locations] for a proof of this lemma. This argument is not new and has found application in several results; see e.g., [31], [34], [35].

We are finally ready to state and prove our first main result.

Theorem 1. *Let $\mathcal{A} = \{A_i\}$ be the set of complex Gaussian random matrices. Given a $\xi > 0$, there exist constants $C, c_0, d_0 > 0$ and $\beta > \alpha > 0$ such that with probability at least $1-\xi$, the following holds*

$$\alpha d(\mathbf{x},\mathbf{y}) \leq \|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2 \leq \beta d(\mathbf{x},\mathbf{y}),$$

provided the number of measurements satisfies $m > Cn$.

Proof. Consider $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Notice that in the case when $d(\mathbf{x},\mathbf{y}) = 0$, then the theorem follows readily. Therefore, in the sequel, we will suppose that $d(\mathbf{x},\mathbf{y}) > 0$.

From Lemma 3, for a fixed $\epsilon > 0$ we have the following

$$\mathbb{P}\left(\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} - 1\right| \geq \epsilon\right) \leq de^{-cm\epsilon} \tag{6}$$

Now, as noted in [34], we know that for any $\delta > 0$, we have the following upper bound on the size of the covering: $|\mathcal{N}_\delta| \leq \left(\frac{12}{\delta}\right)^n$.

Therefore, for a fixed $\epsilon, \delta > 0$, from Lemma 3 and the union bound, we have that

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} - 1\right| > \epsilon\right) \leq de^{-cm\epsilon}\left(\frac{12}{\delta}\right)^n$$

where d, c are the same constants as in Lemma 3. This implies that

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} - 1\right| \leq \epsilon\right) \geq 1 - de^{-cm\epsilon}\left(\frac{12}{\delta}\right)^n.$$

Now, observe that

$$\begin{aligned}&\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} - 1\right| \\ &\geq \sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2}\right| - 1\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\left|\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2}\right| - 1 \leq \epsilon\right) \geq 1 - d_1e^{-c_1m\epsilon}\left(\frac{12}{\delta}\right)^n$$

From covering argument in lemma 4, and,

$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta} d(\mathbf{x},\mathbf{y}) \leq (1+2\delta)\sup_{\mathbf{x},\mathbf{y}\in S^{n-1}} d(\mathbf{x},\mathbf{y})$$

note that,

$$\begin{aligned}&\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}_\delta}\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} \\ &\geq \sup_{\mathbf{x},\mathbf{y}\in S^{n-1}}\sum_{i=1}^n \frac{(1-2\delta)^2}{(1+2\delta)^2 m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2}\end{aligned}$$

Thus we can conclude that,

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in S^{n-1}}\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} \leq \frac{(1+2\delta)^2(1+\epsilon)}{(1-2\delta)^2}\right) \geq 1 - de^{-cm\epsilon}\left(\frac{12}{\delta}\right)^n$$

Similarly, we can prove,

$$\mathbb{P}\left(\inf_{\mathbf{x},\mathbf{y}\in S^{n-1}}\sum_{i=1}^n \frac{1}{m}\frac{|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{d(\mathbf{x},\mathbf{y})^2} \leq \frac{(1-2\delta)^2(1-\epsilon)}{(1+2\delta)^2}\right) \geq 1 - de^{-cm\epsilon}\left(\frac{12}{\delta}\right)^n$$

Hence taking $m \geq Cn$, such that $C > \frac{\log 12d - \log\delta\xi}{c\epsilon}$, we can finally state with probability $\geq 1-\xi$ that,

$$\beta d(\mathbf{x},\mathbf{y}) \geq \|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2 \geq \alpha d(\mathbf{x},\mathbf{y})$$

where α, β are given by,

$$\alpha := \frac{\left((1-2\delta)^2(1-\epsilon\right)}{(1+2\delta)^2}, \quad \beta := \frac{\left((1+2\delta)^2(1+\epsilon\right)}{(1-2\delta)^2}.$$

For obtaining the Theorem statement to be true at least with probability $1-\xi$, the number of measurements required is $m > \frac{\log 12d - \log \delta\xi}{c\epsilon} n$ Notice that we essentially have a choice of the values of δ and ϵ. The closer these are to 0, the stronger the stability result. However, this also implies the more the number of measurements that are needed. $\square$

V. Non-convex loss reformulation and the Optimization Landscape

In the previous section, we established conditions under which the mapping $\mathcal{M}_{\mathcal{A}}$ represents an identifiable measurement model. Our next goal is to actually find a feasible solution (upto a phase shift) in a computationally efficient manner. Toward this end, a natural approach would be to consider minimizing the appropriate ℓ_2 loss as follows

$$\min_{\mathbf{x}} f(\mathbf{x}) = \min_{\mathbf{x}} \frac{1}{m} \sum_{i=1}^{m} \left|\langle A_i \mathbf{x}, \mathbf{x}\rangle - c_i\right|^2. \tag{P2}$$

Unfortunately, it is not hard to see that this optimization problem is non-convex, and in general one might not gradient-based aproaches to converge to a global optimum. On the other hand, nonconvex optimization has received considerable attention lately. In particular, methods like SGD and other gradient based methods have been shows to be astonishingly successful in converging to global minimas in many nonconvex problems [36]–[38]. Arguably, the reason for this is that the optimization landscape of these (somewhat well-behaved) non-convex problems enjoy some advantageous properties which are crucial in the empirical success of gradient based methods [28], [39], [40]. The work in [28] proves that the ℓ_2 loss function for the phase retrieval problem enjoys properties like local minima being global minima and saddle points have a strict negative curvature. This work adds to this rich body of work by demonstrating similarly advantageous properties of the optimization landscape for the quadratic feasibility problem. Before we state our next main result, we should also observe that since the ℓ_2 loss function is not complex differentiable, it is challenging to carry out analysis on the optimization problem above directly. In what follows, we instead use techniques from Wirtinger calculus [2]. Our first step is to define the notion of a strict saddle function.

Definition 2. *Let β, γ, ζ be positive constants. A function f is said to be (β, ζ, γ)-strict saddle if for any $\mathbf{x} \in \mathbb{C}^n$, atleast one of the following is true:*

1) $\|\nabla f(\mathbf{x})\| \geq \beta$
2) $\exists \mathbf{z} \in \mathbb{C}^n$ *such that* $\langle \nabla^2 f(\mathbf{x})\mathbf{z}, \mathbf{z}\rangle \leq -\zeta$
3) $\mathbf{x}$ *is γ-close to a local minima i.e. there is a $\mathbf{w}$ such that $\nabla f(\mathbf{w}) = 0$, $\nabla^2 f(\mathbf{w}) \succeq 0$ and $d(\mathbf{x}, \mathbf{w}) < \gamma$.*

Intuitively, this implies that every $\mathbf{x} \in \mathbb{C}^n$ either violates optimality (condition 1 and 2) or is close to a local optimum. A line of recent work [27], [41], [42] has explored the efficacy of gradient based methods in finding a local optimum of functions satisfying Definition 2.

Our next main result will demonstrate that the function f in (P2) is in fact a strict saddle. However, as the reader might have noticed, in this case it is insufficient for the gradient descent algorithm to simply converge to a local optimum. To address this, we also analyze the optimization landscape of (P2) when our measurement matrices are complex Gaussian, and show that with high probability *every local minimum is in fact a global minimum* (upto the equivalence relation $\sim$). We now state this result formally.

Theorem 2. *Let $\mathcal{A} = \{A_i\}$ be a set of complex Gaussian random matrices. Let c_i's are generated by quadratic measurements of an unknown vector $\mathbf{z}$. Given $\xi > 0$, there exist positive constants C, β, γ, ζ such that the following holds with probability at least $1-\xi$:*

- *the function f in* (P2) *is (β, ζ, γ)-strict saddle, and*
- *every local minimum $\mathbf{w}$ satisfies $d(\mathbf{w}, \mathbf{z}) = 0$*

provided the number of measurements satisfies $m > Cn$.

We refer the reader to [gd:cite full paper] for the details, but we will give here the brief idea of the proof.

Proof sketch. Notice that to show that the function f in (P2) has the strict saddle property, it suffices to only consider the points $\mathbf{x} \in \mathbb{C}^n$ such that $\|\nabla f(\mathbf{x})\| < \beta$ (otherwise, condition 1 of Definition 2 is satisfied). For all such points, we analyze the behavior of the Hessian and establish that there exists a direction $\Delta \in \mathbb{C}^n$ such that the following inequality holds

$$\langle \nabla^2 f(\mathbf{x})\Delta, \Delta\rangle \leq -c_0 \|\mathbf{x}\mathbf{x}^* - \mathbf{z}\mathbf{z}^*\|_F^2, \tag{7}$$

where $c_0 > 0$. Notice that the norm on the right side above is precisely the term $d(\mathbf{x}, \mathbf{z})$. This of course implies that whenever $d(\mathbf{x}, \mathbf{z}) > 0$, there is a direction where the Hessian has a strict negative curvature, and hence such a point cannot be an optimum. In other words, we can conclude that:

1) All local minimas satisfies $d(\mathbf{x}, \mathbf{z}) = 0$
2) All saddle points have a strict negative curvature.

This concludes the proof. $\square$

Finally, we remark that the properties of the optimization landscape that we have established allows one to use any gradient based iterative method to find a global optimum of (P2) (and hence find a solution to the quadratic feasibility problem). Furthermore, our results above also imply that any arbitrary initialization would suffice for the same; this is in sharp contrast to existing works like [31][gd:cite]. Formally,

Corollary 1. *Consider a gradient method applied to function f in* (P2). *Then, for an arbitrary initial point, the method point converges to a global minimum of the loss function f associated with the quadratic feasibility problem.*

We refer the reader to [gd:cite] for a full proof of this result. The one point we would like to remark about is that while the general flow of the argument used to establish this result has appeared in literature before (e.g., [29], [43]), to the best of our knowledge, the present paper is the first to derive such results in the complex domain.

References

[1] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit," in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*. IEEE, 2016, pp. 773–777.

[2] E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[3] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[4] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Applied and Computational Harmonic Analysis*, vol. 36, no. 3, pp. 473–494, 2014.

[5] R. Balan and D. Zou, "Phase retrieval using lipschitz continuous maps," *arXiv preprint arXiv:1403.2301*, 2014.

[6] J. Drenth, *Principles of protein X-ray crystallography*. Springer Science & Business Media, 2007.

[7] T. Dakic, *On the turnpike problem*. Simon Fraser University BC, Canada, 2000.

[8] P. M. Duxbury, L. Granlund, S. Gujarathi, P. Juhas, and S. J. Billinge, "The unassigned distance geometry problem," *Discrete Applied Mathematics*, vol. 204, pp. 117–132, 2016.

[9] S. Huang and I. Dokmanić, "Reconstructing point sets from distance distributions," *arXiv preprint arXiv:1804.02465*, 2018.

[10] J. Park and S. Boyd, "General heuristics for nonconvex quadratically constrained quadratic programming," *arXiv preprint arXiv:1703.07870*, 2017.

[11] L. L. Dines, "On the mapping of quadratic forms," *Bulletin of the American Mathematical Society*, vol. 47, no. 6, pp. 494–498, 1941.

[12] A. Beck and D. Pan, "A branch and bound algorithm for nonconvex quadratic optimization with ball and linear constraints," *Journal of Global Optimization*, vol. 69, no. 2, pp. 309–342, 2017.

[13] A. Beck, "Convexity properties associated with nonconvex quadratic matrix functions and applications to quadratic programming," *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 1–29, 2009.

[14] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM Journal on Optimization*, vol. 17, no. 3, pp. 844–860, 2006.

[15] S. Sahni, "Computationally related problems," *SIAM Journal on Computing*, vol. 3, no. 4, pp. 262–279, 1974.

[16] I. Pólik and T. Terlaky, "A survey of the s-lemma," *SIAM review*, vol. 49, no. 3, pp. 371–418, 2007.

[17] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, "Saving phase: Injectivity and stability for phase retrieval," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 106–125, 2014.

[18] R. Balan and Y. Wang, "Invertibility and robustness of phaseless reconstruction," *Applied and Computational Harmonic Analysis*, vol. 38, no. 3, pp. 469–488, 2015.

[19] A. I. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete & Computational Geometry*, vol. 13, no. 2, pp. 189–202, 1995.

[20] A. Konar and N. D. Sidiropoulos, "Fast approximation algorithms for a class of non-convex qcqp problems using first-order methods," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3494–3509.

[21] ——, "First-order methods for fast feasibility pursuit of non-convex qcqps," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 5927–5941.

[22] A. Konar, "Non-convex quadratically constrained quadratic programming: Hidden convexity, scalable approximation and applications," 2017.

[23] K. Huang and N. D. Sidiropoulos, "Consensus-admm for general quadratically constrained quadratic programming," *Transactions on Signal Processing*, vol. 64, no. 20, pp. 5297–5310, 2016.

[24] Y. S. Tan and R. Vershynin, "Phase retrieval via randomized kaczmarz: theoretical guarantees," *Information and Inference: A Journal of the IMA*, 2017.

[25] S. Bhojanapalli, A. Kyrillidis, and S. Sanghavi, "Dropping convexity for faster semi-definite optimization," in *Conference on Learning Theory*, 2016, pp. 530–582.

[26] C. Jin, S. M. Kakade, and P. Netrapalli, "Provable efficient online matrix completion via non-convex stochastic gradient descent," in *Advances in Neural Information Processing Systems*, 2016, pp. 4520–4528.

[27] P. Jain, P. Kar *et al.*, "Non-convex optimization for machine learning," *Foundations and Trends® in Machine Learning*, vol. 10, no. 3-4, pp. 142–336, 2017.

[28] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *Foundations of Computational Mathematics*, vol. 18, no. 5, pp. 1131–1198, 2018.

[29] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1233–1242.

[30] Y. Wang and Z. Xu, "Generalized phase retrieval: measurement number, matrix recovery and beyond," *Applied and Computational Harmonic Analysis*, 2017.

[31] S. Huang, S. Gupta, and I. Dokmanić, "Solving complex quadratic systems with full-rank random matrices," *arXiv preprint arXiv:1902.05612*, 2019.

[32] J. C. Duchi and F. Ruan, "Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval," *arXiv preprint arXiv:1705.02356*, 2017.

[33] Y. Ye and S. Zhang, "New results on quadratic minimization," *SIAM Journal on Optimization*, vol. 14, no. 1, pp. 245–267, 2003.

[34] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *arXiv preprint arXiv:1011.3027*, 2010.

[35] M. W. Meckes, "Concentration of norms and eigenvalues of random matrices," *Journal of Functional Analysis*, vol. 211, no. 2, pp. 508–524, 2004.

[36] S. S. Du, J. D. Lee, H. Li, L. Wang, and X. Zhai, "Gradient descent finds global minima of deep neural networks," *arXiv preprint arXiv:1811.03804*, 2018.

[37] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval," *arXiv preprint arXiv:1803.07726*, 2018.

[38] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 1724–1732.

[39] A. Bower, L. Jain, and L. Balzano, "The landscape of non-convex quadratic feasibility," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3974–3978.

[40] D. Davis, D. Drusvyatskiy, and C. Paquette, "The nonsmooth landscape of phase retrieval," *arXiv preprint arXiv:1711.03247*, 2017.

[41] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent converges to minimizers," *arXiv preprint arXiv:1602.04915*, 2016.

[42] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht, "First-order methods almost always avoid saddle points," *arXiv preprint arXiv:1710.07406*, 2017.

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

[44] R. Balan, "Reconstruction of signals from magnitudes of redundant representations: The complex case," *Foundations of Computational Mathematics*, vol. 16, no. 3, pp. 677–721, 2016.

[45] J. Cahill, P. G. Casazza, J. Peterson, and L. Woodland, "Phase retrieval by projections," *arXiv preprint arXiv:1305.6226*, 2013.

VI. Appendix

A. Proof of Theorem 1

Proof. Prove **??**$\Rightarrow$2.

'The following result from [30] is quite crucial,

Theorem 3 (Theorem 2.1, [30])**.** *Let* $\mathcal{A} = \{A_i\}_{i=1}^m \subset \mathbf{H}_n^m(\mathbb{C})$. *The following statements are equivalent:*

1) *For a given* $\mathcal{A} = \{A_i\}_{i=1}^m$*, the mapping* $\mathcal{M}_\mathcal{A}$ *has phase retrieval property.*
2) *There exists no nonzero vector* $\mathbf{v}, \mathbf{u} \in \mathbb{C}^n$ *with* $\mathbf{u} \neq ic\mathbf{v}$*,* $c \in \mathbb{R}$*, such that* $\langle A_j \mathbf{u}, \mathbf{v}\rangle = 0$ *for all* $1 \leq j \leq m$*.*

For the mapping $\mathcal{M}_\mathcal{A}$ to be injective, we need the following to be true,

$$\mathcal{M}_\mathcal{A}(\mathbf{x}) = \mathcal{M}_\mathcal{A}(\mathbf{y}) \quad \text{iff} \quad \mathbf{x} \sim \mathbf{y} \tag{8}$$

Thus for $\mathbf{x} \nsim \mathbf{y}$, we can guarantee that $\mathcal{M}_\mathcal{A}(\mathbf{x}) \neq \mathcal{M}_\mathcal{A}(\mathbf{y})$. Then we can also conclude that,

$$\begin{aligned}&\|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2^2 \\ &= \sum_{i=1}^m |(\mathbf{x}^* A_i \mathbf{x}) - (\mathbf{y}^* A_i \mathbf{y})|^2 > 0\end{aligned}$$

From lemma **??**, for every $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ $\exists$ $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ satisfying

$$(\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*) = (\mathbf{u}\mathbf{v}^* + \mathbf{v}\mathbf{u}^*) = [[\mathbf{u}, \mathbf{v}]] \tag{9}$$

Thus,

$$\|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2^2 = \sum_{i=1}^m |\langle A_i, [[\mathbf{u}, \mathbf{v}]]\rangle|^2 \tag{10}$$

In order to find upper and lower bound required in (3), we can define the lower bound α and upper bound β as below,

$$\begin{aligned}\alpha &:= \min_{T \in S^{1,1}, \|T\|_1 = 1} \sum_{i=1}^m |\langle A_i, T\rangle|^2 \\ \beta &:= \max_{T \in S^{1,1}, \|T\|_1 = 1} \sum_{i=1}^m |\langle A_i, T\rangle|^2\end{aligned}$$

Since the set $T \in S^{1,1}, \|T\|_1 = 1$ is compact, the constants α, β exists.

Prove **??**$\Leftarrow$2

We argue the negation that **??**$\nRightarrow$2.

Suppose the mapping $\mathcal{M}_\mathcal{A}$ is not injective, then $\exists \mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ such that,

$$\mathbf{x} \nsim \mathbf{y}, \quad \mathcal{M}_\mathcal{A}(\mathbf{y}) = \mathcal{M}_\mathcal{A}(\mathbf{x}) \tag{11}$$

Hence we can argue that $\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_1 \neq 0$, but $\|\mathcal{M}_\mathcal{A}(\mathbf{y}) - \mathcal{M}_\mathcal{A}(\mathbf{x})\|_2 = 0$. Hence we have proved the negation **??**$\nRightarrow$2. □

B. Proof of lemma 2

Proof. We want to examine the properties of the ratio,

$$V(\mathbf{x}, \mathbf{y}) \;=\; \frac{\|\mathcal{M}_\mathcal{A}(\mathbf{x}) - \mathcal{M}_\mathcal{A}(\mathbf{y})\|_2}{d(\mathbf{x}, \mathbf{y})}$$

This can be easily worked out to be,

$$V^2(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^m |\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2}{\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_1^2}$$

Hence the bi-Lipschitz nature of the mapping $\mathcal{M}_\mathcal{A}$ directly follows from Theorem 1 and lemma **??**. □

VII. High probability bounds

A. Proof of Lemma 3

Proof. Let $A \in \mathbb{C}^{n \times n}$ be a complex Hermitian Gaussian random matrix, i.e.

1) $\forall\, i$, $a_{ii} \sim \mathcal{N}(0, \sigma^2)$.
2) $\forall\, i, j$, $i \neq j$, $a_{ij} \sim \mathcal{N}(0, \frac{\sigma^2}{2}) + i\mathcal{N}(0, \frac{\sigma^2}{2})$.

Define the random variable Y as,

$$\begin{aligned}Y \;&= \frac{1}{m}\sum_{d=1}^m |\langle A_d, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2 \\ &= \frac{1}{m}\sum_{d=1}^m (\langle A_d, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle)(\langle A_d, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle) \\ &= \frac{1}{m}\sum_{d=1}^m \left(\sum_{ij} a_{ij}(x_i\bar{x}_j - y_i\bar{y}_j)\right)\left(\sum_{kl} a_{kl}(x_k\bar{x}_l - y_k\bar{y}_l)\right)\end{aligned}$$

Next we analyze $E[Y]$

$$\begin{aligned}&\mathbb{E}[Y] \\ &= \mathbb{E}\left(\frac{1}{m}\sum_{d=1}^m \left(\sum_{ij} a_{ij}(x_i\bar{x}_j - y_i\bar{y}_j)\right)\left(\sum_{kl} a_{kl}(x_k\bar{x}_l - y_k\bar{y}_l)\right)\right)\end{aligned}$$

For every matrix A_d, we can split the entire summation $(i, j, k, l) \in [1, n]^4$ into the following 4 sets such that:

1) $A := \{(i, j, k, l) | i = j = k = l\}$
2) $B := \{(i, j, k, l) | i = k, j = l\} \cap A^C$
3) $C := \{(i, j, k, l) | i = l, j = k\} \cap A^C$
4) $D := \{(i, j, k, l)\} \cap A^C \cap B^C \cap C^C$

Calculating the expectation of the sum of the elements in each individual sets:

1) For set A,

$$\begin{aligned}&\mathbb{E}\left(\sum_{(i,j,k,l)\in A} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^n |a_{ii}|^2(|x_i|^2 - |y_i|^2)^2\right) \\ &= \sigma^2\sum_{i=1}^n |x_i|^4 + |y_i|^4 - 2|x_i|^2|y_i|^2 \end{aligned}\tag{12}$$

2) For set C,

$$\mathbb{E}\left(\sum_{(i,j,k,l)\in B} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \mathbb{E}\left(\sum_{i,j=1,i\neq j}^{n} |a_{ij}|^2(|x_i|^2|x_j|^2 - y_i\bar{y}_j x_j\bar{x}_i \right.$$
$$\left. - x_i\bar{x}_j y_j\bar{y}_i + |y_i|^2|y_j|^2)\right)$$
$$= \sigma^2 \sum_{i,j=1,i\neq j}^{n} (|x_i|^2|x_j|^2 - y_i\bar{y}_j x_j\bar{x}_i - x_i\bar{x}_j y_j\bar{y}_i + |y_i|^2|y_j|^2) \tag{13}$$

3) For set B, since the matrix A is Hermitian $a_{ij} = \bar{a}_{ji}$

$$\mathbb{E}\left(\sum_{(i,j,k,l)\in C} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \mathbb{E}\left(\sum_{ij}(a_{ij})^2(x_i\bar{x}_j - y_i\bar{y}_j)^2)\right)$$
$$= 0 \tag{14}$$

Notice that $\forall i,j$

$$(a_{ij})^2 \ = ((a^r_{ij})^2 - (a^i_{ij})^2 + ia^r_{ij}a^i_{ij}) \tag{15}$$

Thus,

$$\mathbb{E}\left[(a_{ij})^2\right]$$
$$= \mathbb{E}\left[(a_{ij})^2\right]$$
$$= \mathbb{E}\left[((a^r_{ij})^2 - (a^i_{ij})^2 + ia^r_{ij}a^i_{ij})\right] \tag{16}$$

Since both the real and imaginary parts are independent and from the same distribution, we have that $\mathbb{E}\left[((a^r_{ij})^2 - (a^i_{ij})^2 + ia^r_{ij}a^i_{ij})\right] = 0$

4) For set D, as all the elements $(i,j,k,l) \in D$ make a_{ij}, a_{kl} independent of each other, we have,

$$\mathbb{E}\left(\sum_{(i,j,k,l)\in D} \bar{a}_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right) = 0 \tag{17}$$

Hence we can conclude,

$$\mathbb{E}[Y]$$
$$= \sigma^2 \left(\left(\sum_{i=1}^{n}|x_i|^2\right)^2 + \left(\sum_{i=1}^{n}|y_i|^2\right)^2 - 2\sum_{i=1}^{n}|x_i|^2|y_i|^2\right.$$
$$\left. - \sum_{i,j,i\neq j} y_i\bar{y}_j x_j\bar{x}_i - \sum_{i,j,i\neq j} y_j\bar{y}_i x_i\bar{x}_j\right)$$
$$= \sigma^2\left[\|\mathbf{x}\|_2^4 + \|\mathbf{x}\|_2^4 - |\langle \mathbf{x},\mathbf{y}\rangle|^2\right] \tag{18}$$

From Lemma 3.9 [44], note that $tr\{(\mathbf{xx}^* - \mathbf{yy}^*)^2\} = \left[\|\mathbf{x}\|_2^4 + \|\mathbf{x}\|_2^4 - |\langle \mathbf{x},\mathbf{y}\rangle|^2\right]$, where $tr\{\cdot\}$ represents the trace of a matrix. Since $\mathbf{xx}^* - \mathbf{yy}^*$ is a Hermitian matrix $tr\{(\mathbf{xx}^* - \mathbf{yy}^*)^2\} = \|\mathbf{xx}^* - \mathbf{yy}^*\|_F^2$. Hence we can finally state,

$$\mathbb{E}[Y] = \|\mathbf{xx}^* - \mathbf{yy}^*\|_F^2 \tag{19}$$

We focus our attention on obtaining concentration bounds. Evaluating the behaviour in the individual sets we have,

1) For elements in set A,

$$\sum_{(i,j,k,l)\in A} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)$$
$$- \mathbb{E}\left(\sum_{(i,j,k,l)\in A} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \sum_{i=1}^{n}\left(|a_{ii}|^2 - \sigma^2\right)\left(|x_i|^4 + |y_i|^4 - 2|x_i|^2|y_i|^2\right) \tag{20}$$

We can easily see that $\forall i, |a_{ii}|^2 - \sigma^2$ is a centered subexponential random variable.

2) For elements in set C,

$$\sum_{(i,j,k,l)\in B} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)$$
$$- \mathbb{E}\left(\sum_{(i,j,k,l)\in B} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \sum_{i,j=1,i\neq j}^{n} (|a_{ij}|^2 - \sigma^2)$$
$$(|x_i|^2|x_j|^2 - y_i\bar{y}_j x_j\bar{x}_i - x_i\bar{x}_j y_j\bar{y}_i + |y_i|^2|y_j|^2) \tag{21}$$

We can easily see that $\forall i,j \in [1,n]^2, i\neq j, |a_{ij}|^2 - \sigma^2$ is a centered subexponential random variable.

3) For elements in set B,

$$\sum_{(i,j,k,l)\in C} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)$$
$$- \mathbb{E}\left(\sum_{(i,j,k,l)\in C} a_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \sum_{ij}(a_{ij})^2(x_i)^2(\bar{x}_j)^2 \tag{22}$$

Splitting up $a^2_{ij} = (a^r_{ij})^2 - (a^i_{ij})^2 + ia^r_{ij}a^i_{ij}$ makes it easier to argue that $\forall i,j \in [1,n]^2, i\neq j, (a_{ij})^2$ is a centered subexponential random variable.

4) For elements in set D,

$$\sum_{(i,j,k,l)\in D} \bar{a}_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)$$
$$- \mathbb{E}\left(\sum_{(i,j,k,l)\in D} \bar{a}_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l)\right)$$
$$= \sum_{i,j,k,l\in D} \bar{a}_{ij}a_{kl}(x_i\bar{x}_j - y_i\bar{y}_j)(x_k\bar{x}_l - y_k\bar{y}_l) \tag{23}$$

Similar to the case with the elements of set C we can see that the above is a centered subexponential random variable.

Taking $\sigma^2 = 1$, we have the Bernstein type inequality [34] as,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\frac{1}{m}|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2 - \|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2\right| \leq t\right)$$
$$\geq 1 - c_0\exp\left(-c_1 m \min\left\{\frac{t^2}{K_4^2\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_2^8}, \frac{t}{K_4\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_\infty^4}\right\}\right) \tag{24}$$

for some constants $c_0, c_1 > 0$.

We can introduce the normalized substitute $t = \epsilon\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2$ to conclude,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\frac{1}{m}|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2 - \|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2\right|\right.$$
$$\left.\geq \epsilon\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2\right) \leq c_0\exp^{-c_1 m E(\epsilon)} \tag{25}$$

where $E(\epsilon) := \min\left\{\frac{\epsilon^2}{K^2}, \frac{\epsilon}{K}\right\}$ Note that $\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2$ is the distance metric $d(\cdot,\cdot)$ defined in (2). Hence we can rewrite the high probability result more consicely as,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\frac{1}{m}|\langle A_i, \mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\rangle|^2 - d(\mathbf{x},\mathbf{y})^2\right| \geq \epsilon d(\mathbf{x},\mathbf{y})^2\right)$$
$$\leq c_0\exp^{-c_1 m E(\epsilon)} \tag{26}$$

□

B. Proof Lemma 4

Proof. We aim to connect the supremum over $\mathbf{x}, \mathbf{y} \in S^{n-1}$ to the supremum over $\mathbf{x}, \mathbf{y} \in \mathcal{N}_\delta$.

Since $\mathcal{N}_\delta$ covers S^{n-1}. We can say that $\forall \mathbf{x} \in S^{n-1}$, $\exists \mathbf{u} \in \mathcal{N}_\delta$ such that $\|\mathbf{x} - \mathbf{u}\| \leq \delta$.

Hence $\forall \mathbf{x}_1, \mathbf{x}_2 \in S^{n-1}$, $\exists \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{N}_\delta$ such that,

$$\begin{aligned}
&|\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle - \langle A, \mathbf{y}_1\mathbf{y}_1^* - \mathbf{y}_2\mathbf{y}_2^*\rangle| \\
&= |\langle A\mathbf{x}_1, \mathbf{x}_1\rangle - \langle A\mathbf{y}_1, \mathbf{y}_1\rangle| - \langle A\mathbf{x}_2, \mathbf{x}_2\rangle - \langle A\mathbf{y}_2, \mathbf{y}_2\rangle| \\
&= |\langle A\mathbf{x}_1, \mathbf{x}_1\rangle - \langle A\mathbf{x}_1, \mathbf{y}_1\rangle + \langle A\mathbf{x}_1, \mathbf{y}_1\rangle - \langle A\mathbf{y}_1, \mathbf{y}_1\rangle| \\
&+ |\langle A\mathbf{x}_2, \mathbf{x}_2\rangle - \langle A\mathbf{x}_2, \mathbf{y}_2\rangle + \langle A\mathbf{x}_2, \mathbf{y}_2\rangle - \langle A\mathbf{y}_2, \mathbf{y}_2\rangle| \\
&= |\langle A\mathbf{x}_1, \mathbf{x}_1 - \mathbf{y}_1\rangle + \langle A\mathbf{y}_1, \mathbf{x}_1 - \mathbf{y}_1\rangle| \\
&+ |\langle A\mathbf{x}_2, \mathbf{x}_2 - \mathbf{y}_2\rangle + \langle A\mathbf{y}_2, \mathbf{x}_2 - \mathbf{y}_2\rangle| \\
&\leq 2\|A\|\|\mathbf{x}_1 - \mathbf{y}_1\| + 2\|A\|\|\mathbf{x}_2 - \mathbf{y}_2\| \\
&\leq 4\delta\|A\|
\end{aligned}$$

We can thus conclude,

$$|\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| - |\langle A, \mathbf{y}_1\mathbf{y}_1^* - \mathbf{y}_2\mathbf{y}_2^*\rangle| \leq 4\delta\|A\|$$
$$|\langle A, \mathbf{y}_1\mathbf{y}_1^* - \mathbf{y}_2\mathbf{y}_2^*\rangle| \geq |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| - 4\delta\|A\|$$

And,

$$|\langle A, \mathbf{y}_1\mathbf{y}_1^* - \mathbf{y}_2\mathbf{y}_2^*\rangle| - |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \leq 4\delta\|A\|$$
$$|\langle A, \mathbf{y}_1\mathbf{y}_1^* - \mathbf{y}_2\mathbf{y}_2^*\rangle| \leq |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| + 4\delta\|A\|$$

Taking supremum,

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{N}_\delta} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \\
&\geq \sup_{\mathbf{x}\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| - 4\delta\|A\| \\
&= (2 - 4\delta)\|A\| \\
&= (1 - 2\delta)\sup_{\mathbf{x}\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \\
&\sup_{\mathbf{x}\in\mathcal{N}_\delta} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| \\
&\leq \sup_{\mathbf{x}\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle| + 4\delta\|A\| \\
&= (2 + 4\delta)\|A\| \\
&= (1 + 2\delta)\sup_{\mathbf{x}\in S^{n-1}} |\langle A, \mathbf{x}_1\mathbf{x}_1^* - \mathbf{x}_2\mathbf{x}_2^*\rangle|
\end{aligned}$$

□

Lemma 5. *Let $\{A_i\}_{i=1}^n$ be Gaussian random matrices. Then with probability $1 - 4e^{-cmE(\epsilon)}$ we can say that,*

$$\left|\frac{1}{m}\sum_{d=1}^{m}\langle A_d, \Delta\bar{\Delta}^\top\rangle\langle A_d, \mathbf{x}\bar{\mathbf{x}}^\top\rangle - \langle A_d, \Delta\bar{\mathbf{x}}^\top\rangle\langle A_d, \mathbf{x}\bar{\Delta}^\top\rangle\right|$$
$$\leq \epsilon\|\Delta\|_2^2\|\mathbf{x}\|_2^2 \tag{27}$$

where $E(\epsilon) := \min\left\{\frac{\epsilon^2}{4K_4^2}, \frac{\epsilon}{2K_4}\right\}$

Proof. Evaluating the expression, we can see that,

$$\begin{aligned}
&\frac{1}{m}\sum_{d=1}^{m}\langle A_d, \Delta\bar{\Delta}^\top\rangle\langle A_d, \mathbf{x}\bar{\mathbf{x}}^\top\rangle - \langle A_d, \Delta\bar{\mathbf{x}}^\top\rangle\langle A_d, \mathbf{x}\bar{\Delta}^\top\rangle \\
&= \frac{1}{m}\sum_{d=1}^{m}\left(\sum_{ij} a_{ij}^d\Delta_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}^d\mathbf{x}_k\bar{\mathbf{x}}_l\right) \\
&- \left(\sum_{ij} a_{ij}^d\mathbf{x}_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}^d\Delta_k\bar{\mathbf{x}}_l\right)
\end{aligned} \tag{28}$$

For any A_d, Split the entire summation $(i,j,k,l) \in [1,n]^4$ into the following 4 sets such that:

1) $A := \{(i,j,k,l) | i = j = k = l\}$
2) $B := \{(i,j,k,l) | i = k, j = l\} \cap A^C$
3) $C := \{(i,j,k,l) | i = l, j = k\} \cap A^C$
4) $D := \{(i,j,k,l)\} \cap A^C \cap B^C \cap C^C$

Calculating the expectation of the sum of the elements in each individual sets:

1) For set A,

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{ij} a_{ij}\Delta_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\mathbf{x}_k\bar{\mathbf{x}}_l\right)\right. \\
&\left.- \left(\sum_{ij} a_{ij}\mathbf{x}_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\Delta_k\bar{\mathbf{x}}_l\right)\right] \\
&= \mathbb{E}\left[a_{ii}^2\Delta_i\bar{\Delta}_i\mathbf{x}_i\bar{\mathbf{x}}_j - a_{ii}^2\Delta_i\bar{\Delta}_i\mathbf{x}_i\bar{\mathbf{x}}_i\right] \\
&= \mathbb{E}[0] = 0
\end{aligned} \tag{29}$$

2) For set B,

$$\begin{aligned}&\mathbb{E}\Bigg[\left(\sum_{ij} a_{ij}\Delta_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\mathbf{x}_k\bar{\mathbf{x}}_l\right)\\&-\left(\sum_{ij} a_{ij}\mathbf{x}_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\Delta_k\bar{\mathbf{x}}_l\right)\Bigg]\\&=\mathbb{E}\left[a_{ij}^2\Delta_i\bar{\Delta}_j\mathbf{x}_i\bar{\mathbf{x}}_j - a_{ij}^2\Delta_i\bar{\Delta}_j\mathbf{x}_i\bar{\mathbf{x}}_j\right]\\&=\mathbb{E}[0]=0\end{aligned} \tag{30}$$

3) For set C, since the matrix A is hermitian $a_{ij}=\bar{a}_{ji}$

$$\begin{aligned}&\Bigg[\left(\sum_{ij} a_{ij}\Delta_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\mathbf{x}_k\bar{\mathbf{x}}_l\right)\\&-\left(\sum_{ij} a_{ij}\mathbf{x}_i\bar{\Delta}_j\right)\left(\sum_{kl} a_{kl}\Delta_k\bar{\mathbf{x}}_l\right)\Bigg]\\&=\left[|a_{ij}|^2\Delta_i\bar{\Delta}_j\mathbf{x}_j\bar{\mathbf{x}}_i - |a_{ij}|^2\Delta_j\bar{\Delta}_j\mathbf{x}_i\bar{\mathbf{x}}_i\right]\end{aligned} \tag{31}$$

Notice that $\forall i,j$

$$\begin{aligned}&=|a_{ji}|^2\Delta_j\bar{\Delta}_i\mathbf{x}_i\bar{\mathbf{x}}_j + |a_{ij}|^2\Delta_i\bar{\Delta}_j\mathbf{x}_j\bar{\mathbf{x}}_i\\&-|a_{ji}|^2\Delta_i\bar{\Delta}_i\mathbf{x}_j\bar{\mathbf{x}}_j - |a_{ij}|^2\Delta_j\bar{\Delta}_j\mathbf{x}_i\bar{\mathbf{x}}_i\end{aligned} \tag{32}$$

Since $|a_{ij}|^2=|a_{ji}|^2$. Thus,

$$\begin{aligned}&=|a_{ji}|^2\left[\Delta_j\bar{\Delta}_i\mathbf{x}_i\bar{\mathbf{x}}_j + \Delta_i\bar{\Delta}_j\mathbf{x}_j\bar{\mathbf{x}}_i\right.\\&\quad\left.-\Delta_i\bar{\Delta}_i\mathbf{x}_j\bar{\mathbf{x}}_j - \Delta_j\bar{\Delta}_j\mathbf{x}_i\bar{\mathbf{x}}_i\right]\\&=|a_{ji}|^2\left[\Delta_j\bar{\Delta}_i\mathbf{x}_i\bar{\mathbf{x}}_j + \Delta_i\bar{\Delta}_j\mathbf{x}_j\bar{\mathbf{x}}_i\right.\\&\quad\left.-\|\Delta_i\|_2^2\|\mathbf{x}_j\|_2^2 - \|\Delta_j\|_2^2\|\mathbf{x}_i\|_2^2\right]\\&\le 0\end{aligned} \tag{33}$$

4) For set D, as all the elements $(i,j,k,l)\in D$ make a_{ij},a_{kl} independent of each other, we have,

$$\mathbb{E}\left(\sum a_{ij}a_{kl}\bar{\Delta}_j\bar{\mathbf{x}}_l\left(\Delta_i\mathbf{x}_k-\Delta_k\mathbf{x}_i\right)\right)=0 \tag{34}$$

Hence we can conclude,

$$\mathbb{E}\left[\frac{1}{m}\sum_{d=1}^m\left(\langle A_d,\Delta\bar{\Delta}^\top\rangle\langle A_d,\mathbf{x}\bar{\mathbf{x}}^\top\rangle-\langle A_d,\Delta\bar{\mathbf{x}}^\top\rangle\langle A_d,\mathbf{x}\bar{\Delta}^\top\rangle\right)\right]=0 \tag{35}$$

We focus our attention on obtaining concentration bounds. Evaluating the behaviour in the individual sets we have,

1) For elements in set D,

$$\begin{aligned}&\frac{1}{m}\sum_{d=1}^m\left(\sum_{(i,j,k,l)\in D} a_{ij}^d a_{kl}^d\bar{\Delta}_j\bar{\mathbf{x}}_l\left(\Delta_i\mathbf{x}_k-\Delta_k\mathbf{x}_i\right)\right)\\&-\mathbb{E}\left(\frac{1}{m}\sum_{d=1}^m\sum_{(i,j,k,l)\in D} a_{ij}^d a_{kl}^d\bar{\Delta}_j\bar{\mathbf{x}}_l\left(\Delta_i\mathbf{x}_k-\Delta_k\mathbf{x}_i\right)\right)\\&=\sum_{(i,j,k,l)\in D} a_{ij}^d a_{kl}^d\bar{\Delta}_j\bar{\mathbf{x}}_l\left(\Delta_i\mathbf{x}_k-\Delta_k\mathbf{x}_i\right)\end{aligned} \tag{36}$$

Similar to the case with the elements of set D we can see that the above is a centered subexponential random variable. Hence using Bernstein type inequality [34], we can say that,

$$\begin{aligned}&Pr\left(\left|\frac{1}{m}\sum_{d=1}^m\sum_{(i,j,k,l)\in D} a_{ij}^d a_{kl}^d\bar{\Delta}_j\bar{\mathbf{x}}_l\left(\Delta_i\mathbf{x}_k-\Delta_k\mathbf{x}_i\right)\right|\ge t\right)\\&\le 4\exp\left(-cm\min\left\{\frac{t^2}{4K_4^2\|\Delta\|_2^4\|\mathbf{x}\|_2^4},\frac{t}{2K_4\|\Delta\|_\infty^2\|\mathbf{x}\|_\infty^2}\right\}\right)\end{aligned} \tag{37}$$

where $K_4 := \max_{i,j}\{\|\sum_{i,j,k,l\in D}(\bar{a}_{ij}a_{kl}^d)_{\mathbb{R}}\|_{\psi_1}, \|\sum_{i,j,k,l\in D}(\bar{a}_{ij}a_{kl}^d)_{\mathbb{C}}\|_{\psi_1}\}$ is the subexponential norm.

Thus we can argue that,

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{m}\sum_{d=1}^m\langle A_d,\Delta\bar{\Delta}^\top\rangle\langle A_d,\mathbf{x}\bar{\mathbf{x}}^\top\rangle-\langle A_d,\Delta\bar{\mathbf{x}}^\top\rangle\langle A_d,\mathbf{x}\bar{\Delta}^\top\rangle\right|\ge t\right)\\&\le 4\exp^{-cmD(t)}\end{aligned} \tag{38}$$

where

$$D(t):=\min\left\{\frac{t^2}{4K_4^2\|\Delta\|_2^4\|\mathbf{x}\|_2^4},\frac{t}{2K_4\|\Delta\|_\infty^2\|\mathbf{x}\|_\infty^2}\right\} \tag{39}$$

We can introduce the normalized substitute $t=\epsilon\|\Delta\|_2^2\|\mathbf{x}\|_2^2$ to conclude,

$$\begin{aligned}&\mathbb{P}\ \left(\left|\frac{1}{m}\sum_{d=1}^m\langle A_d,\Delta\bar{\Delta}^\top\rangle\langle A_d,\mathbf{x}\bar{\mathbf{x}}^\top\rangle\right.\right.\\&\left.\left.-\langle A_d,\Delta\bar{\mathbf{x}}^\top\rangle\langle A_d,\mathbf{x}\bar{\Delta}^\top\rangle\right|\ge\epsilon\|\Delta\|_2^2\|\mathbf{x}\|_2^2\right)\\&\le 4\exp^{-cmE(\epsilon)}\end{aligned} \tag{40}$$

where $E(\epsilon):=\min\left\{\frac{\epsilon^2}{4K_4^2},\frac{\epsilon}{2K_4}\right\}$ □

VIII. NON-CONVEX LANSCAPE

A. Supporting Lemmas

Taking inspiration from [29], we can prove the following:

Lemma 6. *Let* $\mathbf{x},\mathbf{x}^*\in\mathbb{C}^n$*. Then,*

$$\|(\mathbf{x}-e^{i\phi}\mathbf{x}^*)(\mathbf{x}-e^{i\phi}\mathbf{x}^*)^*\|_F^2\le 2\|\mathbf{x}\mathbf{x}^*-\mathbf{x}^*(\mathbf{x}^*)^*\| \tag{41}$$

Proof. We first note that,

$$\begin{aligned}&\arg_\theta\min\|\mathbf{x}-e^{i\theta}\mathbf{x}^*\|^2\\&=\arg_\theta\min\left(\mathbf{x}-e^{i\theta}\mathbf{x}^*\right)^*\left(\mathbf{x}-e^{i\theta}\mathbf{x}^*\right)\\&=\arg_\theta\min\|\mathbf{x}\|^2+\|\mathbf{x}^*\|-e^{-i\theta}(\mathbf{x}^*)^*\mathbf{x}-e^{i\theta}\mathbf{x}^*\mathbf{x}^*\\&=\arg_\theta\min\|\mathbf{x}\|^2+\|\mathbf{x}^*\|-2Re(\langle\mathbf{x},e^{i\theta}\mathbf{x}^*\rangle)\end{aligned} \tag{42}$$

The minimum can only be achieved at a point where $\mathbf{x}^*(e^{i\phi}\mathbf{x}^*) \geq 0$. We already know that,

$$\begin{aligned}
&\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\| \\
&= \|\mathbf{x}\Delta^* + \Delta\mathbf{x}^* - \Delta\Delta^*\|_F^2 \\
&= Tr\left((\mathbf{x}\Delta^* + \Delta\mathbf{x}^* - \Delta\Delta^*)^*(\mathbf{x}\Delta^* + \Delta\mathbf{x}^* - \Delta\Delta^*)\right) \\
&= \big(\|\mathbf{x}\Delta^*\|_F^2 + (\langle\mathbf{x},\Delta\rangle)^2 + (\langle\Delta,\mathbf{x}\rangle)^2 + \|\Delta\mathbf{x}^*\|_F^2 \\
&-2\langle\mathbf{x},\Delta\rangle\|\Delta\|_F^2 - 2\langle\Delta,\mathbf{x}\rangle\|\Delta\|_F^2 + \|\Delta\Delta^*\|_F^2\big) \\
&= \big(2\|\langle\mathbf{x},\Delta\rangle\|_F^2 + 2Re((\langle\mathbf{x},\Delta\rangle)^2) \\
&- 4Re(\langle\mathbf{x},\Delta\rangle)\|\Delta\|_F^2 + \|\Delta\Delta^*\|_F^2 \\
&= \big(2\mathbf{x}^*\mathbf{x}\Delta^*\Delta + 2Re((\langle\mathbf{x},\Delta\rangle)^2) \\
&- 4Re(\mathbf{x}^*\Delta\Delta^*\Delta) + \|\Delta\Delta^*\|_F^2
\end{aligned}$$

Since $\mathbf{x}^*\mathbf{x}\Delta^*\Delta = \|\langle\mathbf{x},\Delta\rangle\|_F^2$, its a real quantity.

$$\begin{aligned}
&\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\| \\
&= 2\mathbf{x}^*\left(\mathbf{x}-\Delta\right)\Delta^*\Delta + 2Re((\langle\mathbf{x},\Delta\rangle)^2) \\
&- 2Re(\mathbf{x}^*\Delta\Delta^*\Delta) + \|\Delta\Delta^*\|_F^2 \\
&= 2\mathbf{x}^*\left(\mathbf{x}-\Delta\right)\Delta^*\Delta + (\langle\mathbf{x},\Delta\rangle)^2 + (\langle\Delta,\mathbf{x}\rangle)^2 \\
&- \langle\mathbf{x},\Delta\rangle\|\Delta\|_F^2 - \langle\Delta,\mathbf{x}\rangle\|\Delta\|_F^2 + \|\Delta\Delta^*\|_F^2 \\
&= 2\mathbf{x}^*\left(\mathbf{x}-\Delta\right)\Delta^*\Delta + \left(\langle\mathbf{x},\Delta\rangle - \frac{1}{2}\langle\Delta,\Delta\rangle\right)^2 \\
&+ \left((\langle\Delta,\mathbf{x}\rangle - \frac{1}{2}\langle\Delta,\Delta\rangle\right)^2 + \frac{1}{2}\|\Delta\Delta^*\|_F^2 \\
&= 2\mathbf{x}^*\left(\mathbf{x}-\Delta\right)\Delta^*\Delta + 2Re\left((\langle\Delta,\mathbf{x}-\frac{1}{2}\Delta\rangle\right)^2 + \frac{1}{2}\|\Delta\Delta^*\|_F^2 \\
&= 2\mathbf{x}^*e^{i\phi}\mathbf{x}^*\Delta^*\Delta + \frac{1}{2}Re\left((\langle\Delta,\mathbf{x}+e^{i\phi}\mathbf{x}^*\rangle\right)^2 + \frac{1}{2}\|\Delta\Delta^*\|_F^2
\end{aligned}$$

It can be seen that $Im\left(\langle\Delta,\mathbf{x}+e^{i\phi}\mathbf{x}^*\rangle\right) = 0$. We cannot say this statement for general θ. We also know that $\mathbf{x}^*(e^{i\theta}\mathbf{x}^*) \geq 0$. Hence we can say,

$$\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\| \geq \frac{1}{2}\|\Delta^*\Delta\|_F^2 \tag{43}$$

□

Lemma 7. *For any rank one matrix,*

$$\|\mathbf{x}\mathbf{x}^*\|_F^2 = \|\mathbf{x}\|_2^4 \tag{44}$$

Proof.

$$\begin{aligned}
\|\mathbf{x}\mathbf{x}^*\|_F^2 &= \sum_{i,j=1}^n |x_i\bar{x}_j|^2 = \sum_{i,j=1}^n (x_i\bar{x}_j)^*(x_i\bar{x}_j) = \sum_{i,j=1}^n x_j\bar{x}_i x_i\bar{x}_j \\
&= \sum_{i,j=1}^n |x_j|^2|x_i|^2 = (\sum_{i=1}^n |x_i|^2)^2 = \|\mathbf{x}\|_2^4
\end{aligned} \tag{45}$$

□

B. Wirtinger Calculus

Going with standard arguments from wirtinger calculus [30], note that,

$$\begin{aligned}
f(\mathbf{x}) = g(\mathbf{x},\bar{\mathbf{x}}) &= \sum_{i=1}^m g_i(\mathbf{x},\bar{\mathbf{x}}) \\
&= \frac{1}{2}\sum_{i=1}^n (\bar{\mathbf{x}}^\top A_i\mathbf{x} - b_i)^2
\end{aligned}$$

For the gradient we have,

$$\nabla g(\mathbf{x},\bar{\mathbf{x}}) = \sum_{i=1}^n \begin{bmatrix} (\bar{\mathbf{x}}^\top A_i\mathbf{x} - b_i)A_i\mathbf{x} \\ (\mathbf{x}^\top A_i\bar{\mathbf{x}} - b_i)A_i\bar{\mathbf{x}} \end{bmatrix} \tag{46}$$

For the hessian, we have

$$\nabla^2 g(\mathbf{x},\bar{\mathbf{x}}) = \sum_{i=1}^m \begin{bmatrix} (2\bar{\mathbf{x}}^\top A_i\mathbf{x} - b_i)A_i & (A_i\mathbf{x})(A_i\mathbf{x})^\top \\ (A_i\bar{\mathbf{x}})(A_i\bar{\mathbf{x}})^\top & (2\bar{\mathbf{x}}^\top A_i\mathbf{x} - b_i)A_i \end{bmatrix} \tag{47}$$

For the rest of the write-up, define $\Delta = \mathbf{x} - e^{i\phi}\mathbf{x}^*$ such that $\phi = \min_{\theta\in[0,2\pi]} \|\mathbf{x} - e^{i\theta}\mathbf{x}^*\|$. Notice that the following relation holds,

$$\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^* + \Delta\Delta^* = \mathbf{x}\Delta^* + \Delta\mathbf{x}^* \tag{48}$$

The following can be verified easily,

$$\begin{aligned}
&\langle\nabla g(\mathbf{x}), \begin{bmatrix}\Delta \\ \bar{\Delta}\end{bmatrix}\rangle \\
&= \langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}^*})^\top\rangle\langle A_i, \mathbf{x}\bar{\Delta}^\top + \Delta\bar{\mathbf{x}}^\top\rangle \\
&= \langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}^*})^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top + \Delta\bar{\Delta}^\top\rangle
\end{aligned} \tag{49}$$

$$\begin{aligned}
\begin{bmatrix}\Delta \\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g_i(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta \\ \bar{\Delta}\end{bmatrix} &= (2\mathbf{x}^\top A_i\bar{\mathbf{x}} - b_i)(\bar{\Delta}^\top A_i\Delta + \Delta^\top A_i\bar{\Delta}) \\
&+ \left((\Delta^\top A_i\bar{\mathbf{x}})^2 + (\bar{\Delta}A_i\mathbf{x})^2\right)
\end{aligned} \tag{50}$$

C. Proof of Theorem 2

Proof. Notice that,

$$\begin{aligned}
&(\Delta A\bar{\mathbf{x}})^2 + (\bar{\Delta}^\top A\mathbf{x})^2 \\
&= \left(\langle A, \mathbf{x}\bar{\Delta}^\top + \Delta\bar{\mathbf{x}}^\top\rangle\right)^2 - 2(\Delta^\top A\bar{\mathbf{x}})(\bar{\Delta}^\top A\mathbf{x}) \\
&= \left(\langle A, \mathbf{x}\mathbf{x}^\top - \mathbf{x}^*(\mathbf{x}^*)^\top + \Delta\Delta^\top\rangle\right)^2 \\
&- 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)
\end{aligned} \tag{51}$$

Thus we have,

$$
\begin{aligned}
&\begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g_i(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}\\
&= \langle A_i, 2\mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, 2\Delta\bar{\Delta}^\top\rangle\\
&+ \left(\langle A, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top + \Delta\bar{\Delta}^\top\rangle\right)^2\\
&- 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&= 2\left(\langle A_i, 2\mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\right)\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&- 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&= 2\langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top\rangle\\
&+ 2\left(\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\right)\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\\
&+ \langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&- 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&= 2\langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top\rangle\\
&+ 2\left(\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\right)\\
&- 2\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\\
&+ \langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&- 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&= 2\langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top\rangle - 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&- 3\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\\
&+ 4\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top + \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&= 2\langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top\rangle - 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&- 3\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle\\
&+ 4\langle \nabla g(\mathbf{x},\bar{\mathbf{x}}), \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}\rangle
\end{aligned} \tag{52}
$$

Thus we can say overall that,

$$
\begin{aligned}
&\begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix} = \sum_{i=1}^m \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g_i(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}\\
&= \sum_{i=1}^m \big(2\langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top\rangle - 2(\langle A, \Delta\bar{\mathbf{x}}^\top\rangle)(\langle A, \mathbf{x}\bar{\Delta}^\top\rangle)\\
&-3\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\langle A_i, \mathbf{x}\bar{\mathbf{x}}^\top - \mathbf{x}^*(\bar{\mathbf{x}}^*)^\top\rangle\\
&+ \langle A_i, \Delta\bar{\Delta}^\top\rangle\langle A_i, \Delta\bar{\Delta}^\top\rangle + 4\langle \nabla g(\mathbf{x},\bar{\mathbf{x}}), \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}\rangle\Big)\\
&\le 4\delta\|\Delta\|_2 + \beta\|\Delta\Delta^*\|_F^2\\
&- 3\alpha\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\|_F^2\\
&\le 4\delta\|\Delta\|_2 + 2\beta\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\|_F^2\\
&- 3\alpha\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\|_F^2\\
&\le (2\beta - 3\alpha)\,\|\mathbf{x}\mathbf{x}^* - \mathbf{x}^*(\mathbf{x}^*)^*\|_F^2 + 4\delta\|\Delta\|_2
\end{aligned}
$$

Let $\beta = 1 + c, \alpha = 1 - c$ for some small $c > 0$. If $\mathbf{x}$ is not close to $\mathbf{x}^*$, and $\|\Delta\| \ge C_0\delta$, for sufficiently large $C_0 > 0$, we will have

$$
\begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix} \le (-1+5c)C_0^2\delta^2 + 4C_0\delta^2 \le 0 \tag{53}
$$

A particular set of (c, C_0) which fit the above condition is $c = \frac{1}{20}, C_0 = 10$, then

$$
\begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix}^* \nabla^2 g(\mathbf{x},\bar{\mathbf{x}}) \begin{bmatrix}\Delta\\ \bar{\Delta}\end{bmatrix} \le (-1+5c)C_0^2\delta^2 + 4C_0\delta^2 \le \tag{54}
$$

Since ϵ, δ can be taken to be arbitrarily close to 0, and $\alpha \to \beta \neq 0$, thus we can see that the strict saddle point condition holds. □

IX. Applications

A. Power system state estimation problem

Apart from being a broader class of problems encompassing phase retrieval, the problem setup (P1) also has applications in power system engineering. Given a network of buses and transmission lines, the goal is to estimate complex voltages across all buses from a subset of noisy power and voltage magnitude measurements. In the AC power model, these measurements are quadratically dependent on the voltage values to be determined. Let $\{c_i\}_{i=1}^m$ be the set of measurements and $\{A_i\}_{i=1}^m$ be the corresponding bus admittance value matrices. Then the problem boils down to an estimation problem

$$
\begin{aligned}
&\text{find} \quad \mathbf{x}\\
&\text{s.t.} \quad c_i = \mathbf{x}^* A_i \mathbf{x} + \nu_i \quad \forall i = 1, 2, \ldots, m.
\end{aligned}
$$

where $\nu_i \sim \mathcal{N}(0, \sigma_i^2)$ is gaussian noise associated with the readings. [21]. For details on the problem setup, please refer [1].

B. Fusion Phase retrieval

Let $\{W_i\}_{i=1}^m$ be a set of subspace of $\mathbb{R}^n/\mathbb{C}^n$. Fusion phase retrieval deals with the problem of recovering $\mathbf{x}$ upto a phase ambiguity from the measurements of the form $\{\|P_i\mathbf{x}\|\}_{i=1}^m$, where $P_i : \mathbb{C}^n/\mathbb{R}^n \to W_i$ are projection operators onto the subspaces. [45] had the initial results on this problem with regards to the conditions on the subspaces and minimum number of such subspaces required for successful recovery of $\mathbf{x}$ under phase ambiguity.

C. X-ray crystallography

In X-ray crystallography, especially in crystal twinning [6], the measurements are obtained with orthogonal matrices $Q_i^2 = Q_i$ which again would be solved by out setup.

In the worst case, a feasibility quadratic feasibility problem can be NP-hard, which makes the setup (**??**) we address all the more interesting as we can highlight properties about a subgroup of quadratic feasibility problems and take a shot at providing provably converging algorithm for the same. This question resonates quite closely with many applications of quadratic feasibility as discussed above. In this write-up we have only considered the noiseless system, which later can be extended to noisy system analysis.