Received: 31 May 2018 | Revised: 25 March 2019 | Accepted: 12 April 2019

DOI: 10.1002/nla.2245

RESEARCH ARTICLE

WILEY

Robust quaternion matrix completion with applications to image inpainting

Zhigang Jia[1,2] | Michael K. Ng[3] | Guang-Jing Song[4]

[1]School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, China

[2]The Key Laboratory of Jiangsu Education Big Data Science and Engineering, Jiangsu Normal University, Xuzhou, China

[3]Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[4]School of Mathematics and Information Sciences, Weifang University, Weifang, China

Correspondence
Michael K. Ng, Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.
Email: mng@math.hkbu.edu.hk

Funding information
National Natural Science Foundation of China, Grant/Award Number: 11771188; Natural Science Foundation of the Jiangsu Higher Education Institutions of China, Grant/Award Number: 18KJA110001; HKRGC, Grant/Award Number: GRF 12302715, 12306616, 12200317 and 12300218; HKBU, Grant/Award Number: RC-ICRS/16-17/03

Summary

In this paper, we study robust quaternion matrix completion and provide a rigorous analysis for provable estimation of quaternion matrix from a random subset of their corrupted entries. In order to generalize the results from real matrix completion to quaternion matrix completion, we derive some new formulas to handle noncommutativity of quaternions. We solve a convex optimization problem, which minimizes a nuclear norm of quaternion matrix that is a convex surrogate for the quaternion matrix rank, and the ℓ_1-norm of sparse quaternion matrix entries. We show that, under incoherence conditions, a quaternion matrix can be recovered exactly with overwhelming probability, provided that its rank is sufficiently small and that the corrupted entries are sparsely located. The quaternion framework can be used to represent red, green, and blue channels of color images. The results of missing/noisy color image pixels as a robust quaternion matrix completion problem are given to show that the performance of the proposed approach is better than that of the testing methods, including image inpainting methods, the tensor-based completion method, and the quaternion completion method using semidefinite programming.

KEYWORDS

color images, convex optimization, low rank, matrix recovery, quaternion

1 | INTRODUCTION

In the last decade, there were a lot of research interests and applications in science and engineering for robust recovery of low-rank matrices from incomplete and/or corrupted entries (see, for instance, other works[1–6]). Among these low-rank recovery problems, there are many problems involved in image processing. Usually, the trick is to stack all the image pixels as column vectors of a matrix, and recovery theory and algorithm are applied to the resulting matrix, which is a low-rank or an approximate low-rank matrix. Theoretically, it has been shown that under some mild assumptions, efficient techniques based on convex programming, to minimize the nuclear norm, as an approximation for the matrix rank, can accurately recover low-rank matrices.[2,4,6–8]

Nowadays, color images appear commonly in many image processing applications. A color image contains red, blue, and green channels. In many image processing applications, color image pixels contain missing values and/or gross errors due to many reasons, including transmission loss, sensor failure, and hardware/software malfunction. In the literature, imaging inpainting methods[9–11] have been developed to fill in missing pixels based on isophote completion. In the work

of Chan and Shen,[9] the authors proposed to use a total variation model for inpainting. In the work of Esedoglu and Shen,[10] the authors proposed an inpainting scheme based on the Mumford–Shah–Euler image model by replacing the embedded straight-line curve model with Euler's elastica. The detailed discussion about image inpainting can also be found in the work of Schönlieb.[11] Note that imaging inpainting models are usually developed for gray-level images. For color images, these inpainting methods are applied to red, green, and blue channels separately, and the resulting color image is combined with the inpainting results from the three color channels.

On the other hand, a color image with red, blue, and green channels can be naturally regarded as a third-order tensor. Each frontal slice of this third-order tensor corresponds to a channel of the color image. In general, we can apply tensor-based completion[12–14] and robust tensor principal component analysis[15–17] to deal with this specific tensor. However, the recovery theory for low-rank tensor completion problems is not well established compared with low-rank matrix completion problems. The main reason is that tensor rank has different definitions in the literature. The CANDECOMP/PARAFAC (CP) decomposition[18,19] approximates a tensor as the sum of rank-one outer products, and the minimal number of such decomposition is defined as the CP-rank. However, computing the CP-rank of a specific tensor is NP-hard in general.[20] Tucker decompositions[21] reveal the algebraic structure in the tensor data with the notion of the rank extended to multilinear rank, which is expressed as a vector of ranks of matrix factors. It is clear that Tucker decompositions cannot offer the best low-rank approximation to a tensor.

In this paper, we make use of quaternion matrices to represent color images and study the problem of robust quaternion matrix completion (QMC). The quaternion approach is to encode the red, green, and blue channel pixel values on the three imaginary parts of a quaternion. The use of quaternion for color image representation has been studied in the literature.[22–27] The main advantage is that color images can be studied and processed holistically as a vector field.[22,26,27] Many gray-scale image processing methods have been extended to color image processing using the quaternion algebra, for instance, Fourier transform,[26] wavelet transform,[28] principal component analysis,[29] independent component analysis,[30] and quaternion dictionary learning algorithm.[31–33] Recently, Han et al.[23] have proposed a QMC model of recovering a color image under a low sampling ratio without noise corruption. Their idea is to transform the QMC problem into a constrained trace minimization problem and then apply the semidefinite programming to solve such completion problem. Their experimental results have shown that the QMC method can recover color images quite well.

The main contribution of this paper is to show that one can obtain an exact recovery of a quaternion matrix with high probability by simply solving a convex program where the objective function is a weighted combination of nuclear norm, serving as a convex surrogate for quaternion matrix rank, and the ℓ_1-norm of sparse quaternion matrix entries. The problem is an extension of the real matrix completion problem stated in the work of Candés et al.[4] to quaternion matrices. Here, some new formulas and technical proofs are derived in the paper to handle noncommutativity of quaternion numbers and matrices (see Sections 3.3 and 3.4). In addition, we remark that our results are different from those in the work of Han et al.[23]: (i) we study robust QMC, that is, both missing values and/or noise corruptions are considered; (ii) theoretical results are established in the paper, and there is no theoretical result studied in the work of Han et al.[23]; (iii) we develop a convex optimization problem to solve the robust QMC problem, whereas a semidefinite programming algorithm is employed to solve the completion problem only in the work of Han et al.,[23] and the computational cost of semidefinite optimization is higher than that of the convex optimization approach. Numerical results of robust color image completion are given to show that the performance of the proposed method is better than that of the testing methods, including image inpainting methods,[10,11] the tensor-based completion method,[13] and the quaternion completion method using semidefinite programming.[23]

The rest of this paper is organized as follows. In Section 2, we begin with a brief review of the quaternion matrix. The notations and some preliminaries of quaternion are also introduced. In Section 3, we give our main theoretical results and provide the idea of the proof of our quaternion matrix recovery theorem. The detailed proofs will be presented in the Appendix. In Section 4, we report numerical results to demonstrate our theoretical results. Finally, concluding remarks are given in Section 5.

2 | PRELIMINARIES OF QUATERNION

Quaternion was introduced by Hamilton.[34] Quaternion has one real part and three imaginary parts given by

$$\mathbf{q} = q_r + q_i\mathbf{i} + q_j\mathbf{j} + q_k\mathbf{k},$$

where $q_r, q_i, q_j, q_k \in \mathbb{R}$ and $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ are three imaginary units satisfying

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \quad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}.$$

$\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the fundamental quaternion units. Here, a symbol of boldface indicates that it is a quaternion number, vector, or matrix. When $q_r = 0$, $\mathbf{q}$ is called a pure quaternion. For simplicity, we denote $\mathbb{Q}$ by a set of quaternion numbers.

Let $\mathbf{p}, \mathbf{q} \in \mathbb{Q}$. The sum $\mathbf{p} + \mathbf{q}$ of $\mathbf{p}$ and $\mathbf{q}$ is $(p_r + q_r) + (p_i + q_i)\mathbf{i} + (p_j + q_j)\mathbf{j} + (p_k + q_k)\mathbf{k}$, and their multiplication $\mathbf{pq}$ is given by

$$(p_r q_r - p_i q_i - p_j q_j - p_k q_k) + (p_r q_i + p_i q_r + p_j q_k - p_k q_j)\mathbf{i} + (p_r q_j + p_j q_r + p_k q_i - p_i q_k)\mathbf{j} + (p_r q_k + p_k q_r + p_i q_j - p_j q_i)\mathbf{k}.$$

The conjugate and modulus of $\mathbf{q}$ are respectively defined by

$$\mathbf{q}^* = q_r - q_i\mathbf{i} - q_j\mathbf{j} - q_k\mathbf{k} \quad \text{and} \quad |\mathbf{q}| = \sqrt{q_r^2 + q_i^2 + q_j^2 + q_k^2}.$$

A color image with the spatial resolution of $n_1 \times n_2$ pixels is represented by an $n_1 \times n_2$ quaternion matrix $\mathbf{A}$ in $\mathbb{Q}^{n_1 \times n_2}$ as follows:

$$\mathbf{A}_{ij} = R_{ij}\mathbf{i} + G_{ij}\mathbf{j} + B_{ij}\mathbf{k}, \quad 1 \le i \le n_1, 1 \le j \le n_2,$$

where R_{ij}, G_{ij}, and B_{ij} are the red, green, and blue pixel values, respectively, at the location (i,j) in the image. Note that quaternion matrix–matrix multiplications can be defined similar to classical matrix–matrix multiplication, except that the multiplication between two quaternion numbers is employed. The identity quaternion matrix $\mathbf{I}$ is the same as the classical identity matrix. The inverse $\mathbf{B}$ of a square quaternion matrix $\mathbf{A}$ exists if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. A square quaternion matrix is unitary if $\mathbf{A}^*\mathbf{A} = \mathbf{AA}^* = \mathbf{I}$, where $\mathbf{A}^*$ is the conjugate transpose of $\mathbf{A}$.

Definition 1 (See the work of Zhang[35])**.**
The maximum number of right linearly independent columns of a quaternion matrix $\mathbf{A} \in \mathbb{Q}^{n_1 \times n_2}$ is called the rank of $\mathbf{A}$.

It is clear that the rank of $\mathbf{A}$ can refer to the maximum number of left linearly independent rows of $\mathbf{A}$. Moreover, for any two invertible quaternion matrices $\mathbf{P}$ and $\mathbf{Q}$ of appropriate sizes, $\mathbf{A}$ and $\mathbf{PAQ}$ have the same rank.

Theorem 1 (See the work of Zhang[35])**.**
Let $\mathbf{A} \in \mathbb{Q}^{n_1 \times n_2}$ be of rank r. Then, there exist two unitary quaternion matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_1}] \in \mathbb{Q}^{n_1 \times n_1}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n_2}] \in \mathbb{Q}^{n_2 \times n_2}$ such that

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^*, \tag{1}$$

where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{n_1 \times n_2}$, and σ_i $(i = 1, \ldots, r)$ are positive singular values of $\mathbf{A}$.

Similar to the real/complex matrix case, the Eckart–Young–Mirsky low-rank approximation theorem[36] can be applied to the case of quaternion matrices. Because of such excellent property, it is expected to be useful in studying and analyzing robust color image completion using quaternions.

The inner product between two quaternion matrices $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Tr}(\mathbf{A}^*\mathbf{B}), \tag{2}$$

where $\mathrm{Tr}(\mathbf{A}^*\mathbf{B})$ denotes the trace of $\mathbf{A}^*\mathbf{B}$. Based on this definition, any quaternion matrix $\mathbf{X} \in \mathbb{Q}^{n_1 \times n_2}$ can be expanded as

$$\mathbf{X} = \sum_{i,j} \mathrm{Re}\left(\left\langle e_i e_j^*, \mathbf{X} \right\rangle\right) e_i e_j^* + \mathrm{Re}\left(\left\langle e_i e_j^*\mathbf{i}, \mathbf{X} \right\rangle\right) e_i e_j^*\mathbf{i} + \mathrm{Re}\left(\left\langle e_i e_j^*\mathbf{j}, \mathbf{X} \right\rangle\right) e_i e_j^*\mathbf{j} + \mathrm{Re}\left(\left\langle e_i e_j^*\mathbf{k}, \mathbf{X} \right\rangle\right) e_i e_j^*\mathbf{k}.$$

This expanded form plays a very important role in the proof of our main results. For detailed results related to the quaternion Hilbert spaces, we refer to the works of Ghiloni et al.[37] and Ng.[38]

The norms of the quaternion vector and matrix used in this paper are defined as follows.

Definition 2.

(1) Let $\mathbf{x} = [\mathbf{x}_i] \in \mathbb{Q}^n$ be a quaternion vector. $\|\mathbf{x}\|_1 := \sum_{i=1}^n |\mathbf{x}_i|$, $\|\mathbf{x}\|_2 := \sqrt{\sum_i |\mathbf{x}_i|^2}$, and $\|\mathbf{x}\|_\infty := \max_{1 \le i \le n} |\mathbf{x}_i|$.
(2) Let $\mathbf{A} = [\mathbf{a}_{ij}] \in \mathbb{Q}^{n_1 \times n_2}$ be a quaternion matrix. The ℓ_1-norm $\|\mathbf{A}\|_1 := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |\mathbf{a}_{ij}|$, the ∞-norm $\|\mathbf{A}\|_\infty := \max_{i,j} |\mathbf{a}_{ij}|$, the F-norm $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |\mathbf{a}_{ij}|^2} := \sqrt{\mathrm{Tr}(\mathbf{A}^*\mathbf{A})}$, the spectral norm $\|\mathbf{A}\| := \max\{\sigma_1, \ldots, \sigma_r\}$, and the nuclear norm $\|\mathbf{A}\|_* := \sum_{i=1}^r \sigma_i$, where $\sigma_1, \ldots, \sigma_r$ are nonzero singular values of $\mathbf{A}$.

The nuclear norm $\|\cdot\|_*$ serves as a convex surrogate for quaternion matrix rank, which is shown in Appendix A.1. We can use it to recover low-rank quaternion matrices exactly.

3 | ROBUST QMC

The robust QMC problem aims to recover a low-rank quaternion matrix from its incomplete and corrupted entries. Mathematically, suppose that we observe a quaternion matrix $\mathbf{X} \in \mathbb{Q}^{n_1 \times n_2}$, which is generated as

$$\mathbf{X} = \mathcal{P}_\Omega(\mathbf{L}_0 + \mathbf{S}_0),$$

where $\mathbf{L}_0 \in \mathbb{Q}^{n_1 \times n_2}$ is the target low-rank matrix, $\mathbf{S}_0 \in \mathbb{Q}^{n_1 \times n_2}$ is sparse and acts as the corruption data, and $\mathcal{P}_\Omega$ is the unitary projection onto the linear space of matrices supported on $\Omega \subseteq [1:n_1] \times [1:n_2]$, defined as

$$(\mathcal{P}_\Omega(\mathbf{Y}))_{ij} = \begin{cases} \mathbf{Y}_{ij}, & (i,j) \in \Omega, \\ 0, & (i,j) \notin \Omega, \end{cases} \quad \text{for any} \quad \mathbf{Y} = [\mathbf{Y}_{ij}] \in \mathbb{Q}^{n_1 \times n_2}.$$

Here, $[1:n_1] \times [1:n_2] := \{(i,j) | i = 1, \ldots, n_1, j = 1, \ldots, n_2\}$. If the singular vectors of $\mathbf{L}_0$ satisfy some incoherent conditions, the rank of $\mathbf{L}_0$ is under a given bound, and $\mathbf{S}_0$ is sufficiently sparse, we can recover $\mathbf{L}_0$ and $\mathbf{S}_0$ exactly with a probability near 1 by solving the following minimization problem:

$$\min_{\mathbf{L} \in \mathbb{Q}^{n_1 \times n_2}, \mathbf{S} \in \mathbb{Q}^{n_1 \times n_2}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{subject to} \quad \mathcal{P}_\Omega(\mathbf{L} + \mathbf{S}) = \mathbf{X}. \tag{3}$$

Problem (3) is an extension of the real matrix completion problem stated in the work of Candés et al.[4] to quaternion matrices. In the paper, some new formulas and technical proofs are derived to handle noncommutativity of quaternion numbers and matrices. This model can be applied to treat more complicated completion problems in color image processing.[23] As an application, we can recover a color image with missing pixels from given observed and corrupted color image pixels.

Let the singular value decomposition (SVD) of the quaternion matrix $\mathbf{L}_0$ of rank r be

$$\mathbf{L}_0 = \mathbf{U} \Sigma_r \mathbf{V}^*,$$

where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{Q}^{n_1 \times r}$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{Q}^{n_2 \times r}$ have unitary columns, and $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ consists of all nonzero singular values of $\mathbf{L}_0$. The right column and left row spaces of $\mathbf{L}_0$ are denoted by $\mathbf{U}$ and $\mathbf{V}$, respectively. Here, we consider the following incoherence conditions on the underlying low-rank quaternion matrix.

Definition 3. Let $\mathbf{U}$ be a subspace of $\mathbb{Q}^n$ of dimension r and $\mathcal{P}_\mathbf{U}$ be the unitary projection onto $\mathbf{U}$. Then, the coherence of $\mathbf{U}$ is defined as

$$\mu(\mathbf{U}) = \frac{n}{r} \max_{1 \le i \le n} \|\mathcal{P}_\mathbf{U} e_i\|_2^2,$$

where e_i is the ith unit vector in $\mathbb{R}^n$.

The incoherence assumptions required are given as follows:

$$\max_i \|\mathbf{U}^* e_i\|^2 \le \frac{\mu r}{n_1}, \quad \max_i \|\mathbf{V}^* e_i\|^2 \le \frac{\mu r}{n_2} \tag{4}$$

and

$$\|\mathbf{U}\mathbf{V}^*\|_\infty \le \sqrt{\frac{\mu r}{n_1 n_2}}, \tag{5}$$

where μ is a positive constant. Let us state the main theorem of this paper as follows. For convenience, we always denote $n_{(1)} = \max\{n_1, n_2\}$ and $n_{(2)} = \min\{n_1, n_2\}$.

Theorem 2. *Suppose $\mathbf{L}_0 \in \mathbb{Q}^{n_1 \times n_2}$ obeys conditions (4) and (5), Ω is uniformly distributed among all sets of cardinality m with $m = \rho n_1 n_2$, and each observed entry is corrupted with probability γ independently of other entries. Then, there exists a numerical constant c such that, with a probability of at least $1 - c n_{(1)}^{-10}$, the solution $\hat{\mathbf{L}}$ of (3) with $\lambda = \frac{1}{\sqrt{\rho n_{(1)}}}$ is exact, that is, $\hat{\mathbf{L}} = \mathbf{L}_0$, provided that*

$$\operatorname{rank}(\mathbf{L}_0) \le \frac{\rho_r n_{(2)}}{\mu (\log n_{(1)})^2} \quad \text{and} \quad \gamma \le \gamma_s.$$

Here, ρ_r and γ_s are positive numerical constants.

The proof of this theorem can be found in the following subsections. According to Theorem 2, we can recover a target quaternion matrix with a probability near 1 from a subset of its entries if they are arbitrarily corrupted when the singular vectors $\mathbf{U}$ and $\mathbf{V}$ are reasonably spread, the upper bound of the rank of quaternion matrix $\mathbf{L}_0$ is $\rho_r n_{(2)}\mu^{-1}(\log n_{(1)})^{-2}$, and the corruption term is sufficiently sparse.

3.1 | The proof of Theorem 3.1

There are several concepts and lemmas to be considered in the proof of Theorem 2 in the following subsections.

3.2 | The sampling scheme

To facilitate our proof, similar to the real case used in the work of Candés et al.,[4] we consider the Bernoulli model as a proxy for uniform sampling. The probability of failure under the uniform sampling is bounded by 2 times the probability of failure under the Bernoulli model. In Theorem 2, we consider $\Omega = \{(i,j)|\delta_{ij} = 1\}$ sampled according to the Bernoulli model, where $\{\delta_{ij}\}_{1\le i\le n_1, 1\le j\le n_2}$ is a sequence of independent and identically distributed (0 or 1) Bernoulli random variables with

$$\text{Prob}(\delta_{ij} = 1) = \rho := \frac{m}{n_1 n_2}, \tag{6}$$

where $m = \mathbb{E}|\Omega|$. Here, $\mathbb{E}$ is the expectation operator.

In (3), the available data are of the form $\mathbf{X} = \mathcal{P}_\Omega(\mathbf{L}_0 + \mathbf{S}_0) = \mathcal{P}_\Omega\mathbf{L}_0 + \mathbf{S}_0'$. There are other two index sets of interest. The first one is $\Gamma \subseteq \Omega$, referring to the locations where data are available and clean, that is, $\mathcal{P}_\Gamma\mathbf{X} = \mathcal{P}_\Gamma\mathbf{L}_0$. Here, $\mathcal{P}_\Gamma$ is the unitary projection onto the linear space of matrices supported on $\Gamma \subseteq \Omega$. The second one is $\Omega' = \Omega/\Gamma$, referring to the locations where observed data are unreliable. The locations of Ω, Γ, and Ω' are characterized by the Bernoulli probability distributions with parameters ρ, $\gamma\rho$, and $(1-\gamma)\rho$, where $0 < \gamma < 1$.

It is interesting to note that a quaternion random variable is equivalent to a 4×1 real random vector. Then, we can get the following lemma, which generalizes the main result in the work of Rudelson,[39] from the real field to the quaternion skew field. The proof can be found in Appendix A.2.

Lemma 1. *Let* $\mathbf{x}$ *be a random vector in* $\mathbb{Q}^n$*, satisfying* $\mathbb{E}(\mathbf{x}\mathbf{x}^*) = \mathbf{I}$*. Let* N *be a positive number, and let* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *be independent copies of* $\mathbf{x}$*. Then, there exists a positive number* c_1 *such that*

$$\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^* - \mathbf{I}\right\| \le c_1\sqrt{\frac{\log n}{N}}(\mathbb{E}\|\mathbf{x}\|^{\log N})^{\frac{1}{\log N}},$$

provided that the right-hand side expression of the above equation is smaller than 1.

3.3 | Subgradients of ℓ_1-norm and nuclear norm of quaternion matrix

In the optimalization model, the ℓ_1-norm and nuclear norm are used. In the analysis of the optimizer, their subgradients are required, and they are new in the literature.

The subgradient of any matrix norm of quaternion matrix $\mathbf{A} \in \mathbb{Q}^{n_1\times n_2}$, denoted by $|||\mathbf{A}|||$, can be given by

$$\partial|||\mathbf{A}||| = \left\{\mathbf{G} \in \mathbb{Q}^{n_1\times n_2} : |||\mathbf{B}||| \ge |||\mathbf{A}||| + \text{Re}(\langle\mathbf{G}, \mathbf{B}-\mathbf{A}\rangle), \text{ for all } \mathbf{B} \in \mathbb{Q}^{n_1\times n_2}\right\}.$$

Here, $\text{Re}(\mathbf{A})$ is the real part of quaternion matrix $\mathbf{A}$.

Lemma 2. *Suppose that* $\mathbf{A} \in \mathbb{Q}^{n_1\times n_2}$*, then the subgradient of the* ℓ_1*-norm* $\|\cdot\|_1$ *at* $\mathbf{A}$ *supported on* Ω *is given by*

$$\partial\|\mathbf{A}\|_1 = \{\mathbf{G} \in \mathbb{Q}^{n_1\times n_2} : \mathbf{G} = \texttt{direc}(\mathbf{A}) + \mathbf{F}, \mathcal{P}_\Omega(\mathbf{F}) = 0, \|\mathbf{F}\|_\infty \le 1\}, \tag{7}$$

where $\texttt{direc}(\mathbf{A})$ *is an* $n_1\times n_2$ *matrix with its entries given by* $\left[\frac{\mathbf{a}_{ij}}{|\mathbf{a}_{ij}|}\right]_{n_1\times n_2}$.

Lemma 3. *Suppose that* $\mathbf{A} \in \mathbb{Q}^{n_1\times n_2}$ *has the SVD as in (1). Then,* $\mathbf{G} \in \mathbb{Q}^{n_1\times n_2}$ *is a subgradient of the nuclear norm at* $\mathbf{A}$ *if*

$$\mathbf{G} = \sum_{1\le k\le r}\mathbf{u}_k\mathbf{v}_k^* + \mathbf{F}, \tag{8}$$

where $\mathbf{F}$ *has the following two properties: (a) the right column space of* $\mathbf{F}$ *is unitary to* $\mathbf{U}$*, and the left row space of* $\mathbf{F}$ *is unitary to* $\mathbf{V}$*; (b)* $\|\mathbf{F}\| \leq 1$.

The proofs of the above two lemmas can be found in Appendices A.3 and A.4, respectively.

3.4 | The real part of projection

For a given matrix $\mathbf{A} \in \mathbb{Q}^{n_1 \times n_2}$ with rank r, there exist two unitary matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r]$ obtained from the SVD of $\mathbf{A}$ in (1). Denote $\mathbb{Q}^{n_1 \times n_2} = \mathbb{T} \oplus \mathbb{T}^{\perp}$, where $\mathbb{T}$ is expressed as follows:

$$\mathbb{T} = \{\mathbf{U}\mathbf{Y}^* + \mathbf{Z}\mathbf{V}^* | \mathbf{Y} \in \mathbb{Q}^{n_2 \times r}, \mathbf{Z} \in \mathbb{Q}^{n_1 \times r}\}. \tag{9}$$

Then, the projection $\mathcal{P}_{\mathbb{T}}$ onto $\mathbb{T}$ is given by

$$\mathcal{P}_{\mathbb{T}}(\mathbf{Z}) = \mathcal{P}_{\mathbf{U}}\mathbf{Z} + \mathbf{Z}\mathcal{P}_{\mathbf{V}} - \mathcal{P}_{\mathbf{U}}\mathbf{Z}\mathcal{P}_{\mathbf{V}},$$

where $\mathcal{P}_{\mathbf{U}}$ and $\mathcal{P}_{\mathbf{V}}$ are the unitary projections onto $\mathbf{U}$ and $\mathbf{V}$, respectively. It is clear that $\mathcal{P}_{\mathbb{T}}$ is a linear operator mapping from a matrix to another matrix. Similarly, we define the projection onto $\mathbb{T}^{\perp}$, the complement orthogonal space of $\mathbb{T}$, by

$$\mathcal{P}_{\mathbb{T}^{\perp}}(\mathbf{Z}) = \mathbf{Z} - \mathcal{P}_{\mathbb{T}}(\mathbf{Z}).$$

We remark that, for arbitrary quaternion matrices $\mathbf{A}$ and $\mathbf{B}$, the projection operator $\mathcal{P}_{\mathbb{T}}$ cannot ensure $\langle \mathcal{P}_{\mathbb{T}}(\mathbf{A}), \mathbf{B}\rangle = \langle \mathbf{A}, \mathcal{P}_{\mathbb{T}}(\mathbf{B})\rangle$. It implies that $\mathcal{P}_{\mathbb{T}}$ is not unitary. However, the real part of $\mathcal{P}_{\mathbb{T}}$ can be used because it is idempotent and satisfies (10) in the following lemma.

Lemma 4. *Suppose that* $\mathbf{A}, \mathbf{B} \in \mathbb{Q}^{n_1 \times n_2}$ *and* $\mathbb{T}$ *is defined as (9). Then,*

$$\mathrm{Re}(\langle \mathbf{A}, \mathcal{P}_{\mathbb{T}}(\mathbf{B})\rangle) = \mathrm{Re}(\langle \mathcal{P}_{\mathbb{T}}(\mathbf{A}), \mathbf{B}\rangle). \tag{10}$$

Let $0 < \rho < 1$, $\mu_0 > 0$, and $\epsilon > 0$. In the sequel, we show with high probability that the following inequalities of random quaternion operators hold.

Lemma 5. *Suppose that* Ω *is sampled from the Bernoulli model with parameter* ρ *and the coherence obeys* $\max\{\mu(\mathbf{U}), \mu(\mathbf{V})\} \leq \mu_0$*. Then, there is a numerical constant c such that*

$$\rho^{-1}\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \rho\mathcal{P}_{\mathbb{T}}\| \leq c\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}} \tag{11}$$

with a probability of at least $1 - 3n_{(1)}^{-\beta}$*, where* $\beta > 1$*, provided that* $c\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}} < 1$.

Lemma 6. *Suppose that* Ω *is sampled from the Bernoulli model with parameter* ρ *and the coherence obeys* $\max\{\mu(\mathbf{U}), \mu(\mathbf{V})\} \leq \mu_0$*. Then, with high probability,*

$$\left\|\mathcal{P}_{\mathbb{T}} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}\right\| \leq \varepsilon, \tag{12}$$

provided that $\rho \geq c^2 \frac{\mu_0 r \log n_{(1)}}{\varepsilon^2 n_{(2)}}$*, where c is a numerical constant as given in Lemma 5.*

Lemma 7. *Suppose that* Ω *is sampled from the Bernoulli model with parameter* ρ*. Then, with high probability,* $\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}\|^2 \leq \rho + \varepsilon$*, provided that* $1 - \rho \geq c^2 \frac{\mu_0 r \log n_{(1)}}{\varepsilon^2 n_{(2)}}$*, where c is a numerical constant as given in Lemma 5 and* ε *is given as in Lemma 6. In particular,* $\|\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}\| \leq \frac{1}{2}$ *holds with high probability if* $\rho + \varepsilon \leq \frac{1}{4}$.

Lemma 8. *Suppose that* Ω *is sampled from the Bernoulli model with parameter* ρ*. For* $\mathbf{Z} \in \mathbb{T}$*, then, with high probability,*

$$\left\|\mathbf{Z} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathbf{Z}\right\|_F \leq \varepsilon\|\mathbf{Z}\|_F,$$

where ε *is given as in Lemma 6.*

Lemma 9. *Suppose that* Ω *is sampled from the Bernoulli model with parameter* ρ*. For* $\mathbf{Z} \in \mathbb{T}$*, then, with high probability,*

$$\left\|\mathbf{Z} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathbf{Z}\right\|_{\infty} \leq \varepsilon\|\mathbf{Z}\|_{\infty}, \tag{13}$$

provided that $\rho \geq c^2 \frac{\mu_0 r \log n_{(1)}}{\varepsilon^2 n_{(2)}}$*, where c and* ε *are defined as in Lemmas 5 and 6, respectively.*

Lemma 10. *Suppose that Ω is sampled from the Bernoulli model with parameter ρ. For $\mathbf{Z} \in \mathbb{T}$, then, with high probability,*

$$\|(I - \rho^{-1}\mathcal{P}_\Omega)\boldsymbol{Z}\| \le c'\sqrt{\frac{n\log n}{\rho}}\|\boldsymbol{Z}\|_\infty,$$

for some small numerical constant $c' > 0$ provided that $\rho > c'\frac{\mu_0 \log n_{(1)}}{n_{(2)}}$.

The proofs of Lemmas 5–10 can be found in Appendix A.5.

3.5 | The key steps

Now, we listed the key steps of the proof of Theorem 2. The first step is to show the existence and uniqueness of the optimal solution of (3).

Theorem 3. *Assume that, for any quaternion matrix $\mathbf{M}$,*

$$\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma^\perp}\mathbf{M}\|_F \le n_{(1)}\|\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Gamma^\perp}\mathbf{M}\|_F$$

and take $\lambda \ge \frac{4}{n_{(1)}}$. Then, $(\mathbf{L}_0, \mathbf{S}_0')$ is the unique solution of (3) if there is a pair $(\mathbf{W}, \mathbf{F})$ satisfying

$$\mathbf{U}\mathbf{V}^* + \mathbf{W} + \mathcal{P}_{\mathbb{T}}\mathbf{D} = \lambda(\mathtt{direc}(\mathbf{S}_0') + \mathbf{F}),$$

with $\mathcal{P}_{\mathbb{T}}\mathbf{W} = 0$, $\|\mathbf{W}\| < \frac{1}{2}$, $\mathcal{P}_{\Gamma^\perp}\mathbf{F} = 0$, $\|\mathbf{F}\|_\infty < \frac{1}{2}$, and $\|\mathcal{P}_{\mathbb{T}}\mathbf{D}\|_F \le n_{(1)}^{-2}$.

Here, $\mathcal{P}_{\Gamma^\perp}$ is defined as the unitary projection onto the linear space of matrices supported on $\Gamma^\perp \subseteq [1:n_1] \times [1:n_2]$. The proof of Theorem 3 can be found in Appendix A.6.

The second step is to check the required assumptions in Theorem 2. By using the results in Lemmas 5–10, we show that these assumptions are valid with high probability (see Proposition 5 in Appendix A.6).

The last step of proving the main results in Theorem 2 is to construct a pair $(\mathbf{Y}^L, \mathbf{W}^S)$ obeying

$$(a)\quad \mathcal{P}_{\Gamma^\perp}\mathbf{Y}^L = 0,\quad (b)\quad \|\mathcal{P}_{\mathbb{T}^\perp}\mathbf{Y}^L\| < \frac{1}{4},\quad (c)\quad \|\mathcal{P}_{\mathbb{T}}\mathbf{Y}^L - \mathbf{U}\mathbf{V}^*\|_F \le n_{(1)}^{-2},$$

$$(d)\quad \|\mathcal{P}_{\Gamma}\mathbf{Y}^L\|_\infty < \frac{\lambda}{4},\quad (e)\quad \mathcal{P}_{\mathbb{T}}\mathbf{W}^S = 0,\quad (f)\quad \mathcal{P}_{\Omega'}\mathbf{W}^S = \lambda\ \ \mathtt{direc}(\mathbf{S}_0'),$$

$$(g)\quad \mathcal{P}_{\Omega^\perp}\mathbf{W}^S = 0,\quad (h)\quad \|\mathbf{W}^S\| \le \frac{1}{4},\quad \text{and}\quad (i)\quad \|\mathcal{P}_{\Gamma}\mathbf{W}^S\|_\infty \le \frac{\lambda}{4}$$

with high probability. Once such a pair of matrices $(\mathbf{Y}^L, \mathbf{W}^S)$ is constructed, we can prove that there exist two matrices $\mathbf{W}$ and $\mathbf{F}$ satisfying

$$\mathbf{Y}^L + \mathbf{W}^S = \lambda\left(\mathtt{direc}\left(\mathbf{S}_0'\right) + \mathbf{F}\right) = \mathbf{U}\mathbf{V}^* + \mathbf{W} + \mathcal{P}_{\mathbb{T}}\mathbf{D},$$

where $\mathbf{D}$ is defined as in Theorem 3. The construction of $\mathbf{Y}^L$ and $\mathbf{W}^S$ as well as the proof can be found in Appendix A.7. Hence, the results of Theorem 2 follow.

4 | NUMERICAL EXAMPLES

In this section, we present numerical results on synthetic data and color image completion to validate our main theorem. All these experiments were performed in MATLAB on a Mac Pro with a 2.4-GHz Intel Core i7 processor and 8-GB 1600-MHz DDR3 memory. Here, we solve the convex optimization problem (3) by using the alternating direction method of multipliers (see the work of Candés et al.[4]). Mainly, there are two subproblems to be solved in the iteration of the alternating direction method of multipliers. The first subproblem is to use the singular value thresholding technique[2] to get a low-rank matrix with the application of the Lanczos method for quaternion SVD.[40] The second subproblem is to use the ℓ_1 thresholding technique to get a sparse component.[41] The detailed algorithm of the alternating direction method of multipliers can be found in the work of Yuan and Yang.[42] In the implementation, we apply the real structure-preserving algorithm of quaternion matrices[43,44] for the computational procedure, which only needs real operations.

4.1 | Synthetic data

We first validate the exact recovery phenomenon in Theorem 2 on randomly generated quaternion matrices. We randomly generate a rank-r $n \times n$ quaternion matrix $\mathbf{L}_0$ as a product $\mathbf{L}_0 = \mathbf{U}_r \Sigma_r \mathbf{V}_r^*$, where $\mathbf{U}_r$ and $\mathbf{V}_r$ are $n \times r$ matrices, $\mathbf{U}_r^* \mathbf{U}_r = \mathbf{I}_r$, $\mathbf{V}_r^* \mathbf{V}_r = \mathbf{I}_r$, and Σ_r is a diagonal $r \times r$ matrix with positive diagonal elements. $\mathbf{E} = E_0 + E_1\mathbf{i} + E_2\mathbf{j} + E_3\mathbf{k}$ is a random noise quaternion matrix, where E_t $(t = 0, 1, 2, 3)$ are independently sampled from a standard uniform distribution. A sparse noise matrix $\mathbf{S}_0$ is generated by randomly choosing a support set Ω' of size ℓ uniformly at random and setting $\mathbf{S}_0 = \mathbf{E}$ on Ω'. The sparsity of $\mathbf{S}_0$ is denoted by $\gamma = \ell/n^2$. Let Ω be the set of observed entries. $\rho = |\Omega|/n^2$ is the percentage of observed entries.

We compare our proposed QMC method with the other two real matrix completion methods.

- The first comparison method (RMC1) is to apply the real matrix completion method to the unfolding matrix of size $n \times 3n[X_1, X_2, X_3]$.
- The second comparison method (RMC2) is to apply the real matrix completion method to each matrix X_1, X_2, and X_3 and then combine the results together.

In QMC, we set the value of $\lambda = \frac{1}{\sqrt{\rho n}}$ given in Theorem 2. In RMC1 and RMC2, we set their corresponding values according to their exact recovery results in the work of Candés and Recht,[3] that is, $\lambda = \frac{1}{\sqrt{3\rho n}}$ and $\lambda = \frac{1}{\sqrt{\rho n}}$, respectively. Set the tolerance $\texttt{tol} = 1.0\text{e}{-8}$. The stopping criteria of the alternating direction method of multipliers are that $\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F / \|\mathbf{X}\|_F \leq \texttt{tol}$ and the maximum number of iterations is 5,000. Table 1 reports the results with $\gamma = 0.05$ and $\rho = 0.95$. We see from the Table that QMC performs very well. Low-rank matrices can be recovered exactly, and the corresponding relative errors are very small. However, both RMC1 and RMC2 cannot recover low-rank matrices. Indeed, their recovered ranks are significantly larger than the actual ranks.

To further validate our theoretical results, we check the recovery ability of our algorithm as a function of $\text{rank}(\mathbf{L}_0)$, fractions of gross corruptions γ, and proportion of observed entries ρ. The data are generated randomly as the above-mentioned. We fix the matrix size to be $n = 50$. We set ρ to be different specified values and vary $\text{rank}(\mathbf{L}_0)/n$ and γ to empirically investigate the probability of recovery success. For each pair $(\text{rank}(\mathbf{L}_0)/n, \gamma)$, we simulate 10 test instances and declare a trial to be successful if the recovered quaternion matrix $\mathbf{L}$ satisfies $\|\mathbf{L} - \mathbf{L}_0\|_F / \|\mathbf{L}_0\|_F \leq \texttt{tol}$. Figures 1

TABLE 1 The comparison of different recovery methods for 5% gross errors and 95% observed entries

n	$\text{rank}(\mathbf{L}_0)$	rank(L)			$\|\mathbf{L} - \mathbf{L}_0\|_F / \|\mathbf{L}_0\|_F$		
		QMC	RMC1	RMC2	QMC	RMC1	RMC2
50	2	2	47	50	2.7478e−07	4.5070e−03	9.0428e−04
	4	4	50	50	4.3244e−07	6.3223e−02	3.4267e−02
	6	6	50	50	5.2683e−07	1.3913e−01	1.2127e−01
	8	8	50	50	6.4845e−07	1.7792e−01	1.8226e−01
	10	10	50	50	9.1343e−07	2.0549e−01	2.2281e−01
100	2	2	80	93	2.0858e−07	7.6474e−07	1.3769e−06
	6	6	100	100	3.6925e−07	8.6860e−03	4.7882e−03
	10	10	100	100	5.6842e−07	9.2288e−02	8.7522e−02
	14	14	100	100	7.1509e−07	1.5172e−01	1.5889e−01
	20	20	100	100	8.8028e−07	2.0072e−01	2.1723e−01

Note. QMC = quaternion matrix completion method; RMC1 = first comparison method; RMC2 = second comparison method.

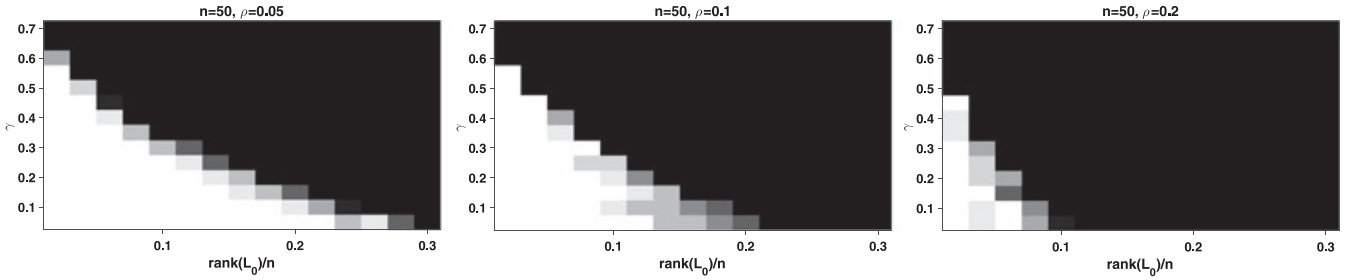

FIGURE 1 Exact recovery for varying $\text{rank}(\mathbf{L}_0)/n$ (x-axis) and gross corruptions γ (y-axis) under different proportions of missing entries

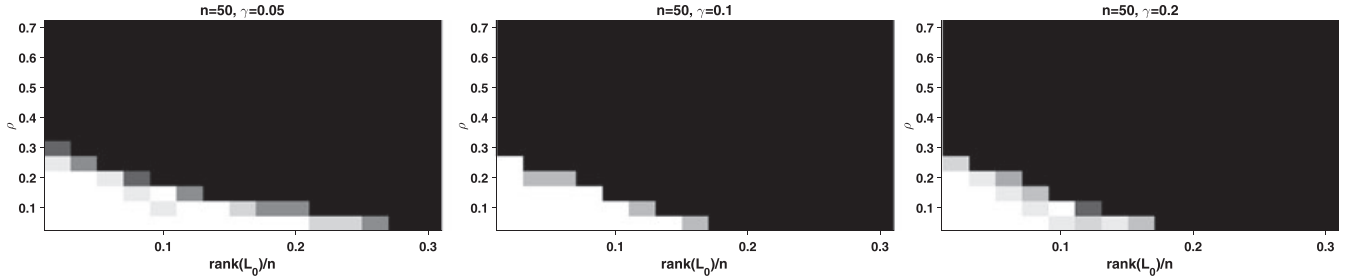

FIGURE 2 Exact recovery for varying $\text{rank}(\mathbf{L}_0)/n$ (x-axis) and missing entries ρ (y-axis)

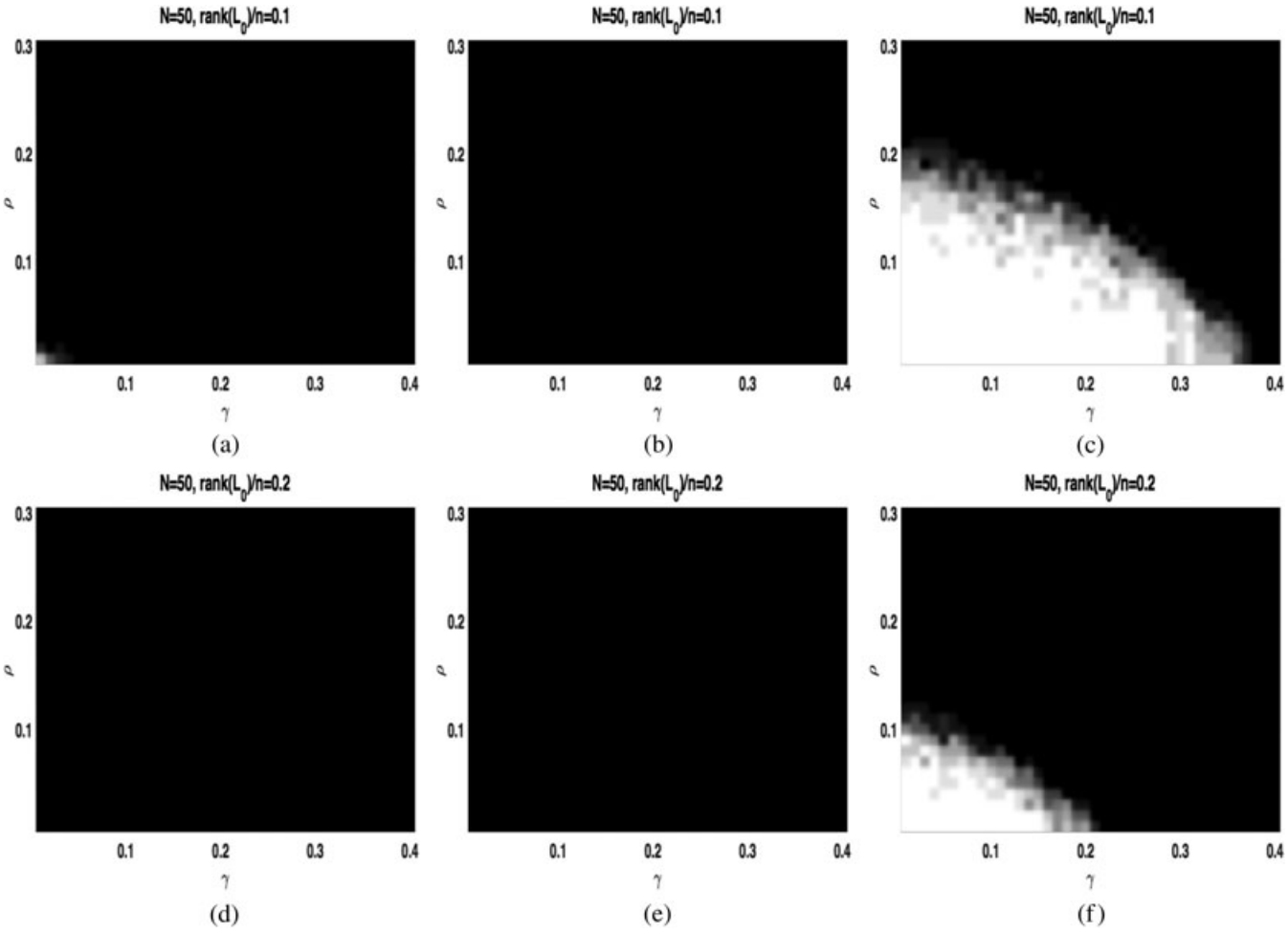

FIGURE 3 Recovery for varying sparsity γ (x-axis) and observed rate ρ (y-axis). (a) First comparison method (RMC1). (b) Second comparison method (RMC2). (c) Quaternion matrix completion method (QMC). (d) RMC1. (e) RMC2. (f) QMC

and 2 report the fraction of perfect recovery for each pair (black = 0% and white = 100%). We see that the smaller the percentage of missing values/the smaller the level of noise is, the larger the region of correct recovery is.

Next, we compare QMC, RMC1, and RMC2 by assuming a Bernoulli model for the support of the sparse term S_0, where each entry of S_0 is corrupted with probability γ. The observed entries are denoted by Ω, and each position appears with probability ρ. For each (γ, ρ) pair, we generate 10 random problems and declare a trial to be successful if the recovered $\mathbf{L}$ satisfies $\|\mathbf{L} - \mathbf{L}_0\|_F / \|\mathbf{L}_0\|_F \leq \texttt{tol}$. Figure 3 plots the fraction of correct recoveries for each pair (γ, ρ). Similarly, we plot different values of $\texttt{rank}((L_0)/n, \gamma)$ and $(\texttt{rank}(L_0)/n, \rho)$ in Figures 4 and 5, respectively. It is clear that the performance of the proposed quaternion method is better than that of the other two methods.

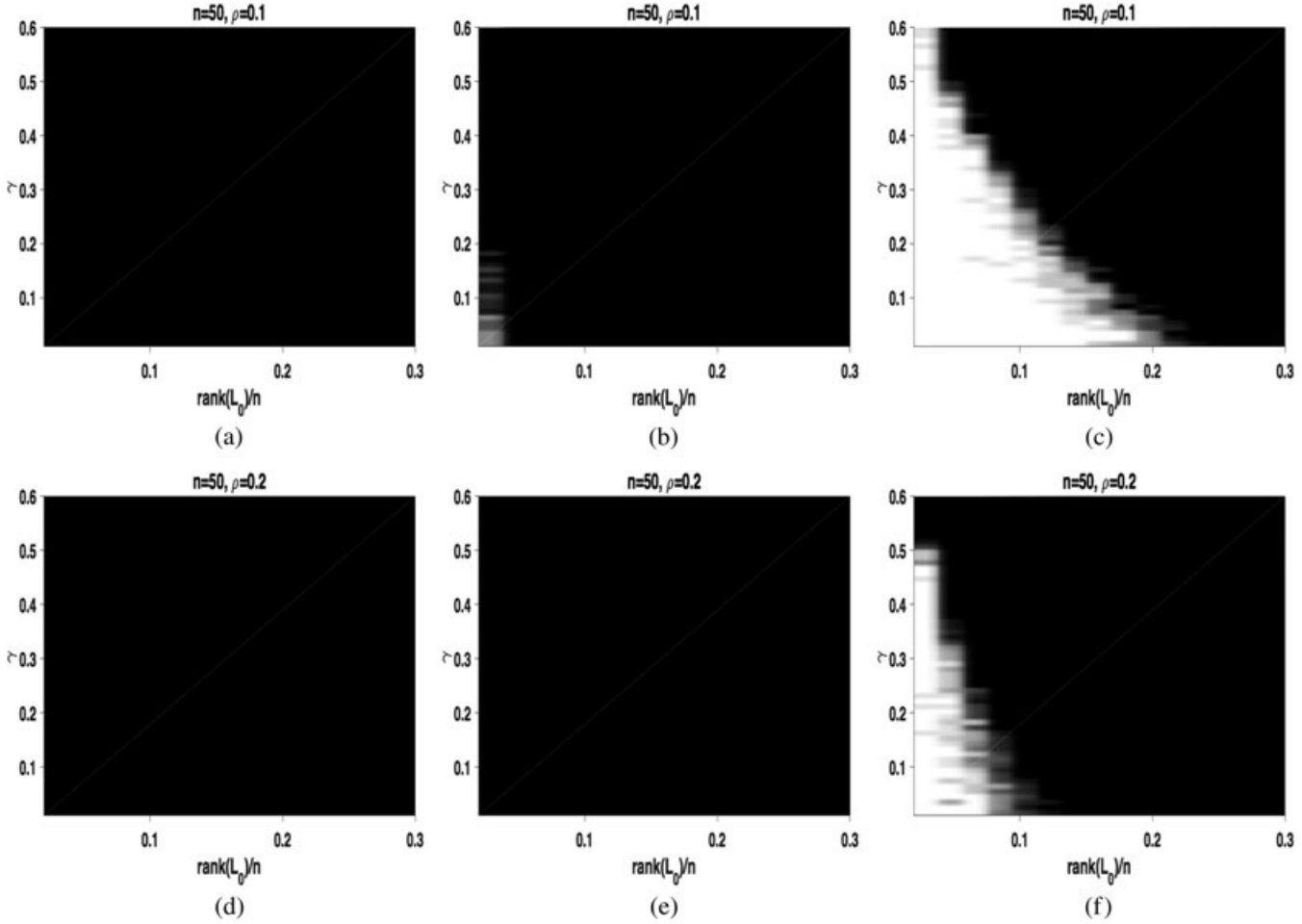

FIGURE 4 Recovery for varying rank $\texttt{rank}(L_0)$ (x-axis) and sparsity γ (y-axis). (a) First comparison method (RMC1). (b) Second comparison method (RMC2). (c) Quaternion matrix completion method (QMC). (d) RMC1. (e) RMC2. (f) QMC

4.2 | Color images

In this experiment, we present color image completion results. The original four color images are shown in Figures 6, 8, 9, and 10.* A standard uniform noise is independently and randomly added into ℓ pixel locations of red, green, and blue channels of color images. The sparsity of noise components is denoted by $\gamma = \ell/(n_1 n_2)$, where the size of the image is $n_1 \times n_2$. Let Ω be the set of observed entries that are generated randomly. $\rho = |\Omega|/(n_1 n_2)$ is the percentage of observed entries.

Similar to the synthetic data case, we compare our proposed QMC method with the other two real matrix completion methods. The first comparison (RMC1) is to apply the real matrix completion method to the unfolding color image of size $n_1 \times 3n_2$, and the second comparison (RMC2) is to apply the real matrix completion method to each color channel and then combine the results together. In QMC, we set the value of $\lambda = \frac{1}{\sqrt{\rho n_{(1)}}}$ given in Theorem 2. In RMC1 and RMC2, we set their corresponding values according to their exact recovery results in the work of Candés and Recht,[3] that is, $\lambda = \frac{1}{\sqrt{3\rho n_{(1)}}}$ and $\lambda = \frac{1}{\sqrt{\rho n_{(1)}}}$, respectively. The stopping criteria of the alternating direction method of multipliers are that the norm of the successive iterates is less than $\texttt{tol}$ and the maximum number of iterations is 5,000.

In Figures 6, 8, 9, and 10, we show the color image completion results for $\rho = 0.2$ and $\gamma = 0.0, 0.1, 0.2$. Both the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are also listed for the comparison of different methods. We find that the performance of PSNR and SSIM by the proposed QMC algorithm is better than those by RMC1 and

*NGC6543a image of size 600 × 600 comes from www.windows2universe.org/the_universe/images/NGC6543a_image.html. It is taken by the Hubble Space Telescope. Pepper image of size 512 × 512 can be used from MATLAB. BSD124084 image of size 321 × 481 is from the Berkeley Segmentation Database.[45] Flower image of size 362 × 500 can be used from MATLAB.

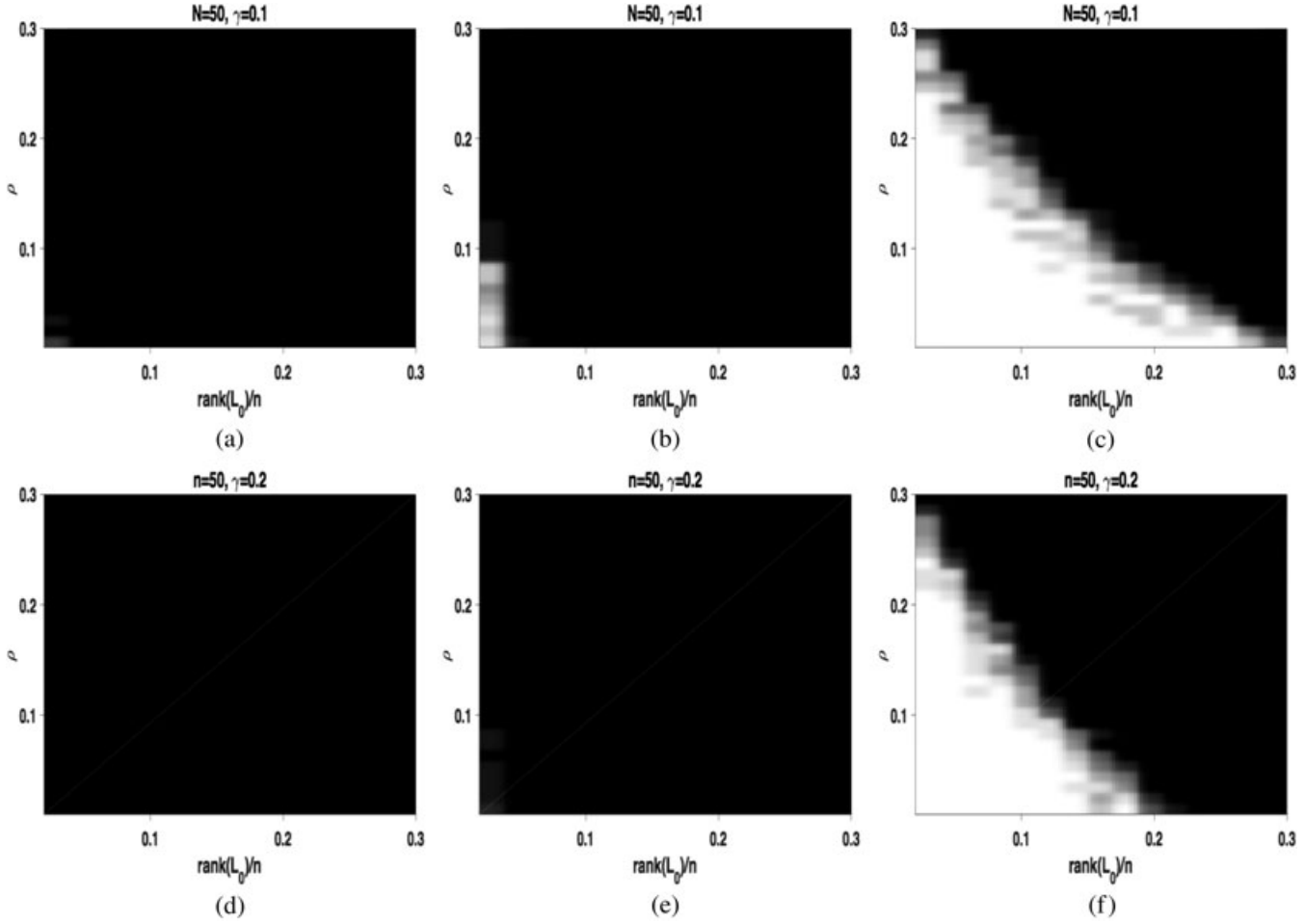

FIGURE 5 Recovery for varying rank $\texttt{rank}(L_0)$ (x-axis) and observed rate ρ (y-axis). (a) First comparison method (RMC1). (b) Second comparison method (RMC2). (c) Quaternion matrix completion method (QMC). (d) RMC1. (e) RMC2. (f) QMC

RMC2. Moreover, we see from the Figures that the color images restored by QMC are visually better than those restored by RMC1 and RMC2. In Figures 7, 11, 12, and 13, corresponding to some zoomed-in regions in Figures 6, 8, 9, and 10, respectively, both the color and the details restored by QMC are visually better than those by RMC1 and RMC2. For instance, the proposed QMC algorithm can provide more clear and detailed restoration in the green stripes in Figure 7, in the boundaries of green and red peppers in Figure 11, in the shapes of red flower (the upper part) and in the small white regions (the lower part) in Figure 12, and in the yellow stigma of flower and the shape of red flower (the upper part) and in the boundary of orange flower (the lower part) in Figure 13.

4.3 | The comparison

In this subsection, we compare the proposed QMC method with image inpainting methods,[10] the tensor-based method,[13] and the recent QMC method by using semidefinite programming.[23] As a comparison, the total variation inpainting model[9,46] is also used to compute the inpainted image. For color images, the image inpainting methods are applied to red, green, and blue channels separately, and the resulting color image is combined with the inpainting results from the three color channels.[11] In the low-rank tensor completion algorithm proposed in the work of Liu et al.,[13] the operator splitting toolbox[47] is employed. The stopping criteria of all these iterative methods are that the norm of the successive iterates is less than `tol` and the maximum number of iterations is 1,000.

The inpainting results are displayed in Figure 14. Figure 14a shows the original color image. Figure 14b shows the observed color image, where 40% color pixels are missed and 1% salt-and-pepper noise is added. Because of the computational requirement, it is very time consuming for semidefinite programming in the quaternion completion method.[23] We choose the block size to be 8×8 for the implementation (see the work of Han et al.[23]). We see from Figure 14 that the

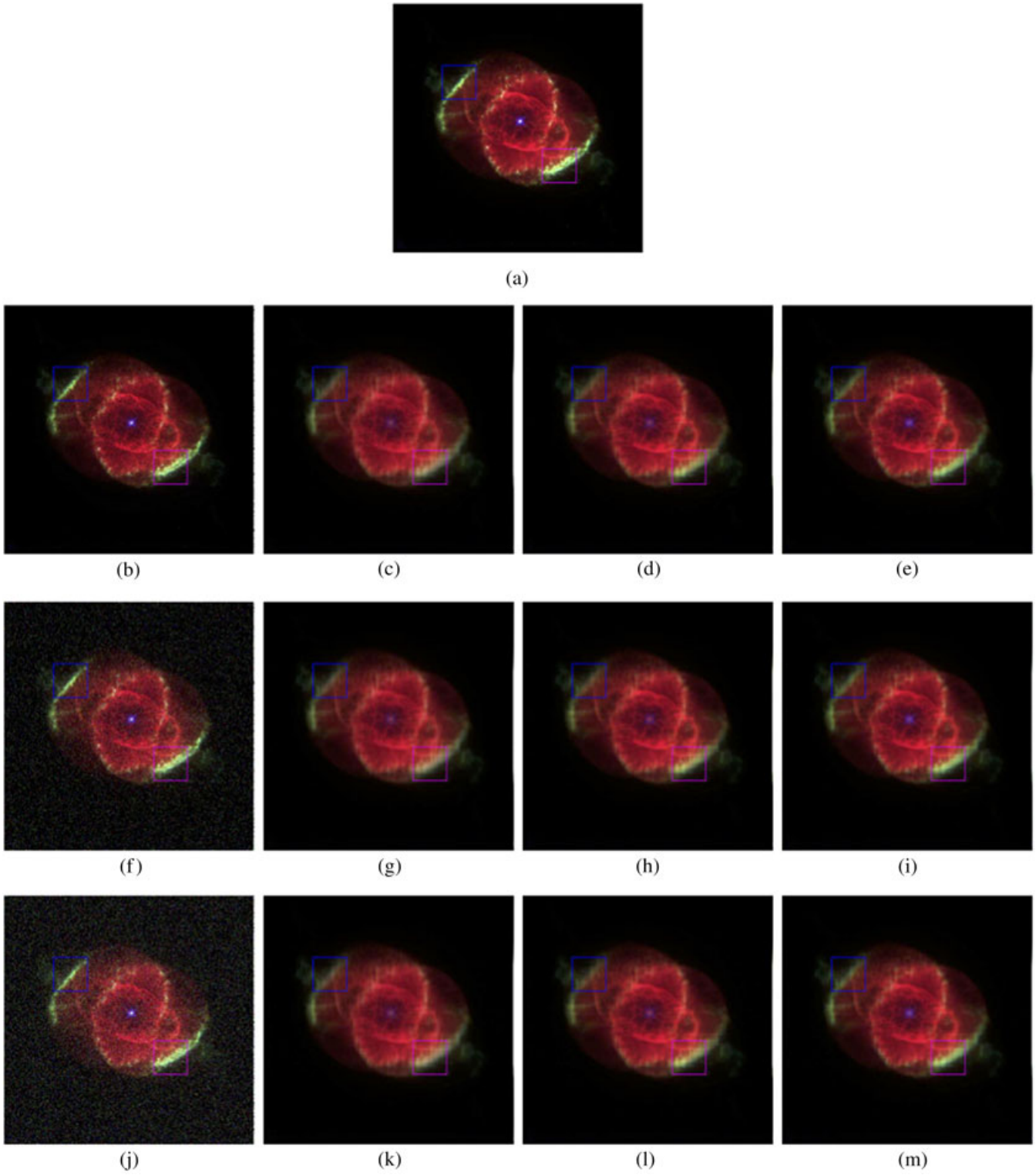

FIGURE 6 Visual comparison of completion results by different methods for the NGC6543a image with $\rho = 0.2$ and $\gamma = 0.0, 0.1, 0.2$ (from top to bottom). Here, the two numbers in each bracket refer to the values of peak signal-to-noise ratio and structural similarity index, respectively. (a) Original. (b) Observed. (c) RMC1(28.96,0.9814). (d) RMC2(29.26,0.9833). (e) QMC(30.20,0.9869). (f) Observed. (g) RMC1(29.05,0.9810). (h) RMC2(29.38,0.9832). (i) QMC(30.22,0.9865). (j) Observed. (k) RMC1(29.07,0.9792). (l) RMC2(29.41,0.9814). (m) QMC(30.17,0.9850). QMC, quaternion matrix completion method; RMC1, first comparison method; RMC2, second comparison method

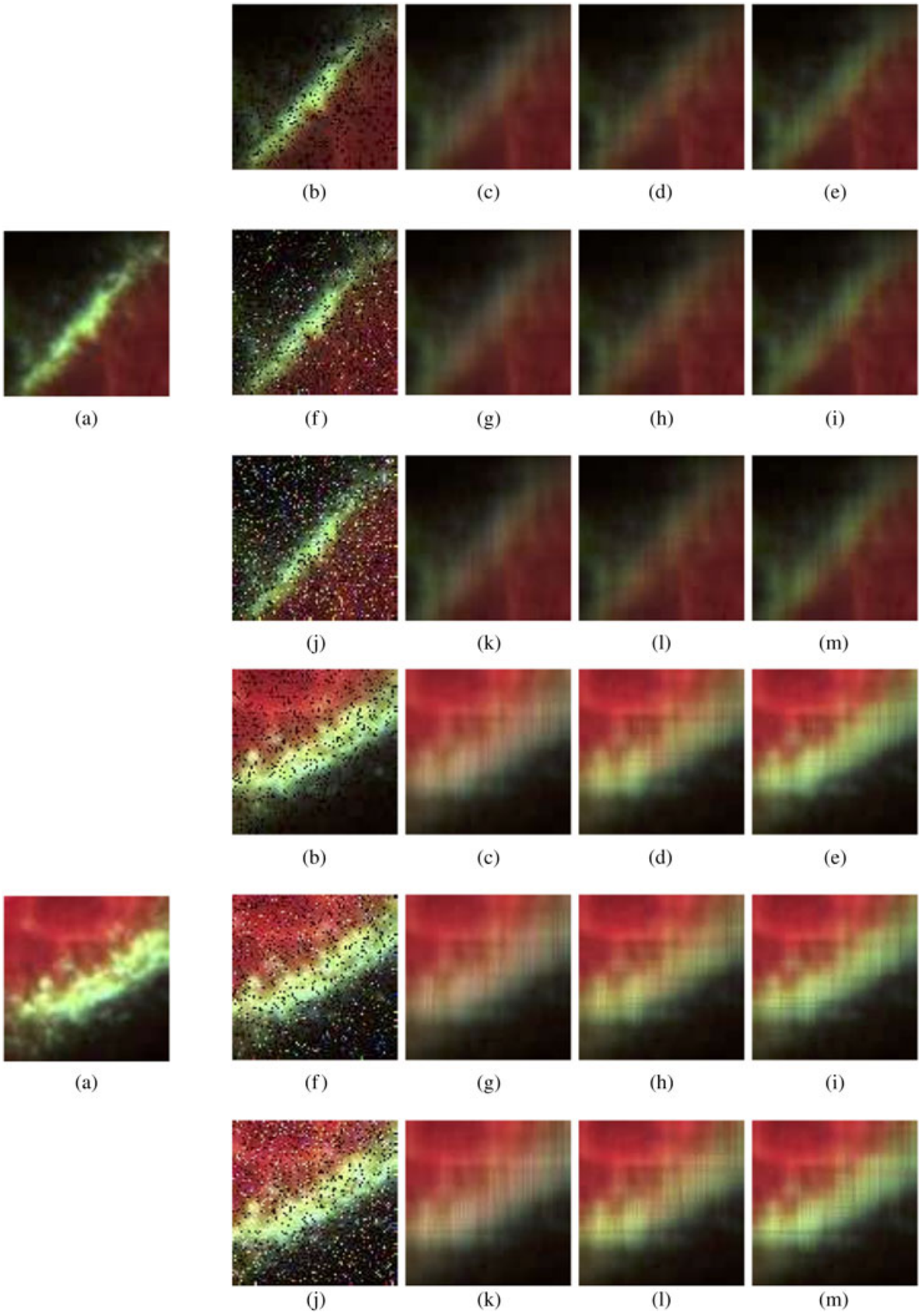

FIGURE 7 Zoomed-in blue part (the upper 13 pictures) and pink part (the lower 13 pictures) of completion results in Figure 6a–m, respectively. (a) Original. (b) Observed. (c) First comparison method (RMC1). (d) Second comparison method (RMC2). (e) Quaternion matrix completion method (QMC). (f) Observed. (g) RMC1. (h) RMC2. (i) QMC. (j) Observed. (k) RMC1. (l) RMC2. (m) QMC

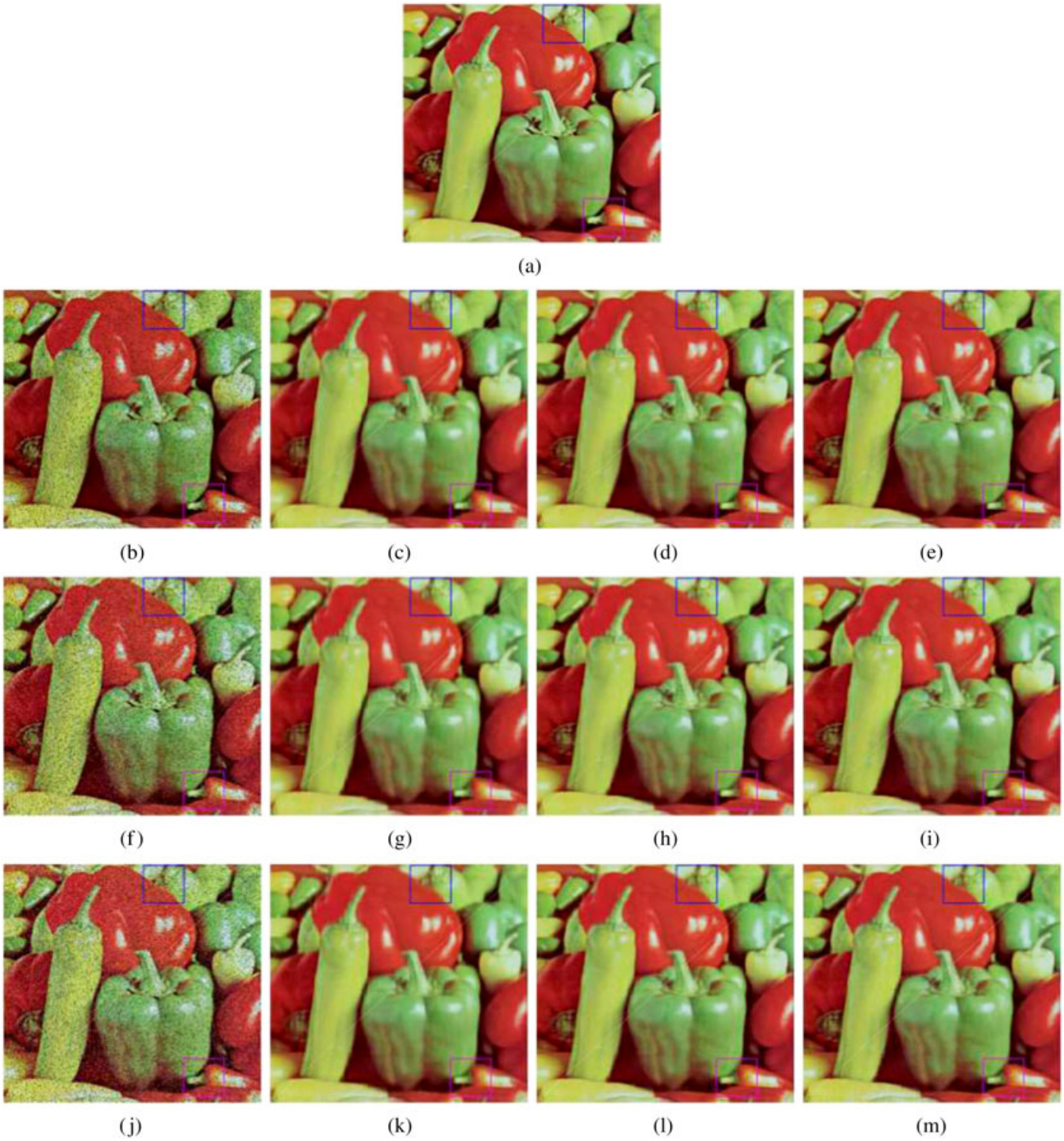

FIGURE 8 Visual comparison of completion results by different methods for the Pepper image with $\rho = 0.2$ and $\gamma = 0.0, 0.1, 0.2$ (from top to bottom). Here, the two numbers in each bracket refer to the values of peak signal-to-noise ratio and structural similarity index, respectively. (a) Original. (b) Observed. (c) RMC1(26.62,0.9620). (d) RMC2(28.20,0.9726). (e) QMC(29.61,0.9807). (f) Observed. (g) RMC1(26.43,0.9602). (h) RMC2(27.82,0.9701). (i) QMC(29.03,0.9772). (j) Observed. (k) RMC1(26.02,0.9553). (l) RMC2(27.26,0.9649). (m) QMC(27.64,0.9652). QMC, quaternion matrix completion method; RMC1, first comparison method; RMC2, second comparison method

FIGURE 9 Visual comparison of completion results by different methods for the BSR124084 image with $\rho = 0.2$ and $\gamma = 0.0, 0.1, 0.2$ (from top to bottom). Here, the two numbers in each bracket refer to the values of peak signal-to-noise ratio and structural similarity index, respectively. (a) Original. (b) Observed. (c) RMC1(25.74,0.9717). (d) RPCA2(26.37,0.9764). (e) QRCIC(26.97,0.9805). (f) Observed. (g) RMC1(25.74,0.9688). (h) RMC2(26.37,0.9733). (i) QMC(26.81,0.9770). (j) Observed. (k) RMC1(25.69,0.9637). (l) RMC2(26.36,0.9677). (m) QMC(26.78,0.9693). QMC, quaternion matrix completion method; RMC1, first comparison method; RMC2, second comparison method

FIGURE 10 Visual comparison of completion results by different methods for the Flower image with $\rho = 0.2$ and $\gamma = 0.0, 0.1, 0.2$ (from top to bottom). Here, the two numbers in each bracket refer to the values of peak signal-to-noise ratio and structural similarity index, respectively. (a) Original. (b) Observed. (c) RMC1(23.02,0.8809). (d) RMC2(23.76,0.8984). (e) QMC(24.61,0.9219). (f) Observed. (g) RMC1(22.90,0.8690). (h) RMC2(23.62,0.8857). (i) QMC(24.26,0.9039). (j) Observed. (k) RMC1(22.76,0.8557). (l) RMC2(23.45,0.8699). (m) QMC(23.89,0.8770). QMC, quaternion matrix completion method; RMC1, first comparison method; RMC2, second comparison method

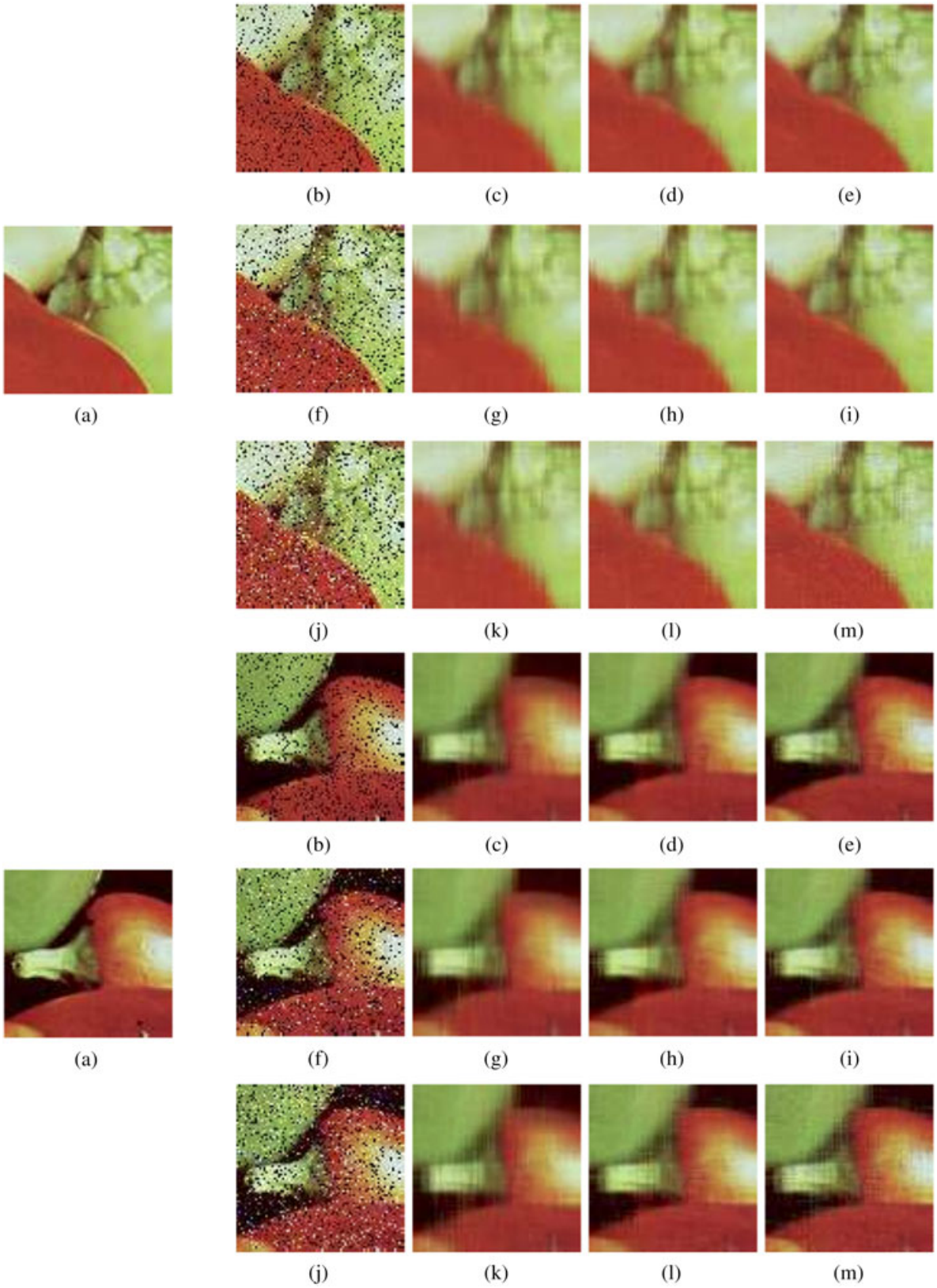

FIGURE 11 Zoomed-in blue part (the upper 13 pictures) and pink part (the lower 13 pictures) of completion results in Figure8a–m, respectively. (a) Original. (b) Observed. (c) First comparison method (RMC1). (d) Second comparison method (RMC2). (e) Quaternion matrix completion method (QMC). (f) Observed. (g) RMC1. (h) RMC2. (i) QMC. (j) Observed. (k) RMC1. (l) RMC2. (m) QMC

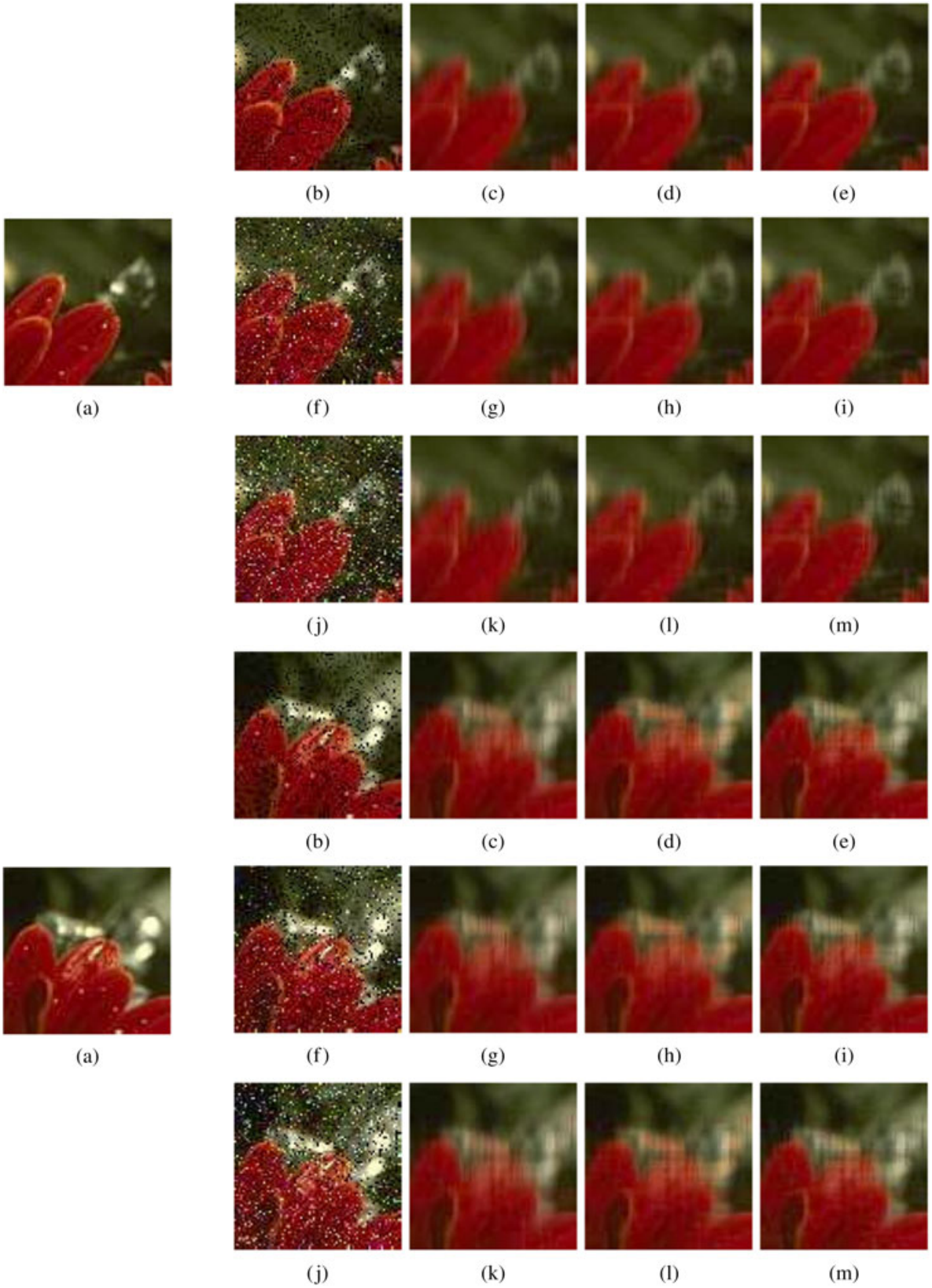

FIGURE 12 Zoomed-in blue part (the upper 13 pictures) and pink part (the lower 13 pictures) of completion results in Figure 9a–m, respectively. (a) Original. (b) Observed. (c) First comparison method (RMC1). (d) Second comparison method (RMC2). (e) Quaternion matrix completion method (QMC). (f) Observed. (g) RMC1. (h) RMC2. (i) QMC. (j) Observed. (k) RMC1. (l) RMC2. (m) QMC

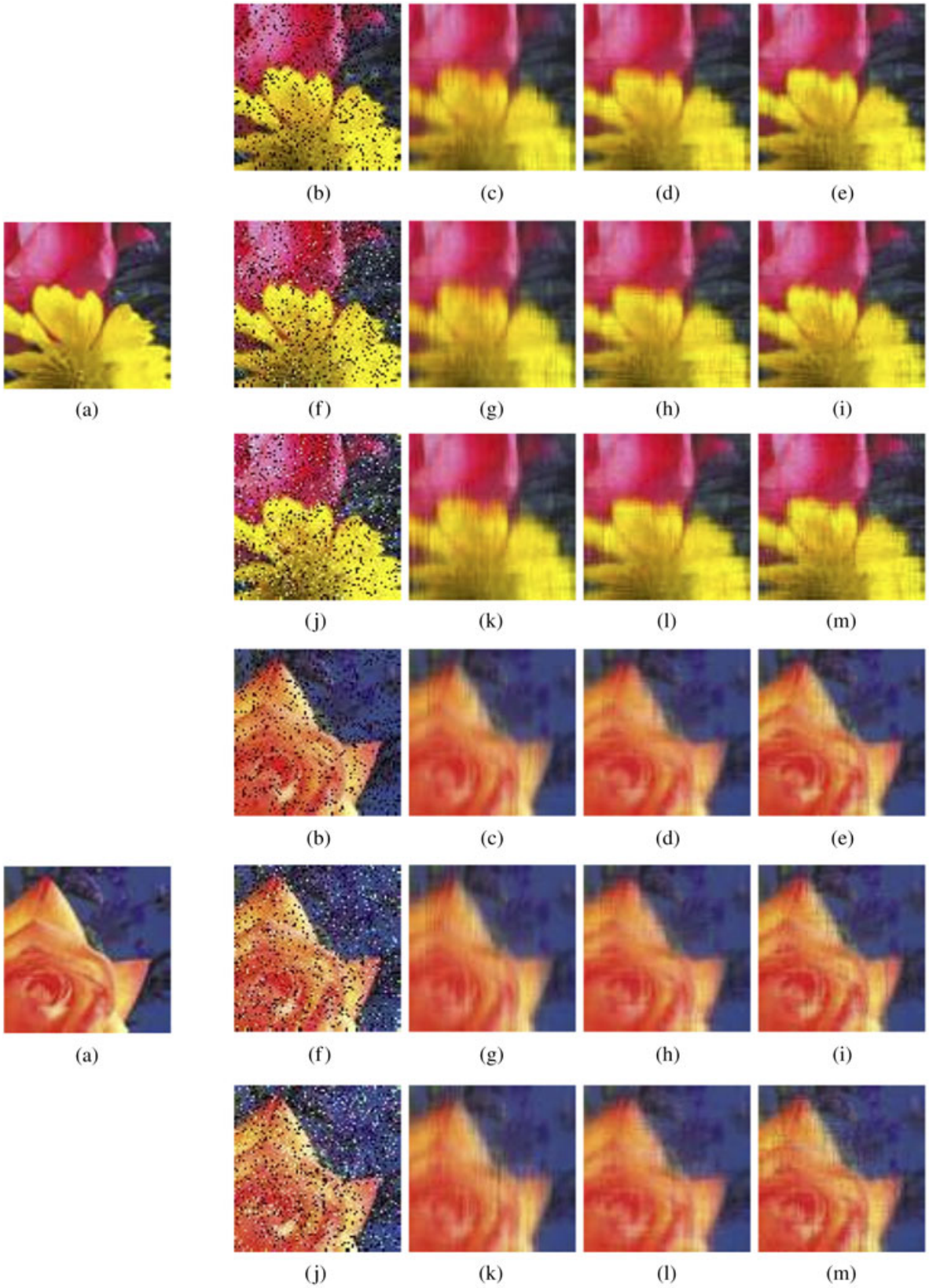

FIGURE 13 Zoomed-in blue part (the upper 13 pictures) and pink part (the lower 13 pictures) of completion results in Figure 10a–m, respectively. (a) Original. (b) Observed. (c) First comparison method (RMC1). (d) Second comparison method (RMC2). (e) Quaternion matrix completion method (QMC). (f) Observed. (g) RMC1. (h) RMC2. (i) QMC. (j) Observed. (k) RMC1. (l) RMC2. (m) QMC

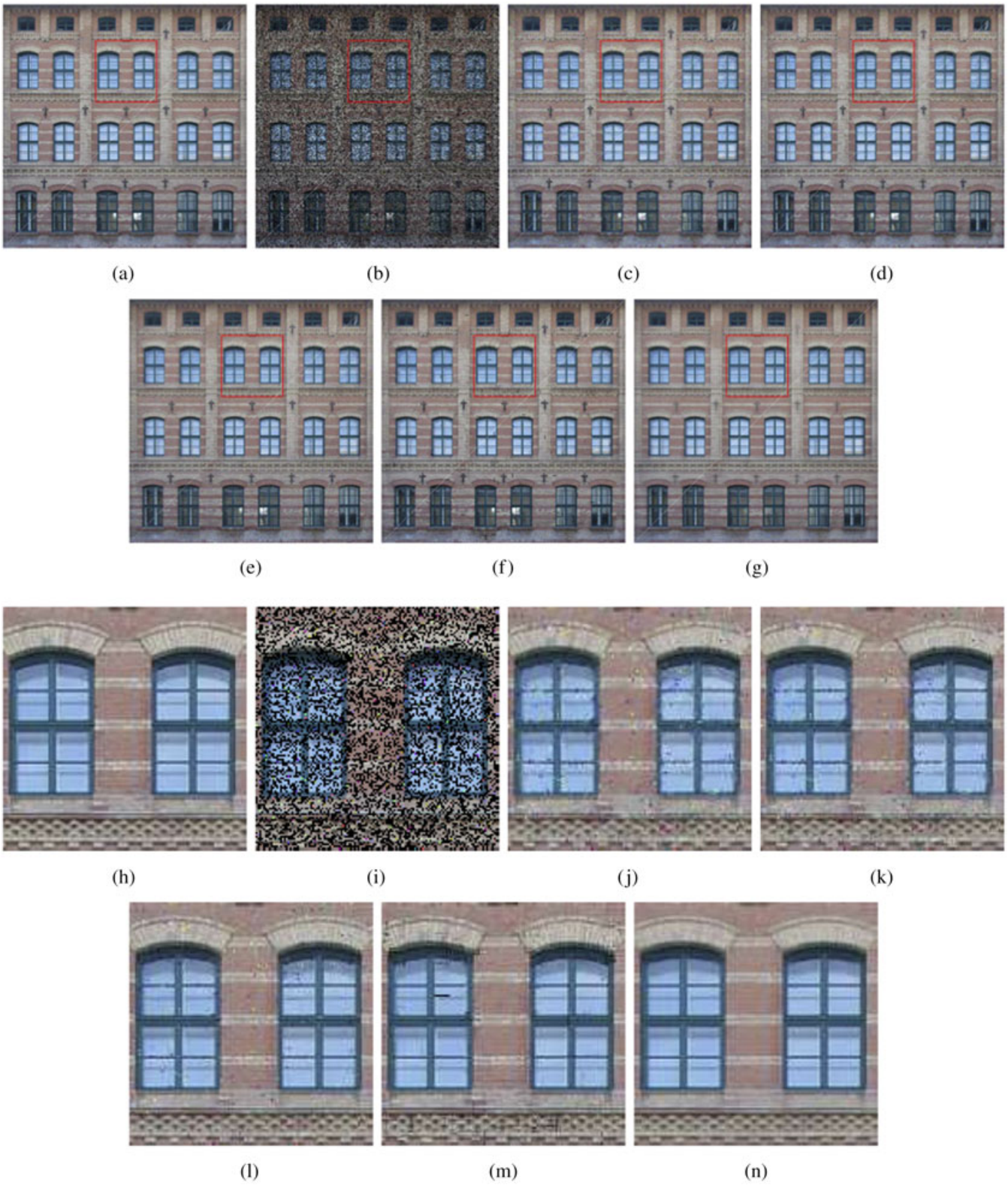

FIGURE 14 The visual comparison. (a) The original color image. (b) The observed color image (40% missing color pixels and 1% salt-and-pepper noise). (c) The Mumford–Shah image inpainting method (MS). (d) The total variation method (TV). (e) The low-rank tensor completion algorithm (TC). (f) The quaternion matrix completion method by semidefinite programming (SDP) with a block size of 8×8. (g) The proposed quaternion matrix completion method (QMC). The two numbers in the bracket refer to the values of peak signal-to-noise ratio and structural similarity index for recovery results. Images (h)–(n) are the zoomed-in parts of (a)–(g), respectively

visual recovery by the proposed quaternion method is better than that by the other testing methods (see Figure 14c–g and their corresponding zoomed-in regions in Figure 14j–n). Moreover, the proposed quaternion method can provide larger PSNR and SSIM values than those by the other testing methods.

5 | CONCLUSION

As a summary, we have studied and analyzed the robust QMC problem. In particular, we have presented a rigorous analysis for provable estimation of quaternion matrix from a random subset of its corrupted entries. Under this quaternion framework, robust color image completion can be solved by a convex optimization problem, which minimizes a nuclear norm of quaternion matrix, which is a convex surrogate for a quaternion matrix rank, and the ℓ_1-norm of sparse quaternion matrix entries. Numerical results have been demonstrated as our theoretical results and shown that the performance of the proposed method for color image completion is better than that of the testing methods using unfolding and performing image completion separately in red, green, and blue channels.

Recently, deep learning approaches have been able to complete images with very large missing parts by using many training examples (see the work of Yu et al.[48] and the references therein). In our future research, we are interested in extending and developing our proposed matrix completion idea to a deep learning framework. We remark that our proposed method in this paper does not use any training example in the image inpainting process. By considering ADMM-Net in the work of Yang et al.,[49] we can employ the QMC in Section 3 and the alternating direction method of multipliers in Section 4 to design a deep learning network for image inpainting by using training examples. We plan to compare this kind of deep network with the other deep learning methods for image inpainting.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under grant 11771188 and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under grant 18KJA110001. HKRGC GRF 12302715, 12306616, 12200317, 12300218, and HKBU RC-ICRS/16-17/03.

CONFLICT OF INTEREST

This work does not have any conflict of interest.

ORCID

Michael K. Ng https://orcid.org/0000-0001-6833-5227

REFERENCES

1. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation. 2003;15(6):1373–1396.
2. Cai J, Candés EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J Optim. 2010;20(4):1956–1982.
3. Candés EJ, Recht B. Exact matrix completion via convex optimization. Found Comput Math. 2009;9(6):717–772.
4. Candés EJ, Li X, Ma Y, Wright J. Robust principal component analysis? J ACM. 2011;58(3). Article No. 11.
5. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. 2000;290(5500):2319–2323.
6. Wright J, Ganesh A, Min K, Ma Y. Compressive principal component pursuit. Inf Inference. 2013;2(1):32–68.
7. Recht B. A simpler approach to matrix completion. J Mach Learn Res. 2011;12(4):3413–3430.
8. Shang F, Liu Y, Cheng J, Cheng H. Robust principal component analysis with missing data. Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management; 2014 Nov 3–7; Shanghai, China. New York, NY: ACM; 2014. p. 1149–1158.
9. Chan T, Shen J. Mathematical models for local non-texture inpainting. SIAM J Appl Math. 2002;62(3):1019–1043.
10. Esedoglu S, Shen J. Digital inpainting based on the Mumford-Shah-Euler image model. Eur J Appl Math. 2002;13(04):353–370.
11. Schönlieb C-B. Partial differential equation methods for image inpainting. Vol. 29. New York, NY: Cambridge University Press; 2015.
12. Gandy S, Recht B, Yamada I. Tensor completion and low-*n*-rank tensor recovery via convex optimization. Inverse Problems. 2011;27(2):025010.

13. Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. IEEE Trans Pattern Anal Mach Intell. 2013;35(1):208–220.
14. Tomioka R, Hayashi K, Kashima H. Estimation of low-rank tensors via convex optimization. arXiv preprint arXiv:1010.0789. 2010.
15. Goldfarb D, Qin Z. Robust low-rank tensor recovery: models and algorithms. SIAM J Matrix Anal Appl. 2013;35(1):225–253.
16. Jiang JQ, Ng MK. Exact tensor completion from sparsely corrupted observations via convex optimization. https://arxiv.org/abs/1708.00601. 2017.
17. Li Y, Yan J, Zhou Y, Yang J. Optimum subspace learning and error correction for tensors. Proceedings of the European Conference on Computer Vision; 2010 Sep 5–11; Heraklion, Greece. Berlin, Germany: Springer; 2010. p. 790–803.
18. Carroll J, Chang J. Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition. Psychometrika. 1970;35(3):283–319.
19. Harshman RA. Foundations of the parafac procedure: Models and conditions for an explanatory multimodal factor analysis. Los Angeles, CA: University of California; 1970.
20. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Review. 2009;51(3):455–500.
21. Tucker L. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;31(3):279–311.
22. Ell TA, Sangwine SJ. Hypercomplex Fourier transforms of color images. IEEE Trans Image Process. 2007;16(1):22–35.
23. Han X, Wu J, Yan L, Senhadji L, Shu H. Color image recovery via quaternion matrix completion. Proceedings of the 6th International Congress on Image and Signal Processing (CISP); 2013 Dec 16–18; Hangzhou, China. Piscataway, NJ: IEEE; 2013. p. 358–362.
24. Le Bihan N, Mars J. Singular value decomposition of quaternion matrices: a new tool for vector-sensor signal processing. Signal Processing. 2004;84(7):1177–1199.
25. Pei S, Chang J, Ding J. Quaternion matrix singular value decomposition and its applications for color image processing. Proceedings of the International Conference on Image Processing; 2003 Sep 14–17; Barcelona, Spain. Piscataway, NJ: IEEE; 2003. p. 805–808.
26. Sangwine SJ. Fourier transforms of colour images using quaternion or hypercomplex numbers. Electronics Letters. 1996;32(21):1979–1980.
27. Subakan ON, Vemuri BC. A quaternion framework for color image smoothing and segmentation. Int J Comput Vis. 2011;91(3):233–250.
28. Fletcher P, Sangwine SJ. The development of the quaternion wavelet transform. Signal Processing. 2017;136:2–15.
29. Zeng R, Wu J, Shao Z, et al. Color image classification via quaternion principal component analysis network. Neurocomputing. 2016;216:416–428.
30. Via J, Vielva L, Santamaria I, Palomar DP. Independent component analysis of quaternion Gaussian vectors. Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop; 2010 Oct 4–7; Jerusalem, Israel. Piscataway, NJ: IEEE; 2010. p. 145–148.
31. Barthlémy Q, Larue A, Mars JI. Color sparse representations for image processing: review, models, and prospects. IEEE Trans Image Process. 2015;24(11):3978–3989.
32. Xu Y, Yu L, Xu H, Zhang H, Nguyen T. Vector sparse representation of color image using quaternion matrix analysis. IEEE Trans Image Process. 2015;24(4):1315–1329.
33. Yu L, Xu Y, Xu H, Zhang H. Quaternion-based sparse representation of color image. Proceedings of the IEEE International Conference on Multimedia and Expo (ICME); 2013 Jul 15–19; San Jose, CA. Piscataway, NJ: IEEE; 2013. p. 1–7.
34. Hamilton WR. Elements of quaternion. London, UK: Longmans, Green, & Co; 1866.
35. Zhang F. Quaternions and matrices of quaternions. Linear Algebra Appl. 1997;251:21–57.
36. Golub GH, Van Loan CF. Matrix computation. 4th ed. Baltimore, MD: The Johns Hopkins University Press; 2013.
37. Ghiloni R, Moretti V, Perotti A. Continuous slice functional calculus in quaternionic Hilbert spaces. Rev Math Phys. 2013;4:1350006.
38. Ng C-K. On quaternionic functional analysis. Math Proc Camb Philos Soc. 2007;143(2):391–406.
39. Rudelson M. Random vectors in the isotropic position. J Funct Anal. 1999;164:60–72.
40. Jia Z, Ng MK, Song G-J. Lanczos method for large-scale quaternion singular value decomposition. Numerical Algorithms. 2018. https://doi.org/10.1007/s11075-018-0621-0
41. Chandrasekaran V, Sanghavi S, Parrilo PA, Willskyc AS. Rank-sparsity incoherence for matrix decomposition. SIAM J Opt. 2011;21(2):572–596.
42. Yuan X, Yang J. Sparse and low-rank matrix decomposition via alternating direction methods. 2009. http://www.optimization-online.org/DB_FILE/2009/11/2447.pdf
43. Jia Z, Wei M, Ling S. A new structure-preserving method for quaternion Hermitian eigenvalue problems. J Comput Appl Math. 2013;239:12–24.
44. Jia Z, Wei M, Zhao M-X, Chen Y. A new real structure-preserving quaternion QR algorithm. J Comput Appl Math. 2018;343:26–48.
45. Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and application to evaluating segmentation algorithms and measuring ecological statistics. Proc IEEE Int Conf Comput Vis. 2001;2(11):416–423.
46. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D Nonlinear Phenom. 1992;60:259–268.
47. Davis D, Yin W. A three-operator splitting scheme and its optimization applications. Set Valued Var Anal. 2017;25(4):829–858. (Three-operator-splitting toolbox: http://www.math.ucla.edu/~wotaoyin/software.html).
48. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang T. Generative image inpainting with contextual attention. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18–22; Salt Lake City, UT. Piscataway, NJ: IEEE; 2018.

49. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. Proceedings of the Neural Information Processing Systems (NIPS); 2016 Dec 5–10; Barcelona, Spain. San Diego, CA: Neural Information Processing Systems Foundation. 2016.
50. Horn RA, Johnson CR. Topics in matrix analysis. Cambridge, UK: Cambridge University Press; 1994. (Corrected reprint of the 1991 original).
51. Lust-Picquard F. Inégalités de Khintchine dans $C_p(1 < p < \infty)$. Comptes Rendus Acad'emie Sci Paris S'erie I. 1986;303(7):289–292.
52. Buchholz A. Operator Khintchine inequality in non-commutative probability. Mathematische Annalen. 2001;319:1–16.
53. Gross D. Recovering low-rank matrices from few coefficients in any basis. IEEE Trans Inf Theory. 2011;57(3):1548–1566.
54. Vershynin R. Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv 1011.3027. 2011.
55. Ledoux M. The concentration of measure phenomenon. Providence, RI: American Mathematical Society; 2001.

How to cite this article: Jia Z, Ng MK, Song G-J. Robust quaternion matrix completion with applications to image inpainting. *Numer Linear Algebra Appl*. 2019;26:e2245. https://doi.org/10.1002/nla.2245

APPENDIX

A.1 | PROOF OF CONVEX ENVELOPE

The von Neumann trace theorem[50] can be generalized to quaternion matrices.

Proposition 1. *Suppose that* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $n \times n$ *quaternion matrices with singular values, that is,*

$$\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n \geq 0, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0,$$

respectively, then

$$\mathrm{Re}(\mathrm{Tr}(\mathbf{AB})) \leq \sum_{j=1}^{n} \rho_j \sigma_j.$$

Proof. It follows from Theorem 1 that there exist two pairs of unitary matrices $\mathbf{U}_i$ and $\mathbf{V}_i$ $(i = 1, 2)$ such that

$$\mathbf{A} = \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^*, \quad \mathbf{B} = \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^*,$$

where $\Sigma_1 = \mathrm{diag}\,(\rho_1, \ldots, \rho_n)$ and $\Sigma_2 = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_n)$. For quaternion matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathrm{Tr}\,(\mathbf{AB}) = \mathrm{Tr}\,(\mathbf{BA})$ is not valid in general. However, $\mathrm{Re}\,(\mathrm{Tr}\,(\mathbf{AB})) = \mathrm{Re}\,(\mathrm{Tr}\,(\mathbf{BA}))$ is always true. Here, $\mathrm{Re}(\cdot)$ is the real part of a quaternion matrix. Based on this property, we have

$$\begin{aligned}
\mathrm{Re}\,(\mathrm{Tr}\,(\mathbf{AB})) &= \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{U}_1 \Sigma_1 \mathbf{V}_1^* \mathbf{U}_2 \Sigma_2 \mathbf{V}_2^*\right)\right) \\
&= \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{V}_2^* \mathbf{U}_1 \Sigma_1 \mathbf{V}_1^* \mathbf{U}_2 \Sigma_2\right)\right) \\
&= \mathrm{Re}\,(\mathrm{Tr}\,(\mathbf{U} \Sigma_1 \mathbf{V} \Sigma_2)) = \mathrm{Re}\left(\sum_{r,s=1}^{n} \mathbf{u}_{r,s} \mathbf{v}_{s,r} \rho_s \sigma_r\right),
\end{aligned}$$

where $\mathbf{U} = (\mathbf{u}_{r,s})_{n\times n} = \mathbf{V}_2^* \mathbf{U}_1$ and $\mathbf{V} = (\mathbf{v}_{r,s})_{n\times n} = \mathbf{V}_1^* \mathbf{U}_2$ are two unitary matrices. Hence,

$$\begin{aligned}
\mathrm{Re}\left(\sum_{r,s=1}^{n} \mathbf{u}_{r,s} \mathbf{v}_{s,r} \rho_s \sigma_r\right) &\leq \left|\mathrm{Re}\left(\sum_{r,s=1}^{n} \mathbf{u}_{r,s} \mathbf{v}_{s,r} \rho_s \sigma_r\right)\right| \\
&\leq \sum_{r,s=1}^{n} \left|\mathbf{u}_{r,s}\right| \left|\mathbf{v}_{s,r}\right| \rho_s \sigma_r \\
&\leq \frac{1}{2} \sum_{r,s=1}^{n} |\mathbf{u}_{r,s}|^2 \rho_s \sigma_r + \frac{1}{2} \sum_{r,s=1}^{n} |\mathbf{v}_{s,r}|^2 \rho_s \sigma_r \leq \sum_{r=1}^{n} \rho_r \sigma_r.
\end{aligned}$$

□

Below is the result that gives the convex envelope of the rank function over the set of quaternion matrices with a bounded norm.

Proposition 2. *Denote* $\mathbf{S} = \{\mathbf{X} \in \mathbb{Q}^{n_1 \times n_2} \mid \|\mathbf{X}\| \leq 1\}$*, then the convex envelope of the function* $\phi(\mathbf{X}) = rank(\mathbf{X})$ *on* $\boldsymbol{S}$ *can be expressed as*

$$\phi_{\text{envo}}(\mathbf{X}) = \|\mathbf{X}\|_* = \sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\mathbf{X}).$$

Proof. We prove this theorem by applying a conjugate function. The conjugate of the quaternion matrix rank function ϕ, with a spectral norm less than or equal to 1, is defined as

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\| \leq 1} (\text{Re}(\langle \mathbf{Y}, \mathbf{X} \rangle) - \text{rank}(\mathbf{X})).$$

Let $n_{(2)} = \min\{n_1, n_2\}$, then it follows from Proposition 1 that

$$\text{Re}(\langle \mathbf{Y}, \mathbf{X} \rangle) \leq \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Y}) \sigma_i(\mathbf{X}), \tag{A1}$$

where $\sigma_i(\mathbf{X})$ denotes the ith largest singular value of $\mathbf{X}$. For any given quaternion matrix $\mathbf{Y}$, there exist unitary matrices $\mathbf{U_Y}$ and $\mathbf{V_Y}$ such that $\mathbf{Y} = \mathbf{U_Y}\Sigma_\mathbf{Y}\mathbf{V_Y^*}$. Choose $\mathbf{X} = \mathbf{U_Y}\Sigma_\mathbf{X}\mathbf{V_Y^*}$, then

$$\begin{aligned}
\text{Re}(\langle \mathbf{Y}, \mathbf{X} \rangle) &= \text{Re}\left(\text{Tr}\left(\mathbf{V_Y}\Sigma_\mathbf{Y}\mathbf{U_Y^*}\mathbf{U_Y}\Sigma_\mathbf{X}\mathbf{V_Y^*}\right)\right) \\
&= \text{Re}\left(\text{Tr}\left(\mathbf{V_Y}\Sigma_\mathbf{Y}\Sigma_\mathbf{X}\mathbf{V_Y^*}\right)\right) \\
&= \text{Tr}\left(\Sigma_\mathbf{Y}\Sigma_\mathbf{X}\right) \\
&= \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Y}) \sigma_i(\mathbf{X}),
\end{aligned}$$

which indicates that the equality in (A1) can be achieved. Then, the conjugate of the quaternion matrix rank function can be rewritten as

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\| \leq 1} \left(\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Y}) \sigma_i(\mathbf{X}) - \texttt{rank}(\mathbf{X}) \right).$$

Now, we first consider separated cases with the rank of $\mathbf{X}$ increasing from 0 to n_2. If $\texttt{rank}(\mathbf{X}) = 0$, then $\phi^*(\mathbf{Y}) = 0$ for all $\mathbf{Y}$. If $\texttt{rank}(\mathbf{X}) = r$ for $1 \leq r \leq n_{(2)}$, then

$$\phi^*(\mathbf{Y}) = \sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r.$$

Hence, $\phi^*(\mathbf{Y})$ can be expressed as

$$\phi^*(\mathbf{Y}) = \max\left\{0, \sigma_1(\mathbf{Y}) - 1, \ldots, \sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r, \ldots, \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Y}) - n_{(2)}\right\}.$$

Furthermore, the sum of all terms that $\sigma_i(\mathbf{Y}) - 1 > 0$ is the largest,

$$\phi^*(\mathbf{Y}) = \begin{cases} 0, & \|\mathbf{Y}\| \leq 1, \\ \sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r, & \sigma_r(\mathbf{Y}) > 1 \text{ and } \sigma_{r+1}(\mathbf{Y}) < 1. \end{cases}$$

The conjugate of $\phi^*(\mathbf{Y})$ is defined as

$$\phi^{**}(\mathbf{Z}) = \sup_{\mathbf{Y}} (\text{Re}(\langle \mathbf{Z}, \mathbf{Y} \rangle) - \phi^*(\mathbf{Y}))$$

for all $\mathbf{Z} \in \mathbb{Q}^{n_1 \times n_2}$. By a similar analysis, suppose that $\mathbf{Z} = \mathbf{U_Z}\Sigma_\mathbf{Z}\mathbf{V_Z^*}$ and choose $\mathbf{Y} = \mathbf{U_Z}\Sigma_\mathbf{Y}\mathbf{V_Z^*}$, then

$$\phi^{**}(\mathbf{Z}) = \sup_{\mathbf{Y}} \left(\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) \sigma_i(\mathbf{Y}) - \phi^*(\mathbf{Y}) \right).$$

We now consider two cases, $\|\mathbf{Z}\| \geq 1$ and $\|\mathbf{Z}\| \leq 1$. For the first case $\|\mathbf{Z}\| \geq 1$, $\sigma_1(\mathbf{Y})$ can be chosen large enough so that $\phi^{**}(\mathbf{Z}) \to \infty$. Then, we have

$$\begin{aligned}\phi^{**}(\mathbf{Z}) &= \sup_{\mathbf{Y}} \left(\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \phi^*(\mathbf{Y}) \right) \\ &= \sup_{\mathbf{Y}} \left(\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \left(\sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r \right) \right) \\ &= \sup_{\mathbf{Y}} \left(\sigma_1(\mathbf{Y})(\sigma_1(\mathbf{Z}) - 1) + \left(\sum_{i=2}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \left(\sum_{i=2}^{r} \sigma_i(\mathbf{Y}) - r \right) \right) \right).\end{aligned}$$

In the second case $\|\mathbf{Z}\| \leq 1$, if $\|\mathbf{Y}\| \leq 1$, then $\phi^*(\mathbf{Y}) = 0$, and the supremum is achieved at $\sigma_i(\mathbf{Y}) = 1, i = 1, \ldots, n_{(2)}$, yielding

$$\phi^{**}(\mathbf{Z}) = \sup \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) = \|\mathbf{Z}\|_*.$$

If $\|\mathbf{Y}\| > 1$, then the argument of the sup is always smaller than the value given above. In fact,

$$\begin{aligned}&\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) \\ =&\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) - \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) + \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) \\ =&\sum_{i=1}^{r} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) + \sum_{i=r+1}^{n_{(2)}} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \sum_{i=1}^{r} (\sigma_i(\mathbf{Y}) - 1) - \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) + \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) \\ =&\sum_{i=1}^{r} (\sigma_i(\mathbf{Z}) - 1)(\sigma_i(\mathbf{Y}) - 1) + \sum_{i=r+1}^{n_{(2)}} \sigma_i(\mathbf{Z})(\sigma_i(\mathbf{Y}) - 1) + \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) \\ \leq&\sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}).\end{aligned}$$

The last inequality can be derived from the following two inequalities:

$$\sum_{i=1}^{r} (\sigma_i(\mathbf{Z}) - 1)(\sigma_i(\mathbf{Y}) - 1) \leq 0 \qquad \text{and} \qquad \sum_{i=r+1}^{n_{(2)}} \sigma_i(\mathbf{Z})(\sigma_i(\mathbf{Y}) - 1) \leq 0.$$

In summary, we have proved that

$$\phi^{**}(\mathbf{Z}) = \sup \sum_{i=1}^{n_{(2)}} \sigma_i(\mathbf{Z}) = \|\mathbf{Z}\|_*$$

over the set $\{\mathbf{Z} \mid \|\mathbf{Z}\| \leq 1\}$. □

A.2 | Proof of Lemma 1

In order to show Lemma 1, we need the following result, which was originally proven by Lust-Picquard[51] and was later sharpened by Buchholz.[52]

Proposition 3. *Let $\{X_i\}_{1 \leq i \leq r}$ be a finite sequence of matrices of the same dimension and $\{\varepsilon_i\}$ be a Rademacher sequence. For each $q \geq 2$,*

$$\left[\mathrm{E}_\epsilon \left\| \sum_i \varepsilon_i X_i \right\|_{S_q}^q \right]^{\frac{1}{q}} \leq C_k \sqrt{q} \max \left\{ \left\| \left(\sum_i X_i^* X_i \right)^{\frac{1}{2}} \right\|_{S_q}, \left\| \left(\sum_i X_i X_i^* \right)^{\frac{1}{2}} \right\|_{S_q} \right\},$$

where $C_k \sqrt{q} = 2^{-\frac{1}{4}} \sqrt{\pi / e}$.

Proof of Lemma 1. Suppose that $1 \le q \le \infty$. The Schatten q-norm of an $n \times n$ quaternion matrix can be defined as

$$\|\mathbf{X}\|_{S_q} = \left(\sum_{i=1}^{n} \sigma_i(\mathbf{X})^q \right)^{\frac{1}{q}},$$

where $\sigma_i(\mathbf{X})$ are the singular values of quaternion matrix $\mathbf{X}$. For $q \ge \log n$, then, by Proposition 3, we have

$$\|\mathbf{X}\|_{S_q} \le \|\mathbf{X}\| \le e\|\mathbf{X}\|_{S_q}.$$

Let $\varepsilon_1, \ldots, \varepsilon_N$ be independent Bernoulli variables taking values $1, -1$ with probability $1/2$, and let $\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}'_1, \ldots, \mathbf{x}'_N$ be independent copies of $\mathbf{x}$. Since $\mathbf{x}_i\mathbf{x}_i^* - \mathbf{x}'_i\mathbf{x}'^*_i$ is Hermitian, we have

$$\begin{aligned}
&\mathrm{E}_{\mathbf{x}} \left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^* - I \right\| \\
&\le \mathrm{E}_{\mathbf{x}}\mathrm{E}_{\mathbf{x}'} \left\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^* - \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}'_i\mathbf{x}'^*_i \right\| \le \mathrm{E}_{\varepsilon}\mathrm{E}_{\mathbf{x}}\mathrm{E}_{\mathbf{x}'} \left\| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \left(\mathbf{x}_i\mathbf{x}_i^* - \mathbf{x}'_i\mathbf{x}'^*_i \right) \right\| \\
&\le 2\mathrm{E}_{\mathbf{x}}\mathrm{E}_{\varepsilon} \left\| \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i\mathbf{x}_i\mathbf{x}_i^* \right\|.
\end{aligned}$$

Applying Proposition 3, we can get

$$\begin{aligned}
\mathrm{E} \left\| \sum_{i=1}^{N} \varepsilon_i\mathbf{x}_i\mathbf{x}_i^* \right\| &\le e \left(\mathrm{E} \left\| \sum_{i=1}^{N} \varepsilon_i\mathbf{x}_i\mathbf{x}_i^* \right\|_{S_q^n}^{p} \right)^{\frac{1}{q}} \le eB_q \left\| \left(\sum_{i=1}^{N} \|\mathbf{x}_i\|^2 \mathbf{x}_i\mathbf{x}_i^* \right)^{\frac{1}{2}} \right\|_{S_q^n} \\
&\le eB_q \left\| \left(\sum_{i=1}^{N} \|\mathbf{x}_i\|^2 \mathbf{x}_i\mathbf{x}_i^* \right)^{\frac{1}{2}} \right\| \le eB_q \max_{i=1,\ldots,N} \|\mathbf{x}_i\| \left\| \sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^* \right\|^{\frac{1}{2}},
\end{aligned}$$

where $0 < B_q \le C\sqrt{q}$. The results follow. □

A.3 | Proof of Lemma 2

Proof. If $\mathbf{A} = 0$, then the result is obvious. We now assume $\mathbf{A} \ne 0$ and $\mathtt{rank}(\mathbf{A}) = r$. Let $\mathbf{Y}$ be given as (7). Then, we have

$$\mathrm{Re}(\langle \mathbf{A}, \mathbf{Y} \rangle) = \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{A}^*\mathbf{Y}\right)\right) = \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{A}^*\left(\mathtt{direc}(\mathbf{A}) + \mathbf{F}\right)\right)\right) = \sum_{1 \le i,j \le r} \left|\mathbf{a}_{ij}\right| = \|\mathbf{A}\|_1$$

and

$$\|\mathbf{Y}\|_\infty = \|\mathtt{direc}(\mathbf{A}) + \mathbf{F}\|_\infty \le \max\left\{\|\mathtt{direc}(\mathbf{A})\|_\infty, \|\mathbf{F}\|_\infty\right\} = 1.$$

It follows that $\mathbf{Y}$ is a subgradient of $\|\cdot\|_1$ at $\mathbf{A}$.

On the other hand, $\mathbf{Y} \in \partial\|\mathbf{A}\|_1$ if and only if $\mathbf{Y} \in \mathbb{V}\left(\mathbf{A} : \|\cdot\|_\infty\right)$, where

$$\mathbb{V}\left(\mathbf{A} : \|\cdot\|_\infty\right) = \left\{\mathbf{G} : \mathbf{G} \in \mathbb{Q}^{n_1 \times n_2}, \mathrm{Re}(\langle \mathbf{A}, \mathbf{G} \rangle) = \|\mathbf{A}\|_1, \|\mathbf{G}\|_\infty \le 1\right\}.$$

Suppose that $\mathbf{A} \in \mathbb{Q}^{n_1 \times n_2}$ has the SVD as (1), then we can get

$$\mathbb{V}\left(\mathbf{A} : \|\cdot\|_\infty\right) = \left\{\mathbf{G} : \mathbf{G} = \mathbf{U}D\mathbf{V}^*, D \in \mathbb{V}\left(\Sigma : \|\cdot\|_\infty\right)\right\}.$$

Since Σ is a real matrix, $D \in \mathbb{V}\left(\Sigma : \|\cdot\|_\infty\right)$ if and only if $D \in \partial\|\Sigma\|_1$, which can be expressed as follows:

$$D = \mathrm{sgn}(\mathbf{U}^*\mathbf{A}\mathbf{V}) + F_1,$$

where $F_1 \in \mathbb{R}^{n_1 \times n_2}$ satisfies $\mathcal{P}_{\Omega(\mathbf{U}^*\mathbf{A}\mathbf{V})}(F_1) = 0$ (the projection of the entries of F_1 on the support of $\mathbf{U}^*\mathbf{A}\mathbf{V}$), $\mathcal{P}_{\Omega(\mathbf{A})}(F_1) = 0$, and $\|F_1\|_\infty \le 1$. The results follow. □

A.4 | Proof of Lemma 3

Proof. If $\mathbf{A} = 0$, then the result is obvious. We now assume $\mathbf{A} \neq 0$. Let $\mathbf{Y}$ be given as (8) satisfying (a) and (b) in Lemma 3. It follows that

$$\begin{aligned}\mathrm{Re}(\langle \mathbf{A}, \mathbf{Y}\rangle) = \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{A}^*\mathbf{Y}\right)\right) &= \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{A}^*\left(\sum_{1\le k\le r}\mathbf{u}_k\mathbf{v}_k^* + \mathbf{F}\right)\right)\right)\\ &= \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{V}\Sigma\mathbf{U}^*\sum_{1\le k\le r}\mathbf{u}_k\mathbf{v}_k^*\right)\right) = \sum_{1\le k\le r}\sigma_i = \|\mathbf{A}\|_*.\end{aligned}$$

Moreover,

$$\|\mathbf{Y}\| = \left\|\sum_{1\le k\le r}\mathbf{u}_k\mathbf{v}_k^* + \mathbf{F}\right\| = \max\left\{\sigma\left(\sum_{1\le k\le r}\mathbf{u}_k\mathbf{v}_k^*\right), \sigma\left(\mathbf{F}\right)\right\} = 1.$$

Then, $\mathbf{Y}$ is a subgradient of the nuclear norm at $\mathbf{A}$.

On the other side, $\mathbf{Y} \in \partial\|\mathbf{A}\|_*$ if and only if $\mathbf{Y} \in \mathbb{V}\left(\mathbf{A} : \|\cdot\|\right)$, where

$$\mathbb{V}\left(\mathbf{A} : \|\cdot\|\right) = \left\{\mathbf{G} : \mathbf{G} \in \mathbb{Q}^{n_1\times n_2}, \mathrm{Re}\left(\langle \mathbf{A}, \mathbf{G}\rangle\right) = \|\mathbf{A}\|_*, \|G\| \le 1\right\}.$$

The SVD of $\mathbf{A} \in \mathbb{Q}^{n_1\times n_2}$ can be expressed as (1); thus,

$$\mathbb{V}\left(\mathbf{A} : \|\cdot\|\right) = \left\{\mathbf{G} : \mathbf{G} = \mathbf{U}D\mathbf{V}^*, D \in \mathbb{V}\left(\Sigma : \|\cdot\|\right)\right\}.$$

Since Σ is a real matrix, $D \in \mathbb{V}\left(\Sigma : \|\cdot\|\right)$ if and only if $D \in \partial\|\Sigma\|_*$, which can be expressed as

$$D = U_1V_1^* + F_1,$$

where the column space of W_1 is unitary to $U_1 = \mathrm{span}\{u_{11}, u_{12}, \ldots, u_{1r}\}$, the row space of F_1 is unitary to $V_1 = \mathrm{spn}\{v_{11}, v_{12}, \ldots, v_{1r}\}$, and the spectral norm of W_1 is less than or equal to 1. Note that u_i and v_i, $i = 1, \ldots, r$, are parts of the standard bases of the vector space, then we can get that $\mathbf{Y} \in \partial\|\mathbf{A}\|_*$ can be expressed as (8) satisfying (a) and (b) in Lemma 3. The result follows. □

A.5 | Proof of Lemmas 5, 6, 7, 8, 9, and 10

First, we show the following proposition, which will be used later.

Proposition 4. *For each pair of quaternion matrices* $\mathbf{W}$ *and* $\mathbf{H}$*,*

$$\mathrm{Re}(\langle \mathbf{W}, \mathbf{H}\rangle) \le \|\mathbf{W}\|\,\|\mathbf{H}\|_*.$$

In addition, for any quaternion matrix $\mathbf{H}$*, there is a matrix* $\mathbf{W}$ *obeying* $\|\mathbf{W}\| = 1$*, which achieves the above equality.*

Proof. Suppose that $\mathbf{H}$ has the SVD as $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^*$, then

$$\mathrm{Re}(\langle \mathbf{W}, \mathbf{H}\rangle) = \mathrm{Re}(\langle \mathbf{W}, \mathbf{U}\Sigma\mathbf{V}^*\rangle) = \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{W}^*\mathbf{U}\Sigma\mathbf{V}^*\right)\right) = \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{V}^*\mathbf{W}^*\mathbf{U}\Sigma\right)\right) \le \|\mathbf{W}\|\,\|\mathbf{H}\|_*.$$

If we take $\mathbf{W} = \mathbf{U}\mathbf{V}^*$, then $\mathrm{Re}(\langle \mathbf{W}, \mathbf{H}\rangle) = \|\mathbf{W}\|\,\|\mathbf{H}\|_*$. □

Proof of Lemma 5. The proof can be divided into two parts. (a) We first prove that there exists a constant c_r' such that

$$\mathbb{E}(\rho^{-1}\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \rho\mathcal{P}_{\mathbb{T}}\|) \le c_r'\sqrt{\frac{\mu_0 n_{(1)} r\log n_{(1)}}{m}}, \tag{A2}$$

provided that the last expression is smaller than 1. (b) Once (a) is satisfied, for each $\lambda > 0$, we have

$$\mathrm{Prob}\left(|Z - \mathbb{E}(Z)| > \lambda\sqrt{\frac{\mu_0 n_{(1)} r\log n_{(1)}}{m}}\right) \le 3\exp\left(-\tau_0'\min\left\{\lambda^2\log n_{(1)}, \lambda\sqrt{\frac{m\log n_{(1)}}{\mu_0 n_{(2)} r}}\right\}\right), \tag{A3}$$

where τ_0' is some positive constant,

$$Z \equiv \rho^{-1}\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \rho\mathcal{P}_{\mathbb{T}}\| = \rho^{-1}\left\|\sum_{i,j}\sum_{\vartheta}(\delta_{ij} - \rho)\mathcal{P}_{\mathbb{T}}\left(e_ie_j^*\vartheta\right)\otimes\mathcal{P}_{\mathbb{T}}\left(e_ie_j^*\vartheta\right)\right\|,$$

and $\vartheta = \{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$.

For (a), suppose that e_i is the ith column of the identity matrix, then

$$\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) = \mathcal{P}_{\mathbf{U}} e_i e_j^* \vartheta + e_i e_j^* \vartheta \mathcal{P}_{\mathbf{V}} - \mathcal{P}_{\mathbf{U}} e_i e_j^* \vartheta \mathcal{P}_{\mathbf{V}}.$$

Recalling Lemma 4 and using $\langle \mathcal{P}_{\mathbf{U}} e_i, \mathcal{P}_{\mathbf{U}} e_i \rangle = \|\mathcal{P}_{\mathbf{U}} e_i\|^2 \le \frac{\mu(\mathbf{U}) r}{n_1}$ and $\langle \mathcal{P}_{\mathbf{V}} e_j, \mathcal{P}_{\mathbf{V}} e_j \rangle = \left\|\mathcal{P}_{\mathbf{V}} e_j\right\|_F^2 \le \frac{\mu(\mathbf{V}) r}{n_2}$, we can derive

$$\begin{aligned}
\left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right)\right\|_F^2 &= \left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right), \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right)\right\rangle \\
&= \left\langle \mathcal{P}_{\mathbf{U}} e_i e_j^* + e_i e_j^* \mathcal{P}_{\mathbf{V}} - \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* + e_i e_j^* \mathcal{P}_{\mathbf{V}} - \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}} \right\rangle \\
&= \langle \mathcal{P}_{\mathbf{U}} e_i e_j^*, \mathcal{P}_{\mathbf{U}} e_i e_j^* \rangle + \langle \mathcal{P}_{\mathbf{U}} e_i e_j^*, e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle - \langle \mathcal{P}_{\mathbf{U}} e_i e_j^*, \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle \\
&\quad + \langle e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* \rangle + \langle e_i e_j^* \mathcal{P}_{\mathbf{V}}, e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle - \langle e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle \\
&\quad - \langle \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* \rangle - \langle \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}}, e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle + \langle \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}} \rangle \\
&= \left\langle \mathcal{P}_{\mathbf{U}} e_i e_j^*, e_i e_j^* \right\rangle + \left\langle e_i e_j^* \mathcal{P}_{\mathbf{V}}, e_i e_j^* \mathcal{P}_{\mathbf{V}} \right\rangle - \left\langle \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}}, \mathcal{P}_{\mathbf{U}} e_i e_j^* \mathcal{P}_{\mathbf{V}} \right\rangle \\
&= \|\mathcal{P}_{\mathbf{U}} e_i\|_2^2 + \left\|\mathcal{P}_{\mathbf{V}} e_j\right\|_2^2 - \|\mathcal{P}_{\mathbf{U}} e_i\|_2^2 \left\|\mathcal{P}_{\mathbf{V}} e_j\right\|_2^2 \\
&\le \frac{2\mu_0 r}{\min\{n_1, n_2\}} = \frac{2\mu_0 r}{n_{(2)}}.
\end{aligned}$$

Similarly, we can get

$$\left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right)\right\|_F^2 \le \frac{2\mu_0 r}{n_{(2)}}, \qquad \left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right)\right\|_F^2 \le \frac{2\mu_0 r}{n_{(2)}}, \quad \text{and} \quad \left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right)\right\|_F^2 \le \frac{2\mu_0 r}{n_{(2)}},$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are three imaginary units, respectively. Next, we will show (A2) is satisfied. Note that any quaternion matrix $\mathbf{X}$ can be expanded as

$$\mathbf{X} = \sum_{i,j} \mathrm{Re}\left(\left\langle e_i e_j^*, \mathbf{X}\right\rangle\right) e_i e_j^* + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{i}, \mathbf{X}\right\rangle\right) e_i e_j^* \mathbf{i} + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{j}, \mathbf{X}\right\rangle\right) e_i e_j^* \mathbf{j} + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{k}, \mathbf{X}\right\rangle\right) e_i e_j^* \mathbf{k},$$

where $\{(e_i e_j^*)_{1 \le i,j \le n}\}$ is the canonical basis. Therefore,

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}}(\mathbf{X}) = \sum_{i,j} &\mathrm{Re}\left(\left\langle e_i e_j^*, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{i}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{i} + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{j}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{j} \\
&+ \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{k}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{k}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X}) = \sum_{i,j} \delta_{ij} &\left(\mathrm{Re}\left(\left\langle e_i e_j^*, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{i}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{i} + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{j}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{j} \right. \\
&\left. + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{k}, \mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{k}\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X}) &= \sum_{i,j} \mathrm{Re}\left(\left\langle e_i e_j^*, \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{i}, \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{i} \\
&\quad + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{j}, \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{j} + \mathrm{Re}\left(\left\langle e_i e_j^* \mathbf{k}, \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}(\mathbf{X})\right\rangle\right) e_i e_j^* \mathbf{k} \\
&= \sum_{i,j} \delta_{ij} \left(\mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right), \mathbf{X}\right\rangle\right) \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right) + \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right), \mathbf{X}\right\rangle\right) \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right) \right. \\
&\quad \left. + \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right), \mathbf{X}\right\rangle\right) \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right) + \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right), \mathbf{X}\right\rangle\right) \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right)\right).
\end{aligned}$$

In other words, we can denote $\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}}$ as

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} = \sum_{i,j} \delta_{ij} &\left(\mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right) + \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right) \right. \\
&\left. + \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right) + \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right)\right).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\mathcal{P}_{\mathbb{T}} = \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\mathbb{T}} = \sum_{i,j} &\mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^*\right) + \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{i}\right) \\
&+ \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{j}\right) + \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \mathbf{k}\right).
\end{aligned}$$

Denote $\vartheta = \{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$, then

$$\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}} - \rho\mathcal{P}_{\mathbb{T}} = \sum_{i,j}\sum_{\vartheta}(\delta_{ij} - \rho)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right).$$

Note that

$$\mathrm{E}\left(\sum_{i,j}\sum_{\vartheta}\left(\delta_{i,j} - \frac{m}{n_1 n_2}\right)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes P_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right) = 0.$$

Applying Lemma 1 to matrices of the form $\mathcal{P}_{\mathbb{T}}(e_i e_j^* \vartheta)$ and using the available bounds on $\|\mathcal{P}_{\mathbb{T}}(e_i e_j^* \vartheta)\|_F$, we have

$$\begin{aligned}
\mathrm{E}(Z) &= \mathrm{E}\left(\rho^{-1}\left\|\sum_{i,j}\sum_{\vartheta}(\delta_{i,j} - \rho)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right\|\right) \\
&= \rho^{-1}\mathrm{E}\left(\left\|\sum_{ij}\sum_{\vartheta}(\delta_{i,j} - \rho)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right\|\right) \\
&\le 4c_1 \frac{n_1 n_2}{m} \frac{\sqrt{m}\sqrt{\log n_1 n_2}}{\sqrt{n_1 n_2}} \sqrt{\frac{2\mu_0 r}{n_{(2)}}} \le c_r' \sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}},
\end{aligned}$$

where c_1 is defined as in Lemma 1 and $c_r' = 8c_1$.

For (b), let $\mathbf{Y} = \sum_{i,j}\mathbf{Y}_{ij}$ with

$$\mathbf{Y}_{i,j} = \rho^{-1}\sum_{\vartheta}\left(\delta_{ij} - \frac{m}{n_1 n_2}\right)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right).$$

Then,

$$\begin{aligned}
Z &= \sup|\langle \mathbf{X}_1, \mathbf{Y}(\mathbf{X}_2)\rangle| \\
&= \sup\rho^{-1}\sum_{i,j}\left|\left\langle \mathbf{X}_1, \sum_{\vartheta}\left(\delta_{ij} - \frac{m}{n_1 n_2}\right)\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right) \otimes \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)(\mathbf{X}_2)\right\rangle\right| \\
&\le \sup\rho^{-1}\sum_{i,j}\sum_{\vartheta}\left(\delta_{ij} - \frac{m}{n_1 n_2}\right)\left|\mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_1\right\rangle\right) \cdot \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_2\right\rangle\right)\right|,
\end{aligned}$$

where the supremum is over a countable collection of matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ obeying $\|\mathbf{X}_1\|_F \le 1$ and $\|\mathbf{X}_2\|_F \le 1$. Moreover, we can get

$$\begin{aligned}
|\langle \mathbf{X}_1, \mathbf{Y}(\mathbf{X}_2)\rangle| &\le \rho^{-1}\left|\sum_{i,j}\left(\delta_{ij} - \frac{m}{n_1 n_2}\right)\right|\sum_{\vartheta}\left|\mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_1\right\rangle\right) \cdot \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_2\right\rangle\right)\right| \\
&\le \rho^{-1}\left|\sum_{i,j}\left(\delta_{ij} - \frac{m}{n_1 n_2}\right)\right|\sum_{\vartheta}\left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right\|_F^2 = \sum_{\vartheta}\rho^{-1}\left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right\|_F^2 \\
&\le 4\frac{n_1 n_2}{m}\frac{2\mu_0 r}{n_{(2)}} \le \frac{8n_{(1)}\mu_0 r}{m}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}|\langle \mathbf{X}_1, \mathbf{Y}(\mathbf{X}_2)\rangle|^2 &= \sum_{i,j}\sum_{\vartheta}\rho^{-1}(1-\rho)\left|\mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_1\right\rangle\right) \cdot \mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_2\right\rangle\right)\right| \\
&\le \sum_{i,j}\sum_{\vartheta}\rho^{-1}(1-\rho)\left\|\mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right)\right\|_F^2\left|\mathrm{Re}\left(\left\langle \mathcal{P}_{\mathbb{T}}\left(e_i e_j^* \vartheta\right), \mathbf{X}_2\right\rangle\right)\right|^2 \\
&\le \sum_{i,j}\sum_{\vartheta}\frac{n_1 n_2}{m}\frac{2\mu_0 r}{n_{(2)}}\left|\mathrm{Re}\left(\left\langle e_i e_j^* \vartheta, \mathcal{P}_{\mathbb{T}}(\mathbf{X}_2)\right\rangle\right)\right|^2 \le \frac{8n_{(1)}\mu_0 r}{m}.
\end{aligned}$$

By Talagrand's concentration inequality, we have

$$\mathrm{Prob}\left(|Z - \mathrm{E}(Z)| > t\right) \le 3\exp\left(-\frac{t}{KB}\log\left(1 + \frac{t}{2}\right)\right) \le 3\exp\left(-\frac{t\log 2}{KB}\min\left\{1, \frac{t}{2}\right\}\right),$$

where $t = \lambda\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}}$, K is a numerical constant, and $B = \frac{8n_{(1)}\mu_0 r}{m}$. Set $\lambda = \sqrt{\frac{\beta}{\theta}}$ and assume that $m > \frac{\beta}{\tau_0'}\mu_0 n_{(1)} r \log n_{(1)}$, then the right-hand side of (A3) is upper bounded by $3n_{(1)}^{-\beta}$. Thus, we can get

$$Z \le c_r'\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}} + \sqrt{\frac{\beta}{\theta}}\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}},$$

which means that (11) is satisfied with $c = c_r' + \sqrt{\frac{\beta}{\theta}}$. □

By Lemma 5, we can prove Lemma 6, which plays an important role in the dual certification.

Proof of Lemma 6. Choosing

$$\varepsilon \ge c\sqrt{\frac{\mu_0 n_{(1)} r \log n_{(1)}}{m}}$$

and recalling $\rho = \frac{m}{n_1 n_2}$, we can get

$$\varepsilon^2 \ge c^2 \frac{\mu_0 n_{(1)} r \log n_{(1)}}{m} \times \frac{n_{(2)}}{n_{(2)}} = c^2 \frac{\mu_0 \rho^{-1} r \log n_{(1)}}{n_{(2)}}.$$

It follows from (11) that if $\rho \ge c^2 \frac{\mu_0 r \log n_{(1)}}{\varepsilon^2 n_{(2)}}$, then (12) is satisfied. □

Proof of Lemma 7. From Lemma 6, we have

$$\left\| \mathcal{P}_{\mathbb{T}} - (1-\rho)^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}} \right\| \le \varepsilon$$

with

$$1 - \rho \ge c^2 \frac{\mu_0 r \log n_{(1)}}{\varepsilon^2 n_{(2)}}.$$

Since $\mathcal{I} = \mathcal{P}_\Omega + \mathcal{P}_{\Omega^\perp}$,

$$\mathcal{P}_{\mathbb{T}} - (1-\rho)^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}} = (1-\rho)^{-1}\left((1-\rho)\mathcal{P}_{\mathbb{T}} - \mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right) = (1-\rho)^{-1}\left(\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}} - \rho\mathcal{P}_{\mathbb{T}}\right).$$

Thus,

$$\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}} = \rho\mathcal{P}_{\mathbb{T}} + (1-\rho)\left(\mathcal{P}_{\mathbb{T}} - (1-\rho)^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right).$$

It follows that

$$\|\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}}\| \le \|\rho\mathcal{P}_{\mathbb{T}}\| + \left\|(1-\rho)\left(\mathcal{P}_{\mathbb{T}} - (1-\rho)^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega^\perp}\mathcal{P}_{\mathbb{T}}\right)\right\| \le \rho + (1-\rho)\varepsilon \le \rho + \varepsilon.$$

Moreover, when $\rho + \varepsilon \le \frac{1}{4}$, we can get $\|\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}}\| \le \frac{1}{2}$. □

Proof of Lemma 8. For two arbitrary quaternion matrices $\mathbf{A}$ and $\mathbf{B}$ with proper sizes,

$$\|\mathbf{AB}\|_F \le \|\mathbf{A}\|\,\|\mathbf{B}\|_F.$$

Then, for any $\mathbf{Z} \in \mathbb{T}$, that is, $\mathbf{Z} = \mathcal{P}_{\mathbb{T}}(\mathbf{Z})$, we get

$$\begin{aligned}\left\|\mathbf{Z} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathbf{Z}\right\|_F &= \left\|\mathcal{P}_{\mathbb{T}}(\mathbf{Z}) - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}}(\mathbf{Z})\right\|_F = \left\|\left(\mathcal{P}_{\mathbb{T}} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}}\right)\mathcal{P}_{\mathbb{T}}(\mathbf{Z})\right\|_F \\ &\le \left\|\left(\mathcal{P}_{\mathbb{T}} - \rho^{-1}P_{\mathbb{T}}\mathcal{P}_\Omega\mathcal{P}_{\mathbb{T}}\right)\right\|\,\|\mathcal{P}_{\mathbb{T}}(\mathbf{Z})\|_F \le \varepsilon\|\mathbf{Z}\|_F.\end{aligned}$$

□

Proof of Lemma 9. Define $\Omega = \{(i,j) : \delta_{ij} = 1\}$, where δ_{ij} is an independent sequence of the Bernoulli variables with parameter ρ. Denote $\mathbf{Z}' = \mathbf{Z} - \rho^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_\Omega\mathbf{Z}$, which also can be given by

$$\mathbf{Z}' = \sum_{i,j}(1 - \rho^{-1}\delta_{ij})\mathcal{P}_{\mathbb{T}}\left(\mathbf{Z}_{ij}e_i e_j^*\right),$$

where $e_i, e_j, i, j = 1, \ldots, n$, are defined as in Lemma 5. Then, $\mathbf{Z}'_{i_0,j_0}$ can be expressed as a sum of independent random variables, that is,

$$\mathbf{Z}'_{i_0 j_0} = \sum_{i,j}\mathbf{Y}_{ij}, \quad \mathbf{Y}_{ij} = \sum_{\vartheta}(1 - \rho^{-1}\delta_{ij})\mathrm{Re}\left(\left\langle e_{i_0}e_{j_0}^*\vartheta, \mathcal{P}_{\mathbb{T}}\left(\mathbf{Z}_{ij}e_i e_j^*\vartheta\right)\right\rangle\right)e_{i_0}e_{j_0}^*\vartheta.$$

It is easy to prove $\mathbb{E}(\mathbf{Y}_{ij}) = 0$, and

$$\sum_{i,j} \mathrm{Var}(\mathbf{Y}_{ij}) = \sum_{i,j} \sum_{\vartheta} \mathbb{E}\Big((1-\rho^{-1}\delta_{ij}) \mathrm{Re}\left(\left\langle e_{i_0} e_{j_0}^* \vartheta, \mathcal{P}_{\mathbb{T}} \left(\mathbf{Z}_{ij} e_i e_j^* \vartheta \right) \right\rangle \right) e_{i_0} e_{j_0}^* \vartheta \Big)^2.$$

Note that

$$\mathbb{E}\left((1-\rho^{-1}\delta_{ij})^2 \right) = \mathbb{E}\left(1 - 2\rho^{-1}\delta_{ij} + \rho^{-2}\delta_{ij}^2 \right) = \frac{\mathbb{E}\left(\rho^2 - 2\rho\delta_{ij} + \delta_{ij}^2 \right)}{\rho^2} = (1-\rho)\,\rho^{-1}.$$

From the proof of Lemma 5, we have

$$|\mathbf{Y}_{ij}| = \left| \sum_{\vartheta} (1-\rho^{-1}\delta_{ij}) \mathrm{Re}\left(\left\langle e_{i_0} e_{j_0}^* \vartheta, \mathcal{P}_{\mathbb{T}} \left(\mathbf{Z}_{ij} e_i e_j^* \vartheta \right) \right\rangle \right) e_{s_0} e_{t_0}^* \vartheta \right| \le 8\rho^{-1} \|\mathbf{Z}\|_\infty \mu_0 r / n_{(2)}$$

and

$$\begin{aligned}
\sum_{i,j} \mathrm{Var}(\mathbf{Y}_{ij}) &\le (1-\rho)\,\rho^{-1} \left(\sum_{i,j} \sum_{\vartheta} \left| \mathrm{Re}\left(\left\langle e_{i_0} e_{j_0}^* \vartheta, \mathcal{P}_{\mathbb{T}} \left(\mathbf{Z}_{ij} e_i e_j^* \vartheta \right) \right\rangle \right) \right|^2 \right) \\
&= (1-\rho)\,\rho^{-1} \left(\sum_{i,j} \sum_{\vartheta} \left| \mathrm{Re}\left(\left\langle \mathbf{Z}_{ij} e_i e_j^* \vartheta, P_{\mathbb{T}} \left(e_{i_0} e_{j_0}^* \vartheta \right) \right\rangle \right) \right|^2 \right) \\
&\le (1-\rho)\,\rho^{-1} \sum_{\vartheta} \left(\|\mathbf{Z}\|_\infty^2 \left\| P_{\mathbb{T}} \left(e_{i_0} e_{j_0}^* \vartheta \right) \right\|_F^2 \right) \le (1-\rho)\,\rho^{-1} \|\mathbf{Z}\|_\infty^2 \frac{8\mu_0 r}{n_{(2)}}.
\end{aligned}$$

Then, by the Bernstein inequality, we have

$$\mathrm{Prob}\left(\left| \mathbf{Z}'_{ij} \right| > \varepsilon \|\mathbf{Z}\|_\infty \right) \le 2\exp\left(-\frac{\varepsilon^2 n_{(1)} \rho}{4(1-\rho+\varepsilon)\,\mu_0 r} \right).$$

It follows that (13) is satisfied when ρ satisfies the assumption. □

Proof of Lemma 10. Note that, for quaternion matrix $\mathbf{Z}$, we have

$$\|\mathbf{Z}\| = \sup_{\|\mathbf{x}\|=1, \|\mathbf{y}\|=1} \sum_{\mathbf{x},\mathbf{y}} \langle \mathbf{Z}, \mathbf{x} \otimes \mathbf{y} \rangle \le \sqrt{n_1 n_2} \|\mathbf{Z}\|_\infty.$$

Then, we can get the results by applying theorem 7 in the work of Recht.[7] □

A.6 | Proof of Theorem 3

Proof of Theorem 3. First, we can get $\Gamma^\perp \cap \mathbb{T} = 0$ by $\|\mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F \le n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F$. Next, we consider a feasible perturbation $(\mathbf{L}_0 + \mathbf{H}_L, \mathbf{S}'_0 - \mathbf{H}_S)$, where $\mathcal{P}_\Omega \mathbf{H}_L = \mathcal{P}_\Omega \mathbf{H}_S$, and show that the objective increases unless $\mathbf{H}_L = \mathbf{H}_S = 0$. In other words, $(\mathbf{L}_0, \mathbf{S}'_0)$ is the unique solution. Let $\mathbf{U}\mathbf{V}^* + \mathbf{W}_0$ be a subgradient of the nuclear norm at $\mathbf{L}_0$ and $\mathrm{direc}(\mathbf{S}'_0) + \mathbf{F}_0$ be a subgradient of the l_1-norm at $\mathbf{S}'_0$. It follows from Lemmas 2 and 3 that

$$\begin{aligned}
\|\mathbf{L}_0 + \mathbf{H}_L\|_* + \lambda \left\| \mathbf{S}'_0 - \mathbf{H}_S \right\|_1 \ge \|\mathbf{L}_0\|_* + \lambda \left\| \mathbf{S}'_0 \right\|_1 + \mathrm{Re}\left(\langle \mathbf{U}\mathbf{V}^* + \mathbf{W}_0, \mathbf{H}_L \rangle \right) \\
- \lambda \mathrm{Re}\left(\left\langle \mathtt{direc}\left(\mathbf{S}'_0 \right) + \mathbf{F}_0, \mathbf{H}_S \right\rangle \right).
\end{aligned}$$

Choose $\mathbf{W}_0 = \mathcal{P}_{\mathbb{T}^\perp}(\mathbf{U}_1 \mathbf{V}_1^*)$, where $\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L = \mathbf{U}_1 \Sigma \mathbf{V}_1^*$, and $\mathbf{F}_0 = -\mathtt{direc}(\mathcal{P}_\Gamma \mathbf{H}_S)$, we can get $\mathrm{Re}\left(\langle \mathbf{W}_0, \mathbf{H}_L \rangle \right) = \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_*$ and $\mathrm{Re}\left(\langle \mathbf{F}_0, \mathbf{H}_S \rangle \right) = -\|\mathcal{P}_\Gamma \mathbf{H}_S\|_1$. Then,

$$\|\mathbf{L}_0 + \mathbf{H}_L\|_* + \lambda \left\| \mathbf{S}'_0 - \mathbf{H}_S \right\|_1 \ge \|\mathbf{L}_0\|_* + \lambda \left\| \mathbf{S}'_0 \right\|_1 + \frac{1}{2} \left(\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_* + \lambda \|\mathcal{P}_\Gamma \mathbf{H}_L\|_1 \right) - \frac{1}{n_{(1)}^2} \|\mathcal{P}_{\mathbb{T}} \mathbf{H}_L\|_F.$$

Note that

$$\begin{aligned}
\|\mathcal{P}_{\mathbb{T}} \mathbf{H}_L\|_F &\le \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_{\Gamma^\perp} \mathbf{H}_L\|_F \le \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{H}_L\|_F \\
&\le \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} (I - \mathcal{P}_\Gamma) \mathbf{H}_L\|_F \\
&\le \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \left(\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_F + \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_\Gamma \mathbf{H}_L\|_F \right) \\
&\le \|\mathcal{P}_{\mathbb{T}} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_F \\
&\le (n_{(1)} + 1) \|\mathcal{P}_\Gamma \mathbf{H}_L\|_F + n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_F.
\end{aligned}$$

Applying $\|\mathcal{P}_\Gamma \mathbf{H}_L\|_F \le \|\mathcal{P}_\Gamma \mathbf{H}_L\|_1$ and $\|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_F \le \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_*$, we can get

$$\|\mathbf{L}_0 + \mathbf{H}_L\|_* + \lambda \|\mathbf{S}_0' - \mathbf{H}_S\|_1 \ge \|\mathbf{L}_0\|_* + \lambda \|\mathbf{S}_0'\|_1 + \left(\frac{1}{2} - \frac{1}{n_{(1)}}\right) \|\mathcal{P}_{\mathbb{T}^\perp} \mathbf{H}_L\|_* + \left(\frac{\lambda}{2} - \frac{n_{(1)}+1}{n_{(1)}^2}\right) \|\mathcal{P}_\Gamma \mathbf{H}_L\|_1.$$

Then, the claim follows from $\Gamma^\perp \cap \mathbb{T} = \{0\}$. □

By using the results in Lemmas 5–10, we show a high probability that the assumption in Theorem 2 is valid.

Proposition 5. *Under the assumptions of Theorem 2, the assumption of Theorem 3 is satisfied. That is, for any quaternion matrix* $\mathbf{M}$, $\|\mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F \le n_{(1)} \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F$ *holds with high probability.*

Proof. Set $\rho_0 = \rho\gamma$, then Γ is characterized by the Bernoulli probability distribution with parameter ρ_0. From Lemma 6, $\|P_\mathbb{T} - \rho_0^{-1} \mathcal{P}_\mathbb{T} \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T}\| \le \frac{1}{2}$ with high probability. Furthermore, from

$$\|\mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F = \|\mathcal{P}_\Gamma \left(I - \mathcal{P}_{\mathbb{T}^\perp}\right) \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F = \|\mathcal{P}_\Gamma \mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F,$$

we have

$$\|\mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F \le \|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F.$$

Moreover,

$$\mathrm{Re}(\langle \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle) = \mathrm{Re}(\langle \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \mathcal{P}_\mathbb{T} \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle)$$

implies that

$$\begin{aligned}
\rho_0^{-1} \|\mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 =& \rho_0^{-1} \mathrm{Re}(\langle \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \mathcal{P}_\Gamma P_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle) \\
=& |\mathrm{Re}(\langle \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \rho_0^{-1} \mathcal{P}_\mathbb{T} \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle)| \\
=& |\mathrm{Re}(\langle P_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle) + \mathrm{Re}(\langle \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}, \left(\rho_0^{-1} P_\mathbb{T} \mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} - \mathcal{P}_\mathbb{T}\right) \mathcal{P}_{\Gamma^\perp} \mathbf{M} \rangle)| \\
\ge& \|\mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 - \frac{1}{2} \|\mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 = \frac{1}{2} \|\mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 .
\end{aligned}$$

Therefore, $\|\mathcal{P}_{\mathbb{T}^\perp} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 \ge \|\mathcal{P}_\Gamma \mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2 \ge \frac{\rho_0}{2} \|\mathcal{P}_\mathbb{T} \mathcal{P}_{\Gamma^\perp} \mathbf{M}\|_F^2$. Note that $\rho_0 = \frac{m\gamma}{n_1 n_2}$, then $\sqrt{\frac{\rho_0}{2}} \ge \frac{\sqrt{\frac{m\gamma}{2}}}{n_{(1)}} \ge \frac{1}{n_{(1)}}$ is satisfied if $m\gamma > 2$. Thus, we can derive the claim. □

A.7 | Construction of $\mathbf{Y}^L$ and $\mathbf{W}^S$

The final step of the proof is to construct $\mathbf{Y}^L$ and $\mathbf{W}^S$.

A.7.1 | Construction of $\mathbf{Y}^L$ via the golfing scheme

We use the similar golf scheme, which was introduced by Gross[53] and modified by Candés et al.,[4] to construct the dual certificate $\mathbf{Y}^L$.

It follows from the sampling scheme that the available and clean data Γ is characterized by the Bernoulli probability distribution with parameter $\rho_0 = \gamma\rho$. Moreover, Γ can also be expressed as $\Gamma = \cup_{1 \le j \le j_0} \Gamma_j$, where $\Gamma_j, j = 1, \ldots, j_0$, are independent and can be characterized by the Bernoulli probability distribution with parameter q. Note that q obeys $\rho_0 = 1 - (1-q)^{j_0}$. Take $j_0 = [3 \log n_{(1)}]$, then $q \ge \frac{\rho_0}{j_0}$. We can inductively define

$$\mathbf{Y}_j = \mathbf{Y}_{j-1} + q^{-1} \mathcal{P}_{\Gamma_j} \mathcal{P}_\mathbb{T} \left(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1}\right),$$

with $\mathbf{Y}_0 = 0$, and set $\mathbf{Y}^L = \mathbf{Y}_{j_0}$.

A.7.2 | Construction of $\mathbf{W}^S$ via the least squares method

Note that unreliable data Ω' is characterized by the Bernoulli probability distribution with parameter $(1-\gamma)\rho$, then it follows from Lemma 7 that $\|\mathcal{P}_{\Omega'} \mathcal{P}_\mathbb{T}\| \le \sqrt{(1-\gamma)\rho + \varepsilon}$ holds with high probability, where ε is a function of γ. Recalling that Ω is characterized by the Bernoulli probability distribution with parameter ρ, then, by Lemma 5, we can derive that the restriction of $\mathcal{P}_\mathbb{T} \mathcal{P}_\Omega \mathcal{P}_\mathbb{T}$ to $\mathbb{T}$ is invertible with high probability. Moreover, $\|(\mathcal{P}_\mathbb{T} \mathcal{P}_\Omega \mathcal{P}_\mathbb{T})^{-1}\| \le 2\rho^{-1}$ follows $\mathcal{P}_\mathbb{T} \mathcal{P}_\Omega \mathcal{P}_\mathbb{T} \ge \frac{\rho}{2} \mathcal{P}_\mathbb{T}$ directly. Since

$$\mathcal{P}_{\Omega'} \left(\mathcal{P}_\mathbb{T} + \mathcal{P}_{\Omega^\perp}\right) \mathcal{P}_{\Omega'} = \mathcal{P}_{\Omega'} \mathcal{P}_\mathbb{T} (\mathcal{P}_\mathbb{T} P_\Omega \mathcal{P}_\mathbb{T})^{-1} \mathcal{P}_\mathbb{T} \mathcal{P}_{\Omega'},$$

$\|\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'}\| \le 2\rho^{-1}(\rho(1-\gamma)+\varepsilon)$ with high probability. We then set

$$\mathbf{W}^S = \lambda(\mathcal{I}-(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp}))(\mathcal{P}_{\Omega'}-\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^{-1}\mathrm{direc}(\mathbf{S}_0').$$

Next, we show $\mathbf{Y}^L$ and $\mathbf{W}^S$ satisfy (a)–(i) in Section 3.5 in detail, respectively.

(a) By the definition of $\mathbf{Y}^L$, we can get $\mathbf{Y}^L \in \Gamma$, then $P_{\Gamma^\perp}\mathbf{Y}^L = 0$ is derived.

(b) Denote $\mathbf{Z}_j = \mathbf{U}\mathbf{V}^* - \mathcal{P}_{\mathbb{T}}(\mathbf{Y}_j)$, then

$$\begin{aligned}
\mathbf{Z}_j &= \mathbf{U}\mathbf{V}^* - \mathcal{P}_{\mathbb{T}}(\mathbf{Y}_j) = \mathbf{U}\mathbf{V}^* - \mathcal{P}_{\mathbb{T}}\left(\mathbf{Y}_{j-1} + q^{-1}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1})\right)\\
&= \mathbf{U}\mathbf{V}^* - \mathcal{P}_{\mathbb{T}}(\mathbf{Y}_{j-1}) - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1})\\
&= P_{\mathbb{T}}(\mathbf{U}\mathbf{V}^*) - \mathcal{P}_{\mathbb{T}}(\mathbf{Y}_{j-1}) - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1})\\
&= \mathcal{P}_{\mathbb{T}}(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1}) - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1})\\
&= (P_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}})(\mathbf{U}\mathbf{V}^* - \mathbf{Y}_{j-1})\\
&= (\mathcal{P}_{\mathbb{T}} - q^{-1}\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}})\mathbf{Z}_{j-1},
\end{aligned}$$

and $\mathbf{Y}^L = \mathbf{Y}_{j_0} = q^{-1}\sum_j \mathcal{P}_{\Gamma_j}\mathbf{Z}_{j-1}$. By Lemmas 8 and 9, if $q \ge c^2\frac{\mu_0 r\log n_{(1)}}{\varepsilon^2 n_{(2)}}$, then

$$\|\mathbf{Z}_j\|_\infty \le e^j\|\mathbf{U}\mathbf{V}^*\|_\infty, \quad \|\mathbf{Z}_j\|_F \le e^j\|\mathbf{U}\mathbf{V}^*\|_F = e^j\sqrt{r}.$$

From Lemma 10 and noting that $\|\mathbf{U}\mathbf{V}^*\|_\infty \le \frac{\sqrt{\mu_0 r}}{n_{(2)}}$ and $j_0 = [3\log n_{(1)}]$, we can deduce

$$\begin{aligned}
\left\|\mathcal{P}_{\mathbb{T}^\perp}\mathbf{Y}_{j_0}\right\| &= \left\|\sum_j q^{-1}\mathcal{P}_{\mathbb{T}^\perp}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{Z}_{j-1})\right\| \le \sum_j \left\|q^{-1}\mathcal{P}_{T^\perp}\mathcal{P}_{\Gamma_j}\mathcal{P}_{\mathbb{T}}(\mathbf{Z}_{j-1})\right\|\\
&\le \sum_j \left\|q^{-1}\mathcal{P}_{\Gamma_j}(\mathbf{Z}_{j-1}) - \mathbf{Z}_{j-1}\right\| = \sum_j \left\|q^{-1}(\mathcal{P}_{\Gamma_j} - \mathcal{I})(\mathbf{Z}_{j-1})\right\|\\
&\le c^2\sqrt{\frac{n_{(1)}\log n_{(1)}}{q}}\sum_j \|\mathbf{Z}_{j-1}\|_\infty \le c^2\sqrt{\frac{n_{(1)}\log n_{(1)}}{q}}\sum_j \varepsilon^j\|\mathbf{U}\mathbf{V}^*\|_\infty\\
&= c^2\sqrt{\frac{n_{(1)}\log n_{(1)}}{q}} \times \frac{\varepsilon(1-\varepsilon^j)}{1-\varepsilon} \times \frac{\sqrt{\mu_0 r}}{n_{(2)}} \le c'\sqrt{\frac{\mu_0 r\log^2 n_{(1)}}{n_{(2)}\rho_0}}.
\end{aligned}$$

When $\rho_0 \ge \frac{16(c')^2\mu_0 r\log^2 n_{(1)}}{n_{(2)}}$, then $\left\|\mathcal{P}_{\mathbb{T}^\perp}Y^L\right\| \le \frac{1}{4}$ can be derived (which is possible provided that ρ_r in Theorem 2 is sufficiently small).

(c) By the definitions of $\mathbf{Y}^L$ and $\mathbf{Z}_{j_0}$, we can get

$$\left\|\mathcal{P}_{\mathbb{T}}\mathbf{Y}^L - \mathbf{U}\mathbf{V}^*\right\|_F = \|\mathbf{Z}_{j_0}\|_F \le e^{j_0}\sqrt{r} \le e^{-3\log n_{(1)}}\sqrt{r} \le n_{(1)}^{-2}.$$

(d) Similarly,

$$\left\|\mathcal{P}_{\Gamma}\mathbf{Y}^L\right\|_\infty \le \|\mathbf{Y}^L\|_\infty \le q^{-1}\|\mathbf{U}\mathbf{V}^*\|_\infty \sum_j e^{-j} \le 3(1-e^{-1})\sqrt{\frac{\mu_0 r\log^2 n_{(1)}}{n_{(2)}^2\rho_0^2}}.$$

Then, we can derive (d) by setting $\lambda = \sqrt{\frac{\gamma}{n_{(2)}\rho_0}}$.

(e),(f),(g) Denote $\mathbf{E} = \mathtt{direc}(\mathbf{S}_0')$, which distributes as

$$\mathbf{E}_{i,j} = \begin{cases} \mathbf{a}_{i,j} = \left(\frac{(\mathbf{S}_0')_{i,j}}{|(\mathbf{S}_0')_{i,j}|}\right), & \text{with probability } \rho(1-\gamma).\\ 0, & \text{with probability} 1-\rho(1-\gamma). \end{cases}$$

Then, we can construct $\mathbf{W}^S$ as

$$\begin{aligned}
\mathbf{W}^S &= \lambda(\mathcal{I}-(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp}))(\mathcal{P}_{\Omega'}-\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^{-1}\mathbf{E}\\
&= (\mathcal{I}-(\mathcal{P}_{\mathbb{T}}+\mathcal{P}_{\Omega^\perp}))\left(\mathbf{W}_0^S + \mathbf{W}_1^S\right),
\end{aligned}$$

where $\mathbf{W}_0^S = \lambda\mathbf{E}$ and $\mathbf{W}_1^S = \lambda\Sigma_{k\geq 1}(\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^k\mathbf{E}$. It is easy to prove that (e), (f), and (g) are satisfied by the definition of $\mathbf{W}^S$.

(h) Because the entries of $\mathbf{E}$ are independent and identically distributed and take zero with probability $1-\rho-\rho\gamma$, by the results in[54], we can get $\|\mathbf{E}\| \leq 4\sqrt{n_{(1)}\rho(1-\gamma)}$ with large probability. In other words, when $\lambda = \frac{1}{\sqrt{n_{(1)}\rho}}$, $\|\mathbf{W}_0^S\| \leq 4\sqrt{1-\gamma} < \frac{1}{8}$ holds if $1-\tau$ is small enough. Denote $\Phi = \sum_{k\geq 1}(\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^k$. Obviously, Φ is Hermitian. In order to consider the norm of the operator, we need the standard covering theory. Known from theorem 4.16 in the work of Ledoux,[55] the size of 1/2-net denoted by $\mathcal{N}$ for Euclidean sphere S^{n-1} is, at most, 6^n. Replacing the Euclidean sphere with the 1/2-net, we obtain

$$\|\Phi(\mathbf{E})\| = \sup_{\mathbf{x},\mathbf{y}\in S^{n-1}} |\langle \mathbf{y}, \Phi(\mathbf{E})\mathbf{x}\rangle| \leq 4 \sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}} |\langle \mathbf{y}, \Phi(\mathbf{E})\mathbf{x}\rangle|.$$

Denote

$$X(\mathbf{x},\mathbf{y}) := |\langle \mathbf{y}, \Phi(\mathbf{E})\mathbf{x}\rangle|,$$

where $(\mathbf{x},\mathbf{y})$ is a pair of fixed unit vectors in $\mathcal{N}\times\mathcal{N}$. Set $\vartheta = \{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$, then we have

$$\begin{aligned} X(\mathbf{x},\mathbf{y}) = |\langle \mathbf{y}, \Phi(\mathbf{E})\mathbf{x}\rangle| = |\mathbf{y}^*\Phi(\mathbf{E})\mathbf{x}| &= \left|\sum_{\vartheta} \mathrm{Re}(\mathrm{Tr}(\mathbf{x}(\mathbf{y}\vartheta)^*\Phi(\mathbf{E})))\vartheta\right| \\ &= \left|\sum_{\vartheta} \mathrm{Re}(\langle \Phi(\mathbf{E}), \mathbf{x}\mathbf{y}^*\vartheta^*\rangle)\vartheta\right| = \left|\sum_{\vartheta} \mathrm{Re}(\langle \Phi(\mathbf{x}\mathbf{y}^*\vartheta^*), \mathbf{E}\rangle)\vartheta\right|. \end{aligned}$$

Denote $\Omega' = \mathrm{supp}(\mathbf{S}_0')$, then by Hoeffding's inequality, we have

$$\mathrm{Prob}(|X(\mathbf{x},\mathbf{y})| > t|\Omega') \leq 2\exp\left(-\frac{2t^2}{\|\Phi(\mathbf{x}\mathbf{y}^*)\|_F^2}\right).$$

It follows from $\|\mathbf{x}\mathbf{y}^*\|_F = 1$ that $\|\Phi(\mathbf{x}\mathbf{y}^*)\|_F \leq \|\Phi\|$. Then,

$$\mathrm{Prob}(\|\Phi(\mathbf{E})\| > t|\Omega') = \mathrm{Prob}\left(4\sup_{\mathbf{x},\mathbf{y}\in\mathcal{N}} |\langle \mathbf{y}, \Phi(\mathbf{E})\mathbf{x}\rangle| > t|\Omega'\right) \leq 2|\mathcal{N}|^2\exp\left(-\frac{t^2}{8\|\Phi\|^2}\right).$$

Moreover, $\|\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'}\| \leq 2\rho^{-1}(\rho(1-\gamma)+\varepsilon)$ implies

$$\|\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\| \leq \sqrt{2(1-\gamma)+2\rho^{-1}\varepsilon} = \sigma,$$

and $\|\Phi\| \leq \Sigma_{k\geq 1}\sigma^{2k} = \frac{\sigma^2}{1-\sigma^2}$. Setting $\alpha = \frac{\sigma^2}{1-\sigma^2}$, then we have

$$\mathrm{Prob}(\lambda\|\Phi(\mathbf{E})\| > t) \leq 2\left|\mathcal{N}\right|^2\exp\left(-\frac{t^2}{2\alpha^2\lambda^2}\right) + \mathrm{Prob}(\|\mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\| \geq \sigma).$$

Since $\|(\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))\mathbf{W}_1^S)\| \leq \lambda\|\Phi(\mathbf{E})\|$, $\|(\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))\mathbf{W}_1^S)\| \leq \lambda < \frac{1}{8}$ can be derived when $2\rho^{-1}(\rho(1-\gamma)+\varepsilon)$, which is equivalent to $1-\gamma$ sufficiently small. Then, when $\lambda = 1/\sqrt{n_{(1)}\rho}$, we can derive $\|\mathbf{W}^S\| \leq \frac{1}{4}$, if setting $1-\gamma$ sufficiently small.

(i) Let

$$\mathbf{X}_{(i,j)} = \sum_{\vartheta}(\mathcal{P}_{\Omega'} - \mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^{-1}\mathcal{P}_{\Omega'}(\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))e_ie_j^T\vartheta.$$

Then, for $(i,j) \in \Gamma$, we have

$$\begin{aligned} \mathbf{W}_{st}^S &= \sum_{\vartheta} \mathrm{Re}\left(\left\langle e_ie_j^T\vartheta, \mathbf{W}^S\right\rangle\right)e_ie_j^T\vartheta \\ &= \lambda\sum_{\vartheta} \mathrm{Re}\left(\left\langle e_ie_j^T\vartheta, (\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))(\mathcal{P}_{\Omega'} - \mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^{-1}E\right\rangle\right)e_ie_j^T\vartheta \\ &= \lambda\mathrm{Re}(\langle \mathbf{X}_{(i,j)}, \mathbf{E}\rangle)e_ie_j^T\vartheta. \end{aligned}$$

Since

$$\mathcal{P}_{\Omega'}(\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))e_ie_j^T\vartheta = \mathcal{P}_{\Omega'}\mathcal{P}_{\mathbb{T}}(\mathcal{P}_{\mathbb{T}}\mathcal{P}_{\Omega}\mathcal{P}_{\mathbb{T}})^{-1}\mathcal{P}_{\mathbb{T}}e_ie_j^T\vartheta$$

and $\|\mathcal{P}_{\mathbb{T}}(e_i e_j^T \vartheta)\|_F^2 \leq \frac{2\mu_0 r}{n_{(2)}}$,

$$\|\mathcal{P}_{\Omega'}(\mathcal{I} - (\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp}))e_i e_j^T \vartheta\|_F \leq \frac{2\sqrt{(1-\gamma)\rho + \varepsilon}}{\rho}\|\mathcal{P}_{\mathbb{T}} e_i e_j^T \vartheta\|_F \leq \frac{2\sqrt{2\mu_0 r((1-\gamma)\rho + \varepsilon)}}{\rho\sqrt{n_{(2)}}}$$

holds with high probability. Moreover, it is easy to prove

$$\|(\mathcal{P}_{\Omega'} - \mathcal{P}_{\Omega'}(\mathcal{P}_{\mathbb{T}} + \mathcal{P}_{\Omega^\perp})\mathcal{P}_{\Omega'})^{-1}\| \leq \frac{1}{1-\eta},$$

where $\eta = 2\rho^{-1}(\rho(1-\gamma) + \varepsilon)$. It follows that

$$\sup_{(i,j)\in\Gamma} \|\mathbf{X}_{(i,j)}\|_F \leq \frac{8\sqrt{2\mu_0 r((1-\gamma)\rho + \varepsilon)}}{\rho(1-\eta)\sqrt{n_{(2)}}}.$$

Applying Hoeffding's inequality again, we can get

$$\text{Prob}\left(\sup_{(i,j)\in\Gamma}\left|\mathbf{W}^S_{ij}\right| > \frac{\lambda}{4}\right) \leq 2n_1 n_2 \exp\left(-\frac{1}{8\sigma^2}\right) + \text{Prob}\left(\sup_{(i,j)\in\Gamma}\|\mathbf{X}_{(i,j)}\|_F > \sigma\right).$$

This shows that $\|\mathcal{P}_\Gamma \mathbf{W}^S\|_\infty \leq \lambda/4$ holds if $1-\gamma$ is sufficiently small.