Appl. Comput. Harmon. Anal. 50 (2021) 1–15

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

The recovery of complex sparse signals from few phaseless measurements

Yu Xia [a,*,1], Zhiqiang Xu [b,c,2]

[a] *Department of Mathematics, Hangzhou Normal University, Hangzhou 311121, China*
[b] *LSEC, Inst. Comp. Math., Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing, 100091, China*
[c] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

ARTICLE INFO

Article history:
Received 25 November 2019
Received in revised form 31 May 2020
Accepted 18 August 2020
Available online 26 August 2020
Communicated by Holger Rauhut

ABSTRACT

We study the stable recovery of complex k-sparse signals from as few phaseless measurements as possible. The main result is to show that one can employ ℓ_1 minimization to stably recover complex k-sparse signals from $m \geq O(k\log(n/k))$ complex Gaussian random quadratic measurements with high probability. To do that, we establish that Gaussian random measurements satisfy the restricted isometry property over rank-2 and sparse matrices with high probability. This paper presents the first theoretical estimation of the measurement number for stably recovering complex sparse signals from complex Gaussian quadratic measurements.

1. Introduction

1.1. Sparse phase retrieval

Suppose that $\mathbf{x}_0 \in \mathbb{F}^n$ is a k-sparse signal, i.e., $\|\mathbf{x}_0\|_0 \leq k$, where $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. We are interested in recovering $\mathbf{x}_0$ from

$$y_j = |\langle \mathbf{a}_j, \mathbf{x}_0 \rangle|^2 + w_j, \quad j = 1, \ldots, m,$$

where $\mathbf{a}_j \in \mathbb{F}^n$ is a measurement vector and $w_j \in \mathbb{R}$ is the noise. This problem is called *sparse phase retrieval* [2,9,12]. Let $\mathcal{A} : \mathbb{F}^{n\times n} \to \mathbb{R}^m$ be a linear map which is defined as

$$\mathcal{A}(X) = (\mathbf{a}_1^* X \mathbf{a}_1, \ldots, \mathbf{a}_m^* X \mathbf{a}_m),$$

* Corresponding author.
E-mail addresses: yxia@hznu.edu.cn (Y. Xia), xuzq@lsec.cc.ac.cn (Z. Xu).
[1] Yu Xia was supported by NSFC grant (11901143), Zhejiang Provincial Natural Science Foundation (LQ19A010008), Department of Education of Zhejiang Province Science Foundation (Y201840082).
[2] Zhiqiang Xu was supported by Beijing Natural Science Foundation (Z180002) and by NSFC grant (11688101).

https://doi.org/10.1016/j.acha.2020.08.001
1063-5203/

where $X \in \mathbb{F}^{n\times n}, \mathbf{a}_j \in \mathbb{F}^n, j = 1, \ldots, m$. By abuse of notation we set

$$\mathcal{A}(\mathbf{x}) := \mathcal{A}(\mathbf{x}\mathbf{x}^*) = (|\langle \mathbf{a}_1, \mathbf{x}\rangle|^2, \ldots, |\langle \mathbf{a}_m, \mathbf{x}\rangle|^2),$$

where $\mathbf{x} \in \mathbb{F}^n$. We also set

$$\tilde{\mathbf{x}}_0 \ := \ \{c\mathbf{x}_0 : |c| = 1, c \in \mathbb{F}\}.$$

The aim of sparse phase retrieval is to recover $\tilde{\mathbf{x}}_0$ from noisy measurements $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{w}$, with $\mathbf{y} = (y_1, \ldots, y_m)^T \in \mathbb{R}^m$ and $\mathbf{w} = (w_1, ..., w_m)^T \in \mathbb{R}^m$. One question in sparse phase retrieval is: *how many measurements $y_j, j = 1, \ldots, m$, are needed to stably recover $\tilde{\mathbf{x}}_0$?* For the case $\mathbb{F} = \mathbb{R}$, in [5], Eldar and Mendelson established that $m = O(k \log(n/k))$ Gaussian random quadratic measurements are enough to stably recover k-sparse signals $\tilde{\mathbf{x}}_0$. For the complex case, Iwen, Viswanathan and Wang suggested a two-stage strategy for sparse phase retrieval and show that $m = O(k \log(n/k))$ measurements can guarantee the stable recovery of $\tilde{\mathbf{x}}_0$ [7]. However, the strategy in [7] requires the measurement matrix to be written as a product of two random matrices. Hence, it still remains open whether one can stably recover arbitrary complex k-sparse signal $\tilde{\mathbf{x}}_0$ from $m = O(k \log(n/k))$ Gaussian random quadratic measurements. One of the aims of this paper is to confirm that $m = O(k \log(n/k))$ Gaussian random quadratic measurements are enough to guarantee the stable recovery of arbitrary complex k-sparse signals. In fact, we do so by employing ℓ_1 minimization.

1.2. ℓ_1 minimization

Set $A := (\mathbf{a}_1, \ldots, \mathbf{a}_m)^T \in \mathbb{F}^{m\times n}$. One classical result in compressed sensing is that one can use ℓ_1 minimization to recover k-sparse signals, i.e.,

$$\underset{\mathbf{x}\in\mathbb{F}^n}{\operatorname{argmin}}\{\|\mathbf{x}\|_1 : A\mathbf{x} = A\mathbf{x}_0\} = \mathbf{x}_0,$$

provided that the measurement matrix A meets the RIP condition [4]. Recall that a matrix A satisfies the k-order RIP condition with RIP constant $\delta_k \in [0, 1)$ if

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \le \|A\mathbf{x}\|_2^2 \le (1 + \delta_k)\|\mathbf{x}\|_2^2$$

holds for all k-sparse vectors $\mathbf{x} \in \mathbb{F}^n$. Using tools from probability theory, one can show that Gaussian random matrices satisfy the k-order RIP with high probability provided $m = O(k \log(n/k))$ [1].

Naturally, one is interested in employing ℓ_1 minimization for sparse phase retrieval. We consider the following model:

$$\underset{\mathbf{x}\in\mathbb{F}^n}{\operatorname{argmin}}\{\|\mathbf{x}\|_1 : |A\mathbf{x}| = |A\mathbf{x}_0|\}. \tag{1.1}$$

Although the constrained conditions in (1.1) are non-convex, the model (1.1) is more amenable to algorithmic recovery. In fact, algorithms have been developed for solving (1.1) [9,15,16]. For the case $\mathbb{F} = \mathbb{R}$, the performance of (1.1) was studied in [11,6,13,8]. Particularly, in [11], it was shown that if $A \in \mathbb{R}^{m\times n}$ is a random Gaussian matrix with $m = O(k \log(n/k))$, then

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}}\{\|\mathbf{x}\|_1 : |A\mathbf{x}| = |A\mathbf{x}_0|\} \ = \ \pm\mathbf{x}_0$$

holds with high probability. The methods developed in [11] heavily depend on $A\mathbf{x}_0$ is a *real* vector and one still does not know the performance of ℓ_1 minimization for recovering complex sparse signals. As mentioned in [11]: "*The extension of these results to hold over* $\mathbb{C}$ *cannot follow the same line of reasoning*". In this paper, we extend the result in [11] to the complex case by employing a new idea on the RIP of quadratic measurements.

1.3. Our contribution

In this paper, we study the performance of ℓ_1 minimization for recovering complex sparse signals from phaseless measurements $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{w}$, where $\|\mathbf{w}\|_2 \le \epsilon$. Particularly, we focus on the model

$$\min_{\mathbf{x}\in\mathbb{C}^n} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{x}) - \mathbf{y}\|_2 \le \epsilon. \tag{1.2}$$

Although the constrained conditions in (1.2) are non-convex, Many numerical experiments were made to demonstrate empirical success of the proposed algorithms. For example, in [9], Moravec, Romberg, and Baraniuk proposed an iterative projection algorithm to solve the noiseless version of (1.2). Furthermore, the ADM algorithm for solving (1.2) was introduced in [16]. However, there are very few results about the theoretical performance of the model.

Our main idea is to lift (1.2) to recover rank-one and sparse matrices, i.e.,

$$\min_{X\in\mathbb{H}^{n\times n}} \|X\|_1 \quad \text{s.t.} \quad \|\mathcal{A}(X) - \mathbf{y}\|_2 \le \epsilon,\ \operatorname{rank}(X) = 1.$$

Throughout this paper, we use $\mathbb{H}^{n\times n}$ to denote the set of Hermitian $n\times n$-matrices. Moreover, we require that $\mathcal{A}$ satisfies the following restricted isometry property over low-rank and sparse matrices:

Definition 1.1. We say that the map $\mathcal{A} : \mathbb{H}^{n\times n} \to \mathbb{R}^m$ satisfies the restricted isometry property of order (r, k) if there exist positive constants c and C such that the inequality

$$c\|X\|_F \le \frac{1}{m}\|\mathcal{A}(X)\|_1 \le C\|X\|_F \tag{1.3}$$

holds for all $X \in \mathbb{H}^{n\times n}$ with $\operatorname{rank}(X) \le r$ and $\|X\|_{0,2} \le k$.

Throughout this paper, we use $\|X\|_{0,2}$ to denote the number of non-zero rows in X. Since X is Hermitian, we have $\|X\|_{0,2} = \|X^*\|_{0,2}$. We next show that a Gaussian random map $\mathcal{A}$ satisfies the RIP of order $(2, k)$ with high probability provided $m \gtrsim k\log(n/k)$. Here we use $A \gtrsim B$ to denote $A \ge C_0 B$, where $C_0 \in \mathbb{R}_+$ is an absolute constant. The notation $\lesssim$ can be defined similarly.

Theorem 1.2. *Assume that the linear measurement* $\mathcal{A}(\cdot)$ *is defined as*

$$\mathcal{A}(X) = (\mathbf{a}_1^* X \mathbf{a}_1, \ldots, \mathbf{a}_m^* X \mathbf{a}_m),$$

with $\mathbf{a}_j$ *independently taken as complex Gaussian random vectors, i.e.,* $\mathbf{a}_j \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I}_{n\times n}) + \mathcal{N}(0, \frac{1}{2}\mathbf{I}_{n\times n})i$. *If*

$$m \gtrsim k\log(n/k),$$

with probability at least $1 - 2\exp(-c_0 m)$, $\mathcal{A}$ *satisfies the restricted isometry property of order* $(2, k)$, *i.e.*

$$0.12\|X\|_F \le \frac{1}{m}\|\mathcal{A}(X)\|_1 \le 2.45\|X\|_F,$$

for all $X \in \mathbb{H}^{n\times n}$ *with* $\operatorname{rank}(X) \le 2$ *and* $\|X\|_{0,2} \le k$ *(also* $\|X^*\|_{0,2} \le k$*).*

In the next theorem, we show that (1.2) can robustly recover complex k-sparse signals from phaseless measurements provided $\mathcal{A}$ satisfies the restricted isometry property of order $(2, 2ak)$ with $a > 0$ being suitably chosen.

Theorem 1.3. *Assume that $\mathcal{A}(\cdot)$ satisfy the RIP condition of order $(2, 2ak)$ with RIP constant $c, C > 0$ satisfying*

$$c - \frac{4C}{\sqrt{a}} - \frac{C}{a} > 0. \tag{1.4}$$

For any k sparse signals $\mathbf{x}_0 \in \mathbb{C}^n$, the solution to (1.2) $\mathbf{x}^\#$ satisfies

$$\|\mathbf{x}^\#(\mathbf{x}^\#)^* - \mathbf{x}_0\mathbf{x}_0^*\|_F \le C_1 \frac{2\epsilon}{\sqrt{m}}, \tag{1.5}$$

where

$$C_1 = \frac{\frac{1}{a} + \frac{4}{\sqrt{a}} + 1}{c - \frac{4C}{\sqrt{a}} - \frac{C}{a}}.$$

Furthermore, we have

$$\min_{c\in\mathbb{C},|c|=1} \|c \cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \le 2\sqrt{2} C_1 \frac{\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}. \tag{1.6}$$

Remark 1.4. According to Lemma 3.2, it obtains that

$$\min_{c\in\mathbb{C},|c|=1} \|c \cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \le \|\mathbf{x}^\# - \mathbf{x}_0\|_2 \le \sqrt{2} \frac{\|\mathbf{x}^\#(\mathbf{x}^\#)^* - \mathbf{x}_0\mathbf{x}_0^*\|_F}{\|\mathbf{x}_0\|_2} \lesssim \frac{\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}.$$

On the other hand, we have

$$\min_{c\in\mathbb{C},|c|=1} \|c \cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \le \|\mathbf{x}^\# - \mathbf{x}_0\|_2 \le \|\mathbf{x}^\#\|_2 + \|\mathbf{x}_0\|_2 \le \|\mathbf{x}^\#\|_1 + \|\mathbf{x}_0\|_2 \le \|\mathbf{x}_0\|_1 + \|\mathbf{x}_0\|_2.$$

Hence, we obtain that

$$\min_{c\in\mathbb{C},|c|=1} \|c \cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \le \min\left\{2\sqrt{2} C_1 \frac{\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}, \|\mathbf{x}_0\|_2 + \|\mathbf{x}_0\|_1\right\}. \tag{1.7}$$

For the case where $\|\mathbf{x}_0\|_2 + \|\mathbf{x}_0\|_1 \le 2\sqrt{2} C_1 \frac{\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}$, we obtain that

$$\begin{aligned} 2\sqrt{2} C_1 \frac{\epsilon}{\sqrt{m}} &\ge \|\mathbf{x}_0\|_2^2 + \|\mathbf{x}_0\|_2 \|\mathbf{x}_0\|_1 \\ &\ge \|\mathbf{x}_0\|_1^2/k + \|\mathbf{x}_0\|_1^2/\sqrt{k} = \|\mathbf{x}_0\|_1^2(1/k + 1/\sqrt{k}), \end{aligned}$$

which implies $\|\mathbf{x}_0\|_1 \le \sqrt{2\sqrt{2} C_1} \cdot \sqrt{\epsilon} \cdot (k/m)^{1/4}$. Noting that

$$\|\mathbf{x}_0\|_2 + \|\mathbf{x}_0\|_1 \le 2\|\mathbf{x}_0\|_1 \le 2\sqrt{2\sqrt{2} C_1} \cdot \sqrt{\epsilon} \cdot \left(\frac{k}{m}\right)^{1/4},$$

we obtain that

$$\min_{c\in\mathbb{C},|c|=1} \|c\cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \lesssim \min\left\{\frac{\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}, \sqrt{\epsilon}\cdot\left(\frac{k}{m}\right)^{1/4}\right\}.$$

Remark 1.5. For $\|\mathbf{x}_0\|_2 \geq 1$, the error bound $\min_{c\in\mathbb{C},|c|=1} \|c\cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \lesssim \frac{\epsilon}{\sqrt{m}}$ presented in Theorem 1.3 is sharp in the sense that there exists $\mathbf{x}_0 \in \mathbb{C}^n$ and $\mathbf{w}\in\mathbb{R}^m$ so that $\min_{c\in\mathbb{C},|c|=1} \|c\cdot \mathbf{x}^\# - \mathbf{x}_0\|_2 \gtrsim \epsilon/\sqrt{m}$ holds with a positive constant probability. Indeed, take $\mathbf{x}_0 = (1,0,...,0)^T \in \mathbb{R}^n$ and $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{w}$ with $\mathbf{w} = (1,...,1) \in \mathbb{R}^m$. Set $\epsilon = \sqrt{10m}$. Assume that $\mathbf{x}^\#$ is a solution to

$$\min_{\mathbf{x}\in\mathbb{C}^n} \|\mathbf{x}\|_1, \qquad \text{s.t.} \quad \|\mathcal{A}(\mathbf{x}) - \mathbf{y}\|_2 \leq \sqrt{10m}.$$

We claim that $\mathbf{x}^\# = \mathbf{0}$ with probability at least $1/2$, which implies that

$$\|\mathbf{x}^\#(\mathbf{x}^\#)^* - \mathbf{x}_0\mathbf{x}_0^*\|_F = 1 \gtrsim \frac{\epsilon}{\sqrt{m}}$$

holds with probability at least $1/2$. To prove $\mathbf{x}^\# = \mathbf{0}$ with probability at least $1/2$, it is enough to show that $\mathbf{P}\{\|\mathcal{A}(\mathbf{0}) - \mathbf{y}\|_2^2 \leq 10m\} \geq 1/2$. Note that

$$\mathbb{E}(\|\mathcal{A}(\mathbf{0}) - \mathbf{y}\|_2^2) = \mathbb{E}\left(\sum_{j=1}^m (|\mathbf{a}_{j,1}|^2+1)^2\right) = \mathbb{E}\left(\sum_{j=1}^m \left(|\mathbf{a}_{j,1}|^4 + 2|\mathbf{a}_{j,1}|^2+1\right)\right) = 5m.$$

According to the Markov inequality, we obtain that

$$\mathbb{P}\{\|\mathcal{A}(\mathbf{0}) - \mathbf{y}\|_2^2 \leq 10m\} \geq 1 - \frac{5m}{10m} = \frac{1}{2}.$$

Hence $\mathbf{x}^\# = 0$ with probability at least $1/2$.

According to Theorem 1.2, if $\mathbf{a}_j, j = 1,\ldots,m$ are complex Gaussian random vectors, then $\mathcal{A}$ satisfies RIP of order $(2, 2ak)$ with constants $c = 0.12$ and $C = 2.45$ with high probability provided $m \gtrsim 2ak\log(n/2ak)$. To guarantee (1.4) holds, it is enough to require $a > (8C/c)^2$. Therefore, combining Theorem 1.2 and Theorem 1.3 with $\epsilon = 0$, we can obtain the following corollary:

Corollary 1.6. *Suppose that $\mathbf{x}_0 \in \mathbb{C}^n$ is a k-sparse signal. Assume that $A = (\mathbf{a}_1,\ldots,\mathbf{a}_m)^T$ where $\mathbf{a}_j, j = 1,\ldots,m$ is Gaussian random vectors, i.e., $\mathbf{a}_j \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I}_{n\times n}) + \mathcal{N}(0, \frac{1}{2}\mathbf{I}_{n\times n})i$. If $m \gtrsim k\log(n/k)$, then*

$$\operatorname*{argmin}_{\mathbf{x}\in\mathbb{C}^n}\{\|\mathbf{x}\|_1 : |A\mathbf{x}| = |A\mathbf{x}_0|\} \;=\; \tilde{\mathbf{x}}_0$$

holds with probability at least $1 - 2\exp(-c_0 m)$. Here $c_0 > 0$ is an absolute constant.

2. Proof of Theorem 1.2

We first introduce a Bernstein-type inequality which plays a key role in our proof.

Lemma 2.1. *[10] Let $\xi_1,\ldots,\xi_m$ be i.i.d. sub-exponential random variables and $K := \max_j \|\xi_j\|_{\psi_1}$. Then for every $\epsilon > 0$, we have*

$$\mathbb{P}\left(\left|\frac{1}{m}\sum_{j=1}^m \xi_j - \frac{1}{m}\mathbb{E}\left(\sum_{j=1}^m \xi_j\right)\right| \geq \epsilon\right) \leq 2\exp\left(-c_0 m \min\left(\frac{\epsilon^2}{K^2}, \frac{\epsilon}{K}\right)\right),$$

where $c_0 > 0$ *is an absolute constant.*

We next introduce some key lemmas needed to prove Theorem 1.2, and then present the proof of Theorem 1.2.

Lemma 2.2. *Assume* z_1*,* z_2*,* z_3 *and* z_4 *are independently drawn from* $\mathcal{N}(0,1)$*. If* $t \in [-1,0]$*, we have*

$$\mathbb{E}|z_1^2 + z_2^2 + tz_3^2 + tz_4^2| = 2\left(\frac{1+t^2}{1-t}\right).$$

Proof. When $t = 0$, we have $\mathbb{E}|z_1^2 + z_2^2 + tz_3^2 + tz_4^2| = \mathbb{E}|z_1^2 + z_2^2| = 2$. If $t \in [-1,0]$, taking coordinates transformation as $z_1 = \rho_1 \cos\theta$, $z_2 = \rho_1 \sin\theta$, $z_3 = \rho_2 \cos\phi$, and $z_4 = \rho_2 \sin\phi$, we obtain that

$$\begin{aligned}
\mathbb{E}|z_1^2 + z_2^2 + tz_3^2 + tz_4^2| &= \left(\frac{1}{2\pi}\right)^2 \int\limits_{\mathbb{R}^4} |z_1^2 + z_2^2 + tz_3^2 + tz_4^2| \exp\left(-\frac{z_1^2+z_2^2+z_3^2+z_4^2}{2}\right) dz_1dz_2dz_3dz_4 \\
&= \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} d\theta \int_0^{2\pi} d\phi \int_0^\infty \int_0^\infty \rho_1\rho_2|\rho_1^2 + t\rho_2^2| \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 \\
&= \int_0^\infty \int_0^\infty \rho_1\rho_2|\rho_1^2 + t\rho_2^2| \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 \\
&= \int\limits_{\rho_1 > \sqrt{-t}\rho_2} \rho_1\rho_2(\rho_1^2 + t\rho_2^2) \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 \\
&\quad + \int\limits_{\rho_1 \le \sqrt{-t}\rho_2} \rho_1\rho_2\left(-t\rho_2^2 - \rho_1^2\right) \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 \\
&= \frac{2}{1-t} + \frac{2t^2}{1-t} = \frac{2(1+t^2)}{1-t}.
\end{aligned}$$

Here, we evaluate the last integrals as follows:

$$\begin{aligned}
&\int\limits_{\rho_1 > \sqrt{-t}\rho_2} \rho_1\rho_2\left(\rho_1^2 + t\rho_2^2\right) \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 \\
=&\int_0^\infty \rho_2 \exp\left(-\frac{\rho_2^2}{2}\right) d\rho_2 \int_{\sqrt{-t}\rho_2}^\infty \rho_1^3 \exp\left(-\frac{\rho_1^2}{2}\right) d\rho_1 + t\int_0^\infty \rho_2^3 \exp\left(-\frac{\rho_2^2}{2}\right) d\rho_2 \int_{\sqrt{-t}\rho_2}^\infty \rho_1 \exp\left(-\frac{\rho_1^2}{2}\right) d\rho_1 \\
=&\int_0^\infty \rho_2 \exp\left(-\frac{\rho_2^2}{2}\right)\left(-t\rho_2^2 \exp\left(\frac{t\rho_2^2}{2}\right) + 2\exp\left(\frac{t\rho_2^2}{2}\right)\right) d\rho_2 + t\int_0^\infty \rho_2^3 \exp\left(-\frac{\rho_2^2}{2}\right) \exp\left(\frac{t\rho_2^2}{2}\right) d\rho_2 \\
=&2\int_0^\infty \rho_2 \exp\left(-\frac{(1-t)\rho_2^2}{2}\right) d\rho_2 = \frac{2}{1-t}.
\end{aligned}$$

We can use the similar method to obtain

$$\int\limits_{\rho_1 \le \sqrt{-t}\rho_2} \rho_1\rho_2\left(-t\rho_2^2 - \rho_1^2\right) \exp\left(-\frac{\rho_1^2+\rho_2^2}{2}\right) d\rho_1 d\rho_2 = \frac{2t^2}{1-t}. \quad \square$$

Lemma 2.3. *Set*

$$\mathcal{X} := \{X \in \mathbb{H}^{n\times n} \mid \|X\|_F = 1,\ rank(X) \le 2,\ \|X\|_{0,2} \le k\}$$

which is equipped with Frobenius norm. The covering number of $\mathcal{X}$ *at scale* $\epsilon > 0$ *is less than or equal to* $\left(\frac{9\sqrt{2}en}{\epsilon k}\right)^{4k+2}$.

Proof. Note that

$$\mathcal{X} = \{X \in \mathbb{H}^{n\times n} \ :\ X = U\Sigma U^*,\ \Sigma \in \Lambda,\ U \in \mathcal{U}\},$$

where

$$\Lambda = \{\Sigma \in \mathbb{R}^{2\times 2} \ :\ \Sigma = \mathrm{diag}(\lambda_1, \lambda_2),\ \lambda_1^2 + \lambda_2^2 = 1\}$$

and

$$\mathcal{U} = \{U \in \mathbb{C}^{n\times 2} \ :\ U^*U = I,\ \|U\|_{0,2} \le k\} = \cup_{\#T=k}\mathcal{U}_T.$$

Here $T \subset \{1, \dots, n\}$, and

$$\mathcal{U}_T := \{U \in \mathbb{C}^{n\times 2} \ :\ U^*U = I,\ U = U_{T,:}\},$$

where $U_{T,:} \subset \mathbb{C}^{n\times 2}$ is the matrix obtained by keeping the rows of U indexed by T and setting all other rows to zero. Note that $\|U\|_F = \sqrt{2}$ for all $U \in \mathcal{U}_T$ and that the real dimension of $\mathcal{U}_T$ is at most $4k$ for any fixed support T with $\#T = k$. We use Q_T to denote an $\epsilon/3$-net of $\mathcal{U}_T$ with $\#Q_T \le (9\sqrt{2}/\epsilon)^{4k}$. Then $Q_\epsilon := \cup_{\#T=k} Q_T$ is an $\epsilon/3$-net of $\mathcal{U}$ with

$$\#Q_\epsilon \le \left(\frac{en}{k}\right)^k \left(\frac{9\sqrt{2}}{\epsilon}\right)^{4k} \le \left(\frac{9\sqrt{2}en}{\epsilon k}\right)^{4k}.$$

We use Λ_ϵ to denote an $\epsilon/3$-net of Λ with $\#\Lambda_\epsilon \le (9/\epsilon)^2$.

Set

$$\mathcal{N}_\epsilon \ :=\ \{U\Sigma U^* \mid U \in Q_\epsilon,\ \text{and}\ \Sigma \in \Lambda_\epsilon\}.$$

Then for any $X = U\Sigma U^* \in \mathcal{X}$, there exists $U_0\Sigma_0U_0^* \in \mathcal{N}_\epsilon$ with $\|U - U_0\|_F \le \epsilon/3$ and $\|\Sigma - \Sigma_0\|_F \le \epsilon/3$. So, we have

$$\begin{aligned}\|U\Sigma U^* - U_0\Sigma_0U_0^*\|_F &\le \|U\Sigma U^* - U_0\Sigma U^*\|_F + \|U_0\Sigma U^* - U_0\Sigma_0U^*\|_F + \|U_0\Sigma_0U^* - U_0\Sigma_0U_0^*\|_F \\ &\le \|U - U_0\|_F\|\Sigma U^*\| + \|U_0\|\|\Sigma - \Sigma_0\|_F\|U^*\| + \|U_0\Sigma_0\|\|U^* - U_0\|_F \\ &\le \epsilon.\end{aligned}$$

Therefore, $\mathcal{N}_\epsilon$ is an ϵ-net of $\mathcal{X}$ with

$$\#\mathcal{N}_\epsilon \le \#\mathcal{Q}_\epsilon \cdot \#\Lambda_\epsilon \le \left(\frac{9\sqrt{2}en}{\epsilon k}\right)^{4k} (9/\epsilon)^2 \le \left(\frac{9\sqrt{2}en}{\epsilon k}\right)^{4k+2}$$

provided that $n \ge k$ and $\epsilon \le 1$. □

We now have the necessary ingredients to prove Theorem 1.2.

Proof of Theorem 1.2. Without loss of generality, we assume that $\|X\|_F = 1$. We first consider $\mathbb{E}\|\mathcal{A}(X)\|_1$. Noting that $\mathrm{rank}(X) \le 2$ and $\|X\|_F = 1$, we can write X in the form of

$$X = \lambda_1 \mathbf{u}_1\mathbf{u}_1^* + \lambda_2 \mathbf{u}_2\mathbf{u}_2^*,$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ satisfying $\lambda_1^2 + \lambda_2^2 = 1$ and $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{C}^n$ satisfying $\|\mathbf{u}_1\|_2 = \|\mathbf{u}_2\|_2 = 1, \langle \mathbf{u}_1, \mathbf{u}_2\rangle = 0$. Therefore, we obtain that

$$\mathbf{a}_k^* X \mathbf{a}_k = \lambda_1 |\mathbf{u}_1^*\mathbf{a}_k|^2 + \lambda_2 |\mathbf{u}_2^*\mathbf{a}_k|^2,$$

where $\mathbf{u}_1^*\mathbf{a}_k$ and $\mathbf{u}_2^*\mathbf{a}_k$ are independently drawn from $\mathcal{N}(0, \frac{1}{2}) + \mathcal{N}(0, \frac{1}{2})i$. Then

$$\frac{1}{m}\|\mathcal{A}(X)\|_1 = \frac{1}{m}\sum_{j=1}^m \left|\lambda_1|\mathbf{u}_1^*\mathbf{a}_j|^2 + \lambda_2|\mathbf{u}_2^*\mathbf{a}_j|^2\right| = \frac{1}{m}\sum_{j=1}^m \xi_j, \tag{2.1}$$

where the ξ_j are independent copies of the following random variable

$$\xi = \left|\lambda_1 z_1^2 + \lambda_1 z_2^2 + \lambda_2 z_3^2 + \lambda_2 z_4^2\right|$$

where $z_1, z_2, z_3, z_4 \sim \mathcal{N}(0, \frac{1}{2})$ are independent. Without loss of generality, we assume that $|\lambda_1| \ge |\lambda_2|$ and hence $|\lambda_1| \in [\frac{\sqrt{2}}{2}, 1]$. Note that ξ can also be rewritten as

$$\xi = |\lambda_1| \left|z_1^2 + z_2^2 + t z_3^2 + t z_4^2\right| \tag{2.2}$$

with $t := \lambda_2/\lambda_1$ satisfying $|t| \le 1$. Since $\frac{1}{m}\mathbb{E}\|\mathcal{A}(X)\|_1 = \mathbb{E}(\xi)$, we first focus on $\mathbb{E}(\xi)$. According to (2.2), we have

$$\mathbb{E}(\xi) \le |\lambda_1| \mathbb{E}(z_1^2 + z_2^2 + z_3^2 + z_4^2) \le 2, \tag{2.3}$$

as $\mathbb{E}(z_j^2) = \frac{1}{2}$ for $j = 1, \dots, 4$. On the other hand, when $t \ge 0$, we obtain that

$$\mathbb{E}(\xi) \ge |\lambda_1| \mathbb{E}(z_1^2 + z_2^2) \ge \frac{\sqrt{2}}{2}. \tag{2.4}$$

For $t \in [-1, 0]$, Lemma 2.2 (note the missing factor two by the slightly different variances of z_i) shows that

$$\mathbb{E}(\xi) = |\lambda_1| \left(\frac{1+t^2}{1-t}\right) \ge 0.57. \tag{2.5}$$

Combining (2.3), (2.4) and (2.5), we obtain that

$$0.57 \;\le\; \mathbb{E}(\xi) \;\le\; 2.$$

Note that ξ is a sub-exponential variable with $\|\xi\|_{\psi_1} \le \sum_{i=1}^4 \|z_i^2\|_{\psi_1} \le \tilde{c}$, where $\|\cdot\|_{\psi_1} := \sup_{p\ge 1} p^{-1}(\mathbb{E}|\cdot|^p)^{1/p}$ denotes the sub-exponential norm. We set

$$\mathcal{X} := \{X \in \mathbb{H}^{n\times n} \;:\; \|X\|_F = 1,\; \mathrm{rank}(X) \le 2,\; \|X\|_{0,2} \le k\},$$

and use $\mathcal{N}_\epsilon$ to denote an ϵ-net of $\mathcal{X}$ with respect to the Frobenius norm $\|\cdot\|_F$, i.e. for any $X \in \mathcal{X}$, there exists $X_0 \in \mathcal{N}_\epsilon$ such that $\|X - X_0\|_F \le \epsilon$. Based on Lemma 2.1, equality (2.1) and a union bound, we obtain that

$$0.57 - \epsilon_0 \ \le\ \frac{1}{m}\|\mathcal{A}(X_0)\|_1 \ \le\ 2+\epsilon_0, \text{ for all } X_0 \in \mathcal{N}_\epsilon \tag{2.6}$$

holds with probability at least $1 - 2\cdot \#\mathcal{N}_\epsilon \cdot \exp(-\frac{c_0}{16} m\epsilon_0^2)$.

Note that $\mathcal{A}$ is continuous at $X \in \mathcal{X}$ and $\mathcal{X}$ is a compact set. We can set

$$U_{\mathcal{A}} := \max_{X\in\mathcal{X}} \frac{1}{m}\|\mathcal{A}(X)\|_1.$$

For any $X \in \mathcal{X}$, there exists $X_0 \in \mathcal{N}_\epsilon$ such that $\|X - X_0\|_F \le \epsilon$ and $\|X - X_0\|_{0,2} \le k$. Without loss of generality, assume that $\mathrm{supp}(X - X_0) \subset [1:k]\times[1:k]$ where $[1:k] := [1,k]\cap\mathbb{Z}$. Note that $\mathrm{rank}(X - X_0) \le 4$. We can use the eigenvalue decomposition to obtain that $(X - X_0)_{[1:k]\times[1:k]} = U\Sigma U^*$ with $U \in \mathbb{C}^{k\times 4}$, and $\Sigma = \mathrm{diag}(\lambda_1, \dots, \lambda_4)$. Take $\Sigma_1 = \mathrm{diag}(\lambda_1, \lambda_2, 0, 0)$ and $\Sigma_2 = \mathrm{diag}(0, 0, \lambda_3, \lambda_4)$. Then $X - X_0 = X_1 + X_2$ where $X_1 = \begin{bmatrix} U\Sigma_1U^*, & \mathbf{0} \\ \mathbf{0}, & \mathbf{0}\end{bmatrix} \in \mathbb{H}^{n\times n}$ and $X_2 = \begin{bmatrix} U\Sigma_2U^*, & \mathbf{0} \\ \mathbf{0}, & \mathbf{0}\end{bmatrix} \in \mathbb{H}^{n\times n}$. If $X_1 = \mathbf{0}$ or $X_2 = \mathbf{0}$, we have $\mathrm{rank}(X - X_0) \le 2$, and

$$\frac{1}{m}\|\mathcal{A}(X - X_0)\|_1 \le U_{\mathcal{A}}\epsilon.$$

Otherwise, $\frac{X_1}{\|X_1\|_F}, \frac{X_2}{\|X_2\|_F} \in \mathcal{X}$ and $\langle X_1, X_2\rangle = \langle \Sigma_1, \Sigma_2\rangle = 0$. Therefore, we can obtain that

$$\begin{aligned}\frac{1}{m}\|\mathcal{A}(X - X_0)\|_1 &= \frac{1}{m}\|\mathcal{A}(X_1 + X_2)\|_1 \le \frac{1}{m}\|\mathcal{A}(X_1)\|_1 + \frac{1}{m}\|\mathcal{A}(X_2)\|_1 \\ &\le U_{\mathcal{A}}\|X_1\|_F + U_{\mathcal{A}}\|X_2\|_F \le \sqrt{2}U_{\mathcal{A}}\|X_1 + X_2\|_F \le \sqrt{2}U_{\mathcal{A}}\epsilon.\end{aligned}$$

Thus

$$\frac{1}{m}\|\mathcal{A}(X)\|_1 \le \frac{1}{m}\|\mathcal{A}(X_0)\|_1 + \frac{1}{m}\|\mathcal{A}(X - X_0)\|_1 \le 2 + \epsilon_0 + \sqrt{2}U_{\mathcal{A}}\epsilon. \tag{2.7}$$

According to the definition of $U_{\mathcal{A}}$, (2.7) implies $U_{\mathcal{A}} \le 2 + \epsilon_0 + \sqrt{2}U_{\mathcal{A}}\epsilon$ and hence which implies that

$$U_{\mathcal{A}} \le \frac{2+\epsilon_0}{1-\sqrt{2}\epsilon}.$$

We also have

$$\frac{1}{m}\|\mathcal{A}(X)\|_1 \ge \frac{1}{m}\|\mathcal{A}(X_0)\|_1 - \frac{1}{m}\|\mathcal{A}(X - X_0)\|_1 \ge 0.57 - \epsilon_0 - \sqrt{2}U_{\mathcal{A}}\epsilon \ge 0.57 - \epsilon_0 - \sqrt{2}\frac{2+\epsilon_0}{1-\sqrt{2}\epsilon}\epsilon.$$

Hence, we obtain that the following holds with probability at least $1 - 2\cdot\#\mathcal{N}_\epsilon\cdot\exp(-\frac{c_0}{16}m\epsilon_0^2)$

$$\left(0.57 - \epsilon_0 - \sqrt{2}\frac{2+\epsilon_0}{1-\sqrt{2}\epsilon}\epsilon\right)\|X\|_F \le \frac{1}{m}\|\mathcal{A}(X)\|_1 \le \left(\frac{2+\epsilon_0}{1-\sqrt{2}\epsilon}\right)\|X\|_F, \text{ for all } X \in \mathcal{X}.$$

Taking $\epsilon = \epsilon_0 = 0.1$, according to Lemma 2.3, we obtain $\#\mathcal{N}_\epsilon \le \left(\frac{90\sqrt{2}en}{k}\right)^{4k+2}$. Thus when $m \ge O(k\log(en/k))$, we obtain that

$$0.12\|X\|_F \le \frac{1}{m}\|\mathcal{A}(X)\|_1 \le 2.45\|X\|_F, \quad \text{for all } X \in \mathcal{X}$$

holds with probability at least $1 - 2\exp(-cm)$. □

3. Proof of Theorem 1.3

In the following, we will use a technical tool based on results in [3,14] which provides convex k-sparse decompositions of certain signals in space.

Lemma 3.1. *[3,14] Suppose that $\mathbf{v} \in \mathbb{R}^p$ satisfying $\|\mathbf{v}\|_\infty \le \theta$, $\|\mathbf{v}\|_1 \le s\theta$ where $\theta > 0$ and $s \in \mathbb{Z}_+$. Then we have*

$$\mathbf{v} = \sum_{i=1}^N \lambda_i \mathbf{u}_i, \qquad 0 \le \lambda_i \le 1, \qquad \sum_{i=1}^N \lambda_i = 1,$$

where $\mathbf{u}_i$ is s-sparse with $(supp(\mathbf{u}_i)) \subset supp(\mathbf{v})$, and

$$\|\mathbf{u}_i\|_1 = \|\mathbf{v}\|_1, \qquad \|\mathbf{u}_i\|_\infty \le \theta.$$

We also need the following lemma:

Lemma 3.2. *If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$, and $\langle \mathbf{x}, \mathbf{y}\rangle \ge 0$, then*

$$\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2 \ge \frac{1}{2}\|\mathbf{x}\|_2^2\|\mathbf{x} - \mathbf{y}\|_2^2.$$

Similarly, we have

$$\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2 \ge \frac{1}{2}\|\mathbf{y}\|_2^2\|\mathbf{x} - \mathbf{y}\|_2^2.$$

Proof. We set $a := \|\mathbf{x}\|_2,\ b := \|\mathbf{y}\|_2$ and $t := \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}$. A simple calculation shows that

$$\|\mathbf{x}\mathbf{x}^* - \mathbf{y}\mathbf{y}^*\|_F^2 - \frac{1}{2}\|\mathbf{x}\|_2^2\|\mathbf{x} - \mathbf{y}\|_2^2 \ = \ h(a, b, t)$$

where

$$h(a, b, t) := a^4 + b^4 - 2(ab)^2t^2 - \frac{1}{2}a^2(a^2 + b^2 - 2abt).$$

Hence, to this end, it is enough to show that $h(a, b, t) \ge 0$ provided $a, b \ge 0$ and $0 \le t \le 1$. For any fixed a and b, $h(a, b, t)$ achieves the minimum for either $t = 0$ or $t = 1$. For $t = 0$, we have

$$h(a, b, 0) = a^4 + b^4 - \frac{1}{2}a^4 - \frac{1}{2}a^2b^2 = \frac{1}{2}(a^2 - \frac{1}{2}b^2)^2 + \frac{7}{8}b^4 \ge 0. \tag{3.1}$$

When $t = 1$, we have

$$\begin{aligned} h(a, b, 1) &= a^4 + b^4 - \frac{1}{2}a^2(a^2 + b^2) - 2(ab)^2 + a^3b \\ &= (a - b)^2(\frac{1}{2}a^2 + b^2 + 2ab) \ge 0 \end{aligned} \tag{3.2}$$

Combining (3.1) and (3.2), we arrive at the conclusion. □

Now we have enough ingredients to prove Theorem 1.3.

Proof of Theorem 1.3. We assume that $\mathbf{x}^\#$ is a solution to (1.2). Noting $\exp(i\theta)\mathbf{x}^\#$ is also a solution to (1.2) for any $\theta \in \mathbb{R}$, in order to apply Lemma 3.2 in (3.10), we assume that

$$\langle \mathbf{x}^\#, \mathbf{x}_0 \rangle \in \mathbb{R} \qquad \text{and} \qquad \langle \mathbf{x}^\#, \mathbf{x}_0 \rangle \geq 0.$$

We consider the programming

$$\min_{X \in \mathbb{H}^{n\times n}} \|X\|_1 \quad s.t. \quad \|\mathcal{A}(X) - \mathbf{y}\|_2 \leq \epsilon, \ \operatorname{rank}(X) = 1. \tag{3.3}$$

Then a simple observation is that $X^\#$ is the solution to (3.3) if and only if $X^\# = \mathbf{x}^\#(\mathbf{x}^\#)^*$.

Set $X_0 := \mathbf{x}_0\mathbf{x}_0^*$ and $H := X^\# - X_0 = \mathbf{x}^\#(\mathbf{x}^\#)^* - \mathbf{x}_0\mathbf{x}_0^*$. Hence, we have to find an upper bound for $\|H\|_F$. Denote $T_0 = \operatorname{supp}(\mathbf{x}_0)$. Set T_1 as the index set which contains the indices of the ak largest elements of $\mathbf{x}^\#_{T_0^c}$ in magnitude, and T_2 contains the indices of the next ak largest elements, and so on. For simplicity, we set $T_{01} := T_0 \cup T_1$ and $\bar{H} := H_{T_{01},T_{01}}$, where $H_{S,T}$ denotes the sub-matrix of H with the row set S and the column set T. We claim that

$$\|H\|_F \leq \|\bar{H}\|_F + \|H - \bar{H}\|_F \leq \left(\frac{1}{a} + \frac{4}{\sqrt{a}} + 1\right)\|\bar{H}\|_F \leq \frac{\frac{1}{a} + \frac{4}{\sqrt{a}} + 1}{c - \frac{4C}{\sqrt{a}} - \frac{C}{a}} \frac{2\epsilon}{\sqrt{m}}, \tag{3.4}$$

which implies the conclusion (1.5). According to Lemma 3.2, we obtain that

$$\min_{c\in\mathbb{C},|c|=1} \|c\cdot\mathbf{x}^\# - \mathbf{x}_0\|_2 \leq \|\mathbf{x}^\# - \mathbf{x}_0\|_2 \leq \sqrt{2}\|H\|_F/\|\mathbf{x}_0\|_2 \leq \frac{\frac{1}{a} + \frac{4}{\sqrt{a}} + 1}{c - \frac{4C}{\sqrt{a}} - \frac{C}{a}} \frac{2\sqrt{2}\epsilon}{\sqrt{m}\|\mathbf{x}_0\|_2}.$$

We next turn to prove (3.4). The first inequality in (3.4) follows from

$$\|H - \bar{H}\|_F \ \leq \ \left(\frac{1}{a} + \frac{4}{\sqrt{a}}\right)\|\bar{H}\|_F \tag{3.5}$$

and the second inequality follows from

$$\|\bar{H}\|_F \leq \frac{1}{c - \frac{4C}{\sqrt{a}} - \frac{C}{a}} \frac{2\epsilon}{\sqrt{m}}. \tag{3.6}$$

To this end, it is enough to prove (3.5) and (3.6).

Step 1: We first present the proof of (3.5). A simple observation is that

$$\begin{aligned} \|H - \bar{H}\|_F &\leq \sum_{i\geq 2, j\geq 2} \|H_{T_i,T_j}\|_F + \sum_{i=0,1}\sum_{j\geq 2} \|H_{T_i,T_j}\|_F + \sum_{j=0,1}\sum_{i\geq 2} \|H_{T_i,T_j}\|_F \\ &= \sum_{i\geq 2, j\geq 2} \|H_{T_i,T_j}\|_F + 2\sum_{i=0,1}\sum_{j\geq 2} \|H_{T_i,T_j}\|_F. \end{aligned} \tag{3.7}$$

We first consider the first term on the right-hand side of (3.7). Note that

$$\begin{aligned} \sum_{i\geq 2, j\geq 2} \|H_{T_i,T_j}\|_F &= \sum_{i\geq 2, j\geq 2} \|\mathbf{x}^\#_{T_i}\|_2 \cdot \|\mathbf{x}^\#_{T_j}\|_2 = \left(\sum_{i\geq 2} \|\mathbf{x}^\#_{T_i}\|_2\right)^2 \leq \frac{1}{ak}\|\mathbf{x}^\#_{T_0^c}\|_1^2 \\ &= \frac{1}{ak}\|H_{T_0^c,T_0^c}\|_1 \leq \frac{1}{ak}\|H_{T_0,T_0}\|_1 \leq \frac{1}{a}\|H_{T_0,T_0}\|_F \leq \frac{1}{a}\|\bar{H}\|_F. \end{aligned} \tag{3.8}$$

Here, the first inequality follows from $\|\mathbf{x}^{\#}_{T_i}\|_2 \le \|\mathbf{x}^{\#}_{T_{i-1}}\|_1/\sqrt{ak}$, for $i \ge 2$. The second inequality is based on $\|H - H_{T_0,T_0}\|_1 \le \|H_{T_0,T_0}\|_1$. Indeed, according to $\|X^{\#}\|_1 \le \|X_0\|_1$, we have

$$\|H - H_{T_0,T_0}\|_1 = \|X^{\#} - X^{\#}_{T_0,T_0}\|_1 \le \|X_0\|_1 - \|X^{\#}_{T_0,T_0}\|_1 \le \|X_0 - X^{\#}_{T_0,T_0}\|_1 = \|H_{T_0,T_0}\|_1.$$

We next turn to $\sum_{i=0,1}\sum_{j\ge 2}\|H_{T_i,T_j}\|_F$. Re-using $\|\mathbf{x}^{\#}_{T_j}\|_2 \le \|\mathbf{x}^{\#}_{T_{j-1}}\|_1/\sqrt{ak}$, we have, for $i \in \{0,1\}$,

$$\sum_{j\ge 2}\|H_{T_i,T_j}\|_F = \|\mathbf{x}^{\#}_{T_i}\|_2 \cdot \sum_{j\ge 2}\|\mathbf{x}^{\#}_{T_j}\|_2 \le \frac{1}{\sqrt{ak}}\|\mathbf{x}^{\#}_{T_0^c}\|_1\|\mathbf{x}^{\#}_{T_i}\|_2 \le \frac{1}{\sqrt{a}}\|\mathbf{x}^{\#}_{T_i}\|_2\|\mathbf{x}^{\#}_{T_{01}} - \mathbf{x}_0\|_2. \tag{3.9}$$

The last inequality is based on $\|\mathbf{x}^{\#}\|_1 \le \|\mathbf{x}_0\|_1$, which leads to

$$\|\mathbf{x}^{\#}_{T_0^c}\|_1 \le \|\mathbf{x}_0\|_1 - \|\mathbf{x}^{\#}_{T_0}\|_1 \le \|\mathbf{x}^{\#}_{T_0} - \mathbf{x}_0\|_1 \le \sqrt{k}\|\mathbf{x}^{\#}_{T_0} - \mathbf{x}_0\|_2 \le \sqrt{k}\|\mathbf{x}^{\#}_{T_{01}} - \mathbf{x}_0\|_2.$$

Substituting (3.8) and (3.9) into (3.7), we obtain that

$$\begin{aligned}\|H - \bar{H}\|_F &\le \sum_{i\ge 2, j\ge 2}\|H_{T_i,T_j}\|_F + \sum_{i=0,1}\sum_{j\ge 2}\|H_{T_i,T_j}\|_F + \sum_{j=0,1}\sum_{i\ge 2}\|H_{T_i,T_j}\|_F \\ &\le \frac{1}{a}\|\bar{H}\|_F + \frac{2\sqrt{2}}{\sqrt{a}}\|\mathbf{x}^{\#}_{T_{01}}\|_2\|\mathbf{x}^{\#}_{T_{01}} - \mathbf{x}_0\|_2 \le \left(\frac{1}{a} + \frac{4}{\sqrt{a}}\right)\|\bar{H}\|_F,\end{aligned} \tag{3.10}$$

where the second inequality is based on $\|\mathbf{x}^{\#}_{T_0}\|_2 + \|\mathbf{x}^{\#}_{T_1}\|_2 \le \sqrt{2}\|\mathbf{x}^{\#}_{T_{01}}\|_2$, and the third inequality follows from Lemma 3.2.

Step 2: We next prove (3.6). Since

$$\|\mathcal{A}(H)\|_2 \le \|\mathcal{A}(X^{\#}) - \mathbf{y}\|_2 + \|\mathcal{A}(X_0) - \mathbf{y}\|_2 \le 2\epsilon,$$

we have

$$\frac{2\epsilon}{\sqrt{m}} \ge \frac{1}{\sqrt{m}}\|\mathcal{A}(H)\|_2 \ge \frac{1}{m}\|\mathcal{A}(H)\|_1 \ge \frac{1}{m}\|\mathcal{A}(\bar{H})\|_1 - \frac{1}{m}\|\mathcal{A}(H - \bar{H})\|_1. \tag{3.11}$$

In order to get a lower bound of $\frac{1}{m}\|\mathcal{A}(\bar{H})\|_1 - \frac{1}{m}\|\mathcal{A}(H - \bar{H})\|_1$, we bound $\frac{1}{m}\|\mathcal{A}(\bar{H})\|_1$ from below and $\frac{1}{m}\|\mathcal{A}(H - \bar{H})\|_1$ from above. As $\text{rank}(\bar{H}) \le 2$ and $\|\bar{H}\|_{0,2} \le (a+1)k$, we obtain by RIP of $\mathcal{A}$ that

$$\frac{1}{m}\|\mathcal{A}(\bar{H})\|_1 \ge c\|\bar{H}\|_F. \tag{3.12}$$

Since $H - \bar{H}$ can be written as

$$H - \bar{H} = (H_{T_0,T_{01}^c} + H_{T_{01}^c,T_0}) + (H_{T_1,T_{01}^c} + H_{T_{01}^c,T_1}) + H_{T_{01}^c,T_{01}^c},$$

we have

$$\frac{1}{m}\|\mathcal{A}(H - \bar{H})\|_1 \le \frac{1}{m}\|\mathcal{A}(H_{T_0,T_{01}^c} + H_{T_{01}^c,T_0})\|_1 + \frac{1}{m}\|\mathcal{A}(H_{T_1,T_{01}^c} + H_{T_{01}^c,T_1})\|_1 + \frac{1}{m}\|\mathcal{A}(H_{T_{01}^c,T_{01}^c})\|_1. \tag{3.13}$$

According to the RIP condition, for $i \in \{0,1\}$, we have

$$\begin{aligned}\frac{1}{m}\|\mathcal{A}(H_{T_i,T_{01}^c}+H_{T_{01}^c,T_i})\|_1 &\leq \sum_{j\geq 2}\frac{1}{m}\|\mathcal{A}(H_{T_i,T_j}+H_{T_j,T_i})\|_1 \leq \sum_{j\geq 2} C\|H_{T_i,T_j}+H_{T_j,T_i}\|_F \\ &\leq C\sum_{j\geq 2}(\|\mathbf{x}_{T_i}^{\#}(\mathbf{x}_{T_j}^{\#})^*\|_F+\|\mathbf{x}_{T_j}^{\#}(\mathbf{x}_{T_i}^{\#})^*\|_F) = 2C\sum_{j\geq 2}\|\mathbf{x}_{T_i}^{\#}\|_2\|\mathbf{x}_{T_j}^{\#}\|_2 \\ &\leq \frac{2C}{\sqrt{a}}\|\mathbf{x}_{T_i}^{\#}\|_2\|\mathbf{x}_{T_{01}}^{\#}-\mathbf{x}_0\|_2,\end{aligned} \tag{3.14}$$

where the first inequality follows from

$$H_{T_i,T_{01}^c}+H_{T_{01}^c,T_i} = \sum_{j\geq 2}(H_{T_i,T_j}+H_{T_j,T_i}) = \sum_{j\geq 2}(\mathbf{x}_{T_i}^{\#}(\mathbf{x}_{T_j}^{\#})^*+\mathbf{x}_{T_j}^{\#}(\mathbf{x}_{T_i}^{\#})^*)$$

and the last inequality is obtained as in (3.9). To bound $\frac{1}{m}\|\mathcal{A}(H_{T_{01}^c,T_{01}^c})\|_1$, note that

$$H_{T_{01}^c,T_{01}^c} = \mathbf{x}_{T_{01}^c}^{\#}(\mathbf{x}_{T_{01}^c}^{\#})^*$$

with $\|\mathbf{x}_{T_{01}^c}^{\#}\|_\infty \leq \|\mathbf{x}_{T_1}^{\#}\|_1/(ak)$. Set $\theta := \max\{\|\mathbf{x}_{T_1}^{\#}\|_1/(ak), \|\mathbf{x}_{T_{01}^c}^{\#}\|_1/(ak)\}$. We assume that $\Phi := \mathrm{Diag}(Ph(\mathbf{x}_{T_{01}^c}^{\#}))$ is the diagonal matrix with diagonal elements being the phase of $\mathbf{x}_{T_{01}^c}^{\#}$, i.e., $\Phi^{-1}\mathbf{x}_{T_{01}^c}^{\#}$ is a real vector. According to Lemma 3.1, we have

$$\Phi^{-1}\mathbf{x}_{T_{01}^c}^{\#} = \sum_{i=1}^N \lambda_i\mathbf{u}_i, \qquad 0\leq\lambda_i\leq 1, \qquad \sum_{i=1}^N\lambda_i = 1,$$

where $\mathbf{u}_i$ is ak-sparse, and

$$\|\mathbf{u}_i\|_1 = \|\mathbf{x}_{T_{01}^c}^{\#}\|_1, \qquad \|\mathbf{u}_i\|_\infty \leq \theta,$$

which leads to

$$\|\mathbf{u}_i\|_2 \leq \sqrt{\|\mathbf{u}_i\|_1\|\mathbf{u}_i\|_\infty} \leq \sqrt{\theta\|\mathbf{x}_{T_{01}^c}^{\#}\|_1}.$$

If $\theta = \|\mathbf{x}_{T_1}^{\#}\|_1/(ak)$, we have

$$\begin{aligned}\|\mathbf{u}_i\|_2 &\leq \sqrt{\frac{\|\mathbf{x}_{T_1}^{\#}\|_1\|\mathbf{x}_{T_{01}^c}^{\#}\|_1}{ak}} = \sqrt{\frac{\|H_{T_1,T_{01}^c}\|_1}{ak}} \\ &\leq \sqrt{\frac{\|H-H_{T_0,T_0}\|_1}{ak}} \leq \sqrt{\frac{\|H_{T_0,T_0}\|_1}{ak}} \leq \sqrt{\frac{\|H_{T_0,T_0}\|_F}{a}} \leq \sqrt{\frac{\|\bar{H}\|_F}{a}}.\end{aligned}$$

If $\theta = \|\mathbf{x}_{T_{01}^c}^{\#}\|_1/(ak)$, we have

$$\begin{aligned}\|\mathbf{u}_i\|_2 &\leq \sqrt{\frac{\|\mathbf{x}_{T_{01}^c}^{\#}\|_1\|\mathbf{x}_{T_{01}^c}^{\#}\|_1}{ak}} = \sqrt{\frac{\|H_{T_{01}^c,T_{01}^c}\|_1}{ak}} \\ &\leq \sqrt{\frac{\|H-H_{T_0,T_0}\|_1}{ak}} \leq \sqrt{\frac{\|H_{T_0,T_0}\|_1}{ak}} \leq \sqrt{\frac{\|H_{T_0,T_0}\|_F}{a}} \leq \sqrt{\frac{\|\bar{H}\|_F}{a}}.\end{aligned}$$

Thus we can obtain that

$$\|\mathbf{u}_i\|_2 \le \sqrt{\frac{\|\bar{H}\|_F}{a}}, \text{ for } i = 1, \ldots, N. \tag{3.15}$$

Since

$$\begin{aligned} H_{T_{01}^c,T_{01}^c} &= \mathbf{x}^\#_{T_{01}^c}(\mathbf{x}^\#_{T_{01}^c})^* = \left(\sum_{i=1}^N \lambda_i \Phi \mathbf{u}_i\right)\left(\sum_{i=1}^N \lambda_i \Phi \mathbf{u}_i\right)^* \\ &= \sum_{i<j} \lambda_i\lambda_j \Phi(\mathbf{u}_i\mathbf{u}_j^* + \mathbf{u}_j\mathbf{u}_i^*)\Phi^{-1} + \sum_i \lambda_i^2 \Phi \mathbf{u}_i\mathbf{u}_i^*\Phi^{-1}, \end{aligned}$$

based on the RIP condition, we can obtain that

$$\begin{aligned} \frac{1}{m}\|\mathcal{A}(H_{T_{01}^c,T_{01}^c})\|_1 &\le \sum_{i<j} C\lambda_i\lambda_j \|(\mathbf{u}_i\mathbf{u}_j^* + \mathbf{u}_j\mathbf{u}_i^*)\|_F + \sum_i C\lambda_i^2\|\mathbf{u}_i\mathbf{u}_i^*\|_F \\ &\le \sum_{i<j} 2C\lambda_i\lambda_j\|\mathbf{u}_i\|_2\|\mathbf{u}_j\|_2 + \sum_i C\lambda_i^2\|\mathbf{u}_i\|_2^2 \\ &\le C\frac{\|\bar{H}\|_F}{a}\left(\sum_i \lambda_i\right)^2 = C\frac{\|\bar{H}\|_F}{a}, \end{aligned} \tag{3.16}$$

where the third line follows from (3.15). Now combining (3.14) and (3.16), we obtain that

$$\begin{aligned} \frac{1}{m}\|\mathcal{A}(H-\bar{H})\|_1 &\le \frac{1}{m}\|\mathcal{A}(H_{T_0,T_{01}^c} + H_{T_{01}^c,T_0})\|_1 + \frac{1}{m}\|\mathcal{A}(H_{T_1,T_{01}^c} + H_{T_{01}^c,T_1})\|_1 + \frac{1}{m}\|\mathcal{A}(H_{T_{01}^c,T_{01}^c})\|_1 \\ &\le \frac{2C}{\sqrt{a}}\|\mathbf{x}^\#_{T_0}\|_2\|\mathbf{x}^\#_{T_{01}} - \mathbf{x}_0\|_2 + \frac{2C}{\sqrt{a}}\|\mathbf{x}^\#_{T_1}\|_2\|\mathbf{x}^\#_{T_{01}} - \mathbf{x}_0\|_2 + C\frac{\|\bar{H}\|_F}{a} \\ &\le \frac{2\sqrt{2}C}{\sqrt{a}}\|\mathbf{x}^\#_{T_{01}}\|_2\|\mathbf{x}^\#_{T_{01}} - \mathbf{x}_0\|_2 + C\frac{\|\bar{H}\|_F}{a} \\ &\le C\left(\frac{4}{\sqrt{a}} + \frac{1}{a}\right)\|\bar{H}\|_F. \end{aligned} \tag{3.17}$$

The last inequality uses Lemma 3.2. Based on (3.12), (3.17) and (3.11), we obtain that

$$\begin{aligned} \frac{2\epsilon}{\sqrt{m}} &\ge \frac{1}{m}\|\mathcal{A}(\bar{H})\|_1 - \frac{1}{m}\|\mathcal{A}(H-\bar{H})\|_1 \\ &\ge c\|\bar{H}\|_F - C\left(\frac{4}{\sqrt{a}} + \frac{1}{a}\right)\|\bar{H}\|_F = \left(c - \frac{4C}{\sqrt{a}} - \frac{C}{a}\right)\|\bar{H}\|_F. \end{aligned}$$

According to the condition (1.4), it implies that

$$\|\bar{H}\|_F \le \frac{1}{c - \frac{4C}{\sqrt{a}} - \frac{C}{a}}\frac{2\epsilon}{\sqrt{m}}.$$

Thus, we arrive at the conclusion (3.6). □

References

[1] Richard Baraniuk, Mark Davenport, Ronald DeVore, Michael Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (3) (2008) 253–263.

[2] T. Tony Cai, Xiaodong Li, Zongming Ma, Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow, Ann. Stat. 44 (5) (2016) 2221–2251.

[3] T. Tony Cai, Anru Zhang, Sparse representation of a polytope and recovery in sparse signals and low-rank matrices, IEEE Trans. Inf. Theory 60 (1) (2014) 122–132.

[4] Emmanuel J. Candès, Justin Romberg, Terence Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.

[5] Yonina C. Eldar, Shahar Mendelson, Phase retrieval: stability and recovery guarantees, Appl. Comput. Harmon. Anal. 36 (3) (2014) 473–494.

[6] Bing Gao, Yang Wang, Zhiqiang Xu, Stable signal recovery from phaseless measurements, J. Fourier Anal. Appl. 22 (4) (2016) 787–808.

[7] Mark Iwen, Aditya Viswanathan, Yang Wang, Robust sparse phase retrieval made easy, Appl. Comput. Harmon. Anal. 42 (1) (2015) 135–142.

[8] Ran Lu, On the strong restricted isometry property of Bernoulli random matrices, J. Approx. Theory 245 (2019) 1–22.

[9] Matthew L. Moravec, Justin K. Romberg, Richard G. Baraniuk, Compressive phase retrieval, Proc. SPIE 6701 (2007).

[10] R. Vershynin, Introduction to the Non-asymptotic Analysis of Random Matrices, Cambridge University Press, 2010.

[11] Vladislav Voroninski, Zhiqiang Xu, A strong restricted isometry property, with an application to phaseless compressed sensing, Appl. Comput. Harmon. Anal. 40 (2) (2016) 386–395.

[12] Gang Wang, Liang Zhang, Georgios B. Giannakis, Mehmet Akcakaya, Jie Chen, Sparse phase retrieval via truncated amplitude flow, IEEE Trans. Signal Process. 66 (2) (2018) 479–491.

[13] Yang Wang, Zhiqiang Xu, Phase retrieval for sparse signals, Appl. Comput. Harmon. Anal. 37 (3) (2014) 531–544.

[14] Guangwu Xu, Zhiqiang Xu, On the ℓ_1-norm invariant convex k-sparse decomposition of signals, J. Oper. Res. Soc. China 1 (2013) 537–541.

[15] P. Schniter, S. Rangan, Compressive phase retrieval via generalized approximate message passing, in: Proceedings of Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, Oct. 2012.

[16] Zai Yang, Cishen Zhang, Lihua Xie, Robust compressive phase retrieval via L1 minimization with application to image reconstruction, arXiv:1302.0081.