Supplementary Material for Exact and Stable Recovery of Pairwise Interaction Tensors

Shouyuan Chen **Michael R. Lyu** **Irwin King**
The Chinese University of Hong Kong
{sychen,lyu,king}@cse.cuhk.edu.hk

Zenglin Xu
Purdue University
xu218@purdue.edu

1 Main results

For convenience, we first restate the convex program for recovery of pairwise interaction tensors and our main results in both noiseless and noisy cases.

Exact recovery in the absence of noise. We propose to recover matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and therefore tensor $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ using the following convex program,

$$\begin{aligned} \underset{\mathbf{X}\in S_A,\mathbf{Y}\in S_B,\mathbf{Z}\in S_C}{\text{minimize}} \quad & \sqrt{n_3}\,\|\mathbf{X}\|_* + \sqrt{n_1}\,\|\mathbf{Y}\|_* + \sqrt{n_2}\,\|\mathbf{Z}\|_* \\ \text{subject to} \quad & X_{ij} + Y_{jk} + Z_{ki} = T_{ijk}, \quad (i,j,k) \in \Omega. \end{aligned} \tag{1}$$

We show that, under the incoherence conditions, the above nuclear norm minimization method successful recovers a pairwise interaction tensor $\mathcal{T}$ when the number of observations m is $O(nr\log^2 n)$ with high probability.

Theorem 1. *Let $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ be a pairwise interaction tensor $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $\mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C$ as defined in Proposition 1. Without loss of generality assume that $9 \le n_1 \le n_2 \le n_3$. Suppose we observed m entries of $\mathcal{T}$ with the locations sampled uniformly at random from $[n_1] \times [n_2] \times [n_3]$ and also suppose that each of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ is (μ_0, μ_1)-incoherent. Then, there exists a universal constant C, such that if*

$$m > C \max\{\mu_1^2, \mu_0\} n_3 r \beta \log^2(6n_3),$$

where $r = \max\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B}), \mathrm{rank}(\mathbf{C})\}$ and $\beta > 2$ is a parameter, the minimizing solution $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ for program Eq. (1) is unique and satisfies $\mathbf{X} = \mathbf{A}, \mathbf{Y} = \mathbf{B}, \mathbf{Z} = \mathbf{C}$ with probability at least $1 - \log(6n_3)6n_3^{2-\beta} - 3n_3^{2-\beta}$.

Stable recovery in the presence of noise. Our noisy model assumes that we observe

$$\hat{T}_{ijk} = T_{ijk} + \sigma_{ijk}, \quad \text{for all } (i,j,k) \in \Omega, \tag{2}$$

where σ_{ijk} is a noise term which maybe deterministic or stochastic. We assume σ has bounded energy on Ω and specifically that $\|\mathcal{P}_\Omega(\sigma)\|_F \le \epsilon_1$ for some $\epsilon_1 > 0$, where $\mathcal{P}_\Omega(\cdot)$ denotes the restriction on Ω. We derive the error bound of the following quadratically-constrained convex program which recover $\mathcal{T}$ from observations under this assumption.

$$\begin{aligned} \underset{\mathbf{X}\in S_A,\mathbf{Y}\in S_B,\mathbf{Z}\in S_C}{\text{minimize}} \quad & \sqrt{n_3}\,\|\mathbf{X}\|_* + \sqrt{n_1}\,\|\mathbf{Y}\|_* + \sqrt{n_2}\,\|\mathbf{Z}\|_* \\ \text{subject to} \quad & \left\|\mathcal{P}_\Omega(\mathrm{Pair}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})) - \mathcal{P}_\Omega(\hat{\mathcal{T}})\right\|_F \le \epsilon_2. \end{aligned} \tag{3}$$

Theorem 2. *Let $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $\mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C$. Let Ω be the set of observations as described in Theorem 1. Suppose we observe $\hat{T}_{ijk}$ for $(i,j,k) \in \Omega$ as defined in Eq. (2) and also assume that $\|\mathcal{P}_\Omega(\sigma)\|_F \le \epsilon_1$ holds. Denote the reconstruction error of the optimal* *solution* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *of convex program Eq.* (3) *as* $\mathbf{E} = \mathrm{Pair}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \mathcal{T}$. *Also assume that* $\epsilon_1 \leq \epsilon_2$. *Then, we have*

$$\|\mathbf{E}\|_* \leq 5\sqrt{\frac{2rn_1n_2^2}{8\beta\log(n_1)}}(\epsilon_1 + \epsilon_2),$$

with probability at least $1 - \log(6n_3)6n_3^{2-\beta} - 3n_3^{2-\beta}$.

We will also prove Proposition 1 in the later part of the supplemetary material.

Proposition 1. *For any pairwise interaction tensor* $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$*, there exists unique matrices* $\mathbf{A}' \in S_A, \mathbf{B}' \in S_B, \mathbf{C}' \in S_C$ *such that* $\mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \mathrm{Pair}(\mathbf{A}', \mathbf{B}', \mathbf{C}')$ *where we define* $S_B = \{\mathbf{M} \in \mathbb{R}^{n_2 \times n_3} : \mathbf{1}^T\mathbf{M} = \mathbf{0}^T\}$*,*$S_C = \{\mathbf{M} \in \mathbb{R}^{n_3 \times n_1} : \mathbf{1}^T\mathbf{M} = \mathbf{0}^T\}$ *and* $S_A = \{\mathbf{M} \in \mathbb{R}^{n_1 \times n_2} : \mathbf{1}^T\mathbf{M} = \left(\frac{1}{n_2}\mathbf{1}^T\mathbf{M}\mathbf{1}\right)\mathbf{1}^T\}$.

2 Proof of Theorem 1

Sampling model. Recall that Theorem 1 assumed that Ω is sampled uniformly at random from the collection of all set of size m. This uniform sampling model turns out to be awkward to deal with. Following the strategy of [4, 5], we use the sampling with replacement model on Ω as a proxy for uniform sampling. This differs from the earlier approach by [2] where the authors used a Bernoulli sampling model as a proxy for uniform sampling model. The sampling with replacement model has enabled a significant simplification on the proof and therefore we shall follow this model in the rest of our proof. Specifically, we consider the case where the index of each observation is sampled independently and uniformly from the set $[n_1] \times [n_2] \times [n_3]$. Note that, in expectation, the sampling with replacement model is the same with uniform sampling model. It may appear to be troublesome since the sampling with replacement model can lead to duplicated entires. However, the following lemma allows us to bound the probability of failure when sampling with replacement by the likelihood of error under uniform sampling model.

Lemma 1. *([5, Proposition 3.1]) The probability that the recovery algorithm Eq.* (1) *fails when* Ω *is sampled uniformly from the collection of sets of size* m *is no larger than the probability that the algorithm fails when each index of* Ω *is sampled independently and uniformly.*

Proof. The proof is similar to [3, Section ii.C] and [5, Proposition 3.1]. Let Ω' be a collection of indices sampled independent and uniformly from the set $[n_1] \times [n_2] \times [n_3]$. Also denote Ω_k as a set of entries of size k sampled uniformly at random from all sets of entries of size k. Then, we have

$$\begin{aligned}
\Pr(\text{Failure}(\Omega')) &= \sum_{k=0}^{m} \Pr(\text{Failure}(\Omega') | |\Omega'| = k) \Pr(|\Omega'| = k) \\
&= \sum_{k=0}^{m} \Pr(\text{Failure}(\Omega_k)) \Pr(|\Omega'| = k) \\
&\geq \Pr(\text{Failure}(\Omega_m)) \sum_{k=0}^{m} \Pr(|\Omega'| = k) \\
&= \Pr(\text{Failure}(\Omega_m)).
\end{aligned}$$

□

Therefore, the probability of failure when sampling with replacement is larger than that under uniform sampling model. Hence, we only need to upper bound the failure probability under sampling with replacement model to prove Theorem 1. In the rest of this paper, we will consider solely sampling with replacement model.

Preliminaries. In order to present the proof, we require several additional notations. We shall slightly abuse the notation and denote $\mathbf{e}_k$ be the kth standard basis vector, equal to 1 in kth entry and 0 everywhere else. Denote $\delta_{ij} = \mathbf{e}_i\mathbf{e}_j^T$ be the matrix which equals to 1 in entry (i, j) and 0 in other entries. The dimension of $\mathbf{e}_k$ and δ_{ij} shall be clear from context.

Let $\Omega = \{(a_i, b_i, c_i)\}_{i\in[m]}$, where each (a_i, b_i, c_i) is sampled independently and uniformly at random from $[n_1] \times [n_2] \times [n_3]$. We define the operator $\mathcal{R}_\Omega : \mathbb{R}^{n_1\times n_2} \otimes \mathbb{R}^{n_2\times n_3} \otimes \mathbb{R}^{n_3\times n_1} \to \mathbb{R}^m$ to be

$$\mathcal{R}_\Omega(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \sum_{i=1}^{m} \frac{1}{\sqrt{n_3}} \langle \mathbf{X}, \delta_{a_i b_i} \rangle + \frac{1}{\sqrt{n_1}} \langle \mathbf{Y}, \delta_{b_i c_i} \rangle + \frac{1}{\sqrt{n_2}} \langle \mathbf{Z}, \delta_{c_i a_i} \rangle . \tag{4}$$

Then, the original convex program Eq. (1) can be reformulated as

$$\begin{aligned} \underset{\mathbf{X}\in S_A, \mathbf{Y}\in S_B, \mathbf{Z}\in S_C}{\text{minimize}} \quad & \|\mathbf{X}\|_* + \|\mathbf{Y}\|_* + \|\mathbf{Z}\|_* \\ \text{subject to} \quad & \mathcal{R}_\Omega(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \mathbf{t}, \end{aligned} \tag{5}$$

where $t_i = T_{a_i b_i c_i}$ is the ith observation of $\mathcal{T}$. Note that the scaling coefficients on $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ have been incorporated into $\mathcal{R}_\Omega$.

In order to further simplify the notations, we consider the following block diagonal matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{X} & & \\ & \mathbf{Y} & \\ & & \mathbf{Z} \end{bmatrix},$$

or more compactly $\mathbf{M} = \operatorname{diag}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. It is clear that $\|\mathbf{M}\|_* = \|\mathbf{X}\|_* + \|\mathbf{Y}\|_* + \|\mathbf{Z}\|_*$. Now, denote $\delta_{ab}^{(A)} \triangleq \operatorname{diag}(\delta_{ab}, \mathbf{0}_{n_2\times n_3}, \mathbf{0}_{n_3\times n_1})$ where $\delta_{ab}^{(A)}$ is a $n_1 \times n_2$ matrix and $\mathbf{0}_{n_2\times n_3}$ and $\mathbf{0}_{n_3\times n_1}$ are zero matrices of size $n_2 \times n_3$ and $n_3 \times n_1$, respectively. Similarly, we define define $\delta_{bc}^{(B)} \triangleq \operatorname{diag}(\mathbf{0}_{n_1\times n_2}, \delta_{bc}, \mathbf{0}_{n_3\times n_1})$ and $\delta_{ca}^{(C)} \triangleq \operatorname{diag}(\mathbf{0}_{n_1\times n_2}, \mathbf{0}_{n_2\times n_3}, \delta_{ca})$. Now, we have

$$\mathbf{M} = \sum_{ab} \left\langle \mathbf{X}, \delta_{ab}^{(A)} \right\rangle + \sum_{bc} \left\langle \mathbf{Y}, \delta_{bc}^{(B)} \right\rangle + \sum_{ca} \left\langle \mathbf{Z}, \delta_{ca}^{(C)} \right\rangle .$$

Then, we may equivalently define $\mathcal{R}_\Omega$ by

$$\begin{aligned} \mathcal{R}_\Omega(\mathbf{M}) &= \sum_{i=1}^{m} \frac{1}{\sqrt{n_3}} \left\langle \mathbf{M}, \delta_{a_i b_i}^{(A)} \right\rangle + \frac{1}{\sqrt{n_1}} \left\langle \mathbf{M}, \delta_{b_i c_i}^{(B)} \right\rangle + \frac{1}{\sqrt{n_2}} \left\langle \mathbf{M}, \delta_{c_i a_i}^{(C)} \right\rangle \\ &= \sum_{i=1}^{m} \left\langle \mathbf{M}, \frac{1}{\sqrt{n_3}} \delta_{a_i b_i}^{(A)} + \frac{1}{\sqrt{n_1}} \delta_{b_i c_i}^{(B)} + \frac{1}{\sqrt{n_2}} \delta_{c_i a_i}^{(C)} \right\rangle \\ &\triangleq \sum_{i=1}^{m} \langle \mathbf{M}, \sigma_{a_i b_i c_i} \rangle , \end{aligned} \tag{6}$$

where in the last equation, we have defined

$$\sigma_{a_i b_i c_i} \triangleq \frac{1}{\sqrt{n_3}} \delta_{a_i b_i}^{(A)} + \frac{1}{\sqrt{n_1}} \delta_{b_i c_i}^{(B)} + \frac{1}{\sqrt{n_2}} \delta_{c_i a_i}^{(C)}. \tag{7}$$

Note that we have $\mathcal{R}_\Omega(\mathbf{M}) = \mathcal{R}_\Omega(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Therefore, we can further rewrite the convex program as

$$\begin{aligned} \underset{\mathbf{M}\in S}{\text{minimize}} \quad & \|\mathbf{M}\|_* \\ \text{subject to} \quad & \mathcal{R}_\Omega(\mathbf{M}) = \mathbf{t}, \end{aligned} \tag{8}$$

where we have define the linear subspace S as the product space of S_A, S_B and S_C, namely,

$$S = \{\operatorname{diag}(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C\}.$$

Hence, $\mathbf{M} \in S$ if and only if $\mathbf{M}$ is a block diagonal matrix $\operatorname{diag}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ and $\mathbf{X} \in S_A$, $\mathbf{Y} \in S_B$ and $\mathbf{Z} \in S_C$. For convenience, we also define the orthogonal complement $S^\perp$ by

$$S^\perp = \{\operatorname{diag}(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \mathbf{A} \in S_A^\perp, \mathbf{B} \in S_B^\perp, \mathbf{C} \in S_C^\perp\}.$$

Indeed, the convex program Eq. (8), despite the constraint $\mathbf{M} \in S$, seems to be very similar to the standard nuclear norm heuristic to matrix completion problem. However, we found the major challenge here is that the observation operator $\mathcal{R}_\Omega$ is non-orthogonal. Previously, Gross et al. [4]

showed that the nuclear norm heuristic leads to exact recovery when the observation operator is orthogonal. The orthogonality of observation operator is critical to their argument and therefore their proof cannot be directly applied to our problem. In this work, we extend their technique to deal with the non-orthogonal operator $\mathcal{R}_\Omega$. It turns out that the constraint $\mathbf{M} \in S$ plays a vital role in the argument which is unknown in previous work of matrix completion.

In the rest of paper, we shall consider the following singular value decompositions of $\mathbf{A}, \mathbf{B}, \mathbf{C}$

$$\mathbf{A} = \mathbf{U}_A \Sigma_A \mathbf{V}_A^T, \quad \mathbf{B} = \mathbf{U}_B \Sigma_B \mathbf{V}_B^T, \quad \mathbf{C} = \mathbf{U}_C \Sigma_C \mathbf{V}_C^T.$$

Recall that we have defined $S_B = \{\mathbf{M} \in \mathbb{R}^{n_2 \times n_3} : \mathbf{1}^T\mathbf{M} = \mathbf{0}^T\}$, $S_C = \{\mathbf{M} \in \mathbb{R}^{n_3 \times n_1} : \mathbf{1}^T\mathbf{M} = \mathbf{0}^T\}$ and $S_A = \{\mathbf{M} \in \mathbb{R}^{n_1 \times n_2} : \mathbf{1}^T\mathbf{M} = \left(\frac{1}{n_2}\mathbf{1}^T\mathbf{M}\mathbf{1}\right)\mathbf{1}^T\}$. Now, we introduce the orthogonal decompositions of $S_A = T_A \oplus T_A^\perp$, $S_B = T_B \oplus T_B^\perp$ and $S_C = T_C \oplus T_C^\perp$, where T_A is the linear space $T_A = \{\mathbf{U}_A\mathbf{Y}^T + \mathbf{X}\mathbf{V}_A^T : \forall \mathbf{X}, \mathbf{Y}\} \cap S_A$ and $T_A^\perp$ is the orthogonal complement (and respectively $T_B, T_C, T_B^\perp, T_C^\perp$) are defined similarly. Analogous to the definition of S, we define subspace T as

$$T = \{\text{diag}(\mathbf{A}, \mathbf{B}, \mathbf{C}) : \mathbf{A} \in T_A, \mathbf{B} \in T_B, \mathbf{C} \in T_C\}.$$

We also denote the orthogonal complement of T as $T^\perp$ which is defined in a similar way.

Further, the orthogonal projection operator onto $\mathcal{T}_A$ is given by

$$\mathcal{P}_{T_A}(\mathbf{Z}) = \mathbf{P}_{U_A}\mathcal{P}_{S_A}(\mathbf{Z}) + \mathcal{P}_{S_A}(\mathbf{Z})\mathbf{P}_{V_A} - \mathbf{P}_{U_A}\mathcal{P}_{S_A}(\mathbf{Z})\mathbf{P}_{V_A},$$

where $\mathbf{P}_{U_A}$, $\mathbf{P}_{V_A}$ are the orthogonal projections onto U_A and V_A respectively and $\mathcal{P}_{S_A}$ is the orthogonal projection onto S_A. By simple calculation, we can derive

$$\mathcal{P}_{S_A}(\mathbf{A}) = \mathbf{A} - \frac{1}{n_1}\mathbf{1}^T\mathbf{1}\mathbf{A} + \frac{1}{n_1 n_2}(\mathbf{1}^T\mathbf{A}\mathbf{1})\mathbf{1}^T\mathbf{1}.$$

Similarly, we can derive the orthogonal projection operator $\mathcal{P}_{T_B}$ (respectively $\mathcal{P}_{T_C}$) onto $\mathcal{T}_B$ (respectively $\mathcal{T}_C$) as $\mathcal{P}_{T_B}(\mathbf{Z}) = \mathbf{P}_{U_B}\mathcal{P}_{S_B}(\mathbf{Z}) + \mathcal{P}_{S_B}(\mathbf{Z})\mathbf{P}_{V_B} - \mathbf{P}_{U_B}\mathcal{P}_{S_B}(\mathbf{Z})\mathbf{P}_{V_B}$ and $\mathcal{P}_{T_C}(\mathbf{Z}) = \mathbf{P}_{U_C}\mathcal{P}_{S_C}(\mathbf{Z}) + \mathcal{P}_{S_C}(\mathbf{Z})\mathbf{P}_{V_C} - \mathbf{P}_{U_C}\mathcal{P}_{S_C}(\mathbf{Z})\mathbf{P}_{V_C}$, where $\mathcal{P}_{S_B}(\mathbf{B}) = \mathbf{B} - \frac{1}{n_2}\mathbf{1}^T\mathbf{1}\mathbf{B}$ and $\mathcal{P}_{S_C}(\mathbf{C}) = \mathbf{C} - \frac{1}{n_3}\mathbf{1}^T\mathbf{1}\mathbf{C}$.

In addition, we also consider the orthogonal decomposition $\mathbb{R}^{n_1 \times n_2} = S_A \oplus S_A^\perp$ (respectively S_B,$S_B^\perp$, S_C, $S_C^\perp$). The orthogonal projection operator $\mathcal{P}_{S_A^\perp}$, $\mathcal{P}_{S_B^\perp}$ and $\mathcal{P}_{S_C^\perp}$ are given by

$$\mathcal{P}_{S_A^\perp}(\mathbf{A}) = (\mathcal{I} - \mathcal{P}_{S_A})(\mathbf{A}) = \frac{1}{n_1}\mathbf{1}^T\mathbf{1}\mathbf{A} - \frac{1}{n_1 n_2}(\mathbf{1}^T\mathbf{A}\mathbf{1})\mathbf{1}^T\mathbf{1},$$

and

$$\mathcal{P}_{S_B^\perp}(\mathbf{B}) = \frac{1}{n_2}\mathbf{1}^T\mathbf{1}\mathbf{B}, \quad \mathcal{P}_{S_C^\perp}(\mathbf{C}) = \frac{1}{n_3}\mathbf{1}^T\mathbf{1}\mathbf{C}.$$

Moreover, we can derive the orthogonal projection operators $\mathcal{P}_S$ as $\mathcal{P}_S(\text{diag}(\mathbf{X}, \mathbf{Y}, \mathbf{Z})) = \text{diag}(\mathcal{P}_{S_A}(\mathbf{X}), \mathcal{P}_{S_B}(\mathbf{Y}), \mathcal{P}_{S_C}(\mathbf{Z}))$. The orthogonal projection operators $\mathcal{P}_T, \mathcal{P}_{S^\perp}$ and $\mathcal{P}_{T^\perp}$ can also be derived similarly.

To proceed, we shall need one additional tool, the non-commutative Bernstein inequality.

Theorem 3. *(Non-commutative Bernstein inequality [5, Theorem 3.2]) Let* $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *be independent zero mean random matrices of dimension* $d_1 \times d_2$*. Suppose* $\rho_k^2 = \max\{\left\|\mathbf{E}[\mathbf{X}_k\mathbf{X}_k^T]\right\|, \left\|\mathbf{E}[\mathbf{X}_k^T\mathbf{X}_k]\right\|\}$ *and* $\|\mathbf{X}_k\| \leq M$ *almost surely for every* k*. Then, for any* $\tau > 0$*,*

$$\Pr\left[\left\|\sum_{k=1}^m \mathbf{X}_k\right\| > \tau\right] \leq (d_1 + d_2)\exp\left(\frac{-\tau^2/2}{\sum_{k=1}^m \rho_k^2 + M\tau/3}\right).$$

We omit the proof of the non-commutative Bernstein inequality. For details, readers may refer to [5, Appendix A] and [1]. Furthermore, the righthand side is always less than $(d_1 + d_2)\exp(-\frac{3}{8}\tau^2/(\sum_{k=1}^m \rho_k^2))$ when $\tau \leq \frac{1}{M}\sum_{k=1}^m \rho_k^2$. In our proof, we will solely rely on the condensed version of non-commutative Bernstein inequality.

We are now ready to state the proof of Theorem 1. First, in the following theorem, we show that if there exists a "dual certificate", the solution to convex program Eq. (1) is unique and exactly recovers the pairwise interaction tensor.

Theorem 4. *Let $r = \max\{r_1, r_2, r_3\}$. Let $\mathbf{W} = \mathrm{diag}(\mathbf{U}_A\mathbf{V}_A^T, \mathbf{U}_B\mathbf{V}_B^T, \mathbf{U}_C\mathbf{V}_C^T)$ be a block diagonal matrix. Suppose that there exists a "dual certificate" $\mathbf{F} \in \mathrm{range}(\mathcal{R}_\Omega^*)$ such that*

$$\left\|\mathcal{P}_T(\mathbf{F}) - \mathbf{W}\right\|_F \leq \sqrt{\frac{r}{2n_3}}, \quad \left\|\mathcal{P}_{T^\perp}(\mathbf{F})\right\| < \frac{1}{2}$$

And also suppose that

$$\frac{1}{2}\left\|\mathcal{P}_{T^\perp}(\mathbf{E})\right\|_* > \sqrt{\frac{r}{2n_3}}\left\|\mathcal{P}_T(\mathbf{E})\right\|_F$$

holds for any $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$. Then, the $\mathbf{A}, \mathbf{B}, \mathbf{C}$ is the unique minimizing solution of Eq. (1).

Proof. Let $\mathbf{M} = \mathrm{diag}\left(\sqrt{n_3}\mathbf{A}, \sqrt{n_1}\mathbf{B}, \sqrt{n_2}\mathbf{C}\right)$ be a block diagonal matrix. By the definition of nuclear norm, we have $\|\mathbf{M}\|_* = \sqrt{n_3}\|\mathbf{A}\|_* + \sqrt{n_1}\|\mathbf{B}\|_* + \sqrt{n_2}\|\mathbf{C}\|_*$. Now, consider for any block diagonal matrix $\mathbf{E} = \mathrm{diag}(\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C)$ such that $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$. Pick $\mathbf{U}_{A^\perp}$ and $\mathbf{V}_{A^\perp}$ such that $[\mathbf{U}_A, \mathbf{U}_{A^\perp}]$ and $[\mathbf{V}_A, \mathbf{V}_{A^\perp}]$ are unitary matrices and $\left\langle \mathbf{U}_{A^\perp}\mathbf{V}_{A^\perp}^T, \mathcal{P}_{T_A^\perp}(\mathbf{E}_A)\right\rangle = \left\|\mathcal{P}_{T_A^\perp}(\mathbf{E})\right\|_*$. Also pick $\mathbf{U}_{B^\perp}, \mathbf{V}_{B^\perp}, \mathbf{U}_{C^\perp}, \mathbf{V}_{C^\perp}$ similarly. Let $\mathbf{W}_\perp = \mathrm{diag}(\mathbf{U}_{A^\perp}\mathbf{V}_{A^\perp}^T, \mathbf{U}_{B^\perp}\mathbf{V}_{B^\perp}^T, \mathbf{U}_{C^\perp}\mathbf{V}_{C^\perp}^T)$. We have $\mathbf{W}_\perp \in T^\perp$ and $\langle \mathbf{W}_\perp, \mathcal{P}_{T^\perp}(\mathbf{E})\rangle = \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_*$. Also note the fact that $\langle \mathbf{F}, \mathbf{E}\rangle = 0$ since $\mathbf{F} \in \mathrm{range}(\mathcal{R}_\Omega^T)$ and $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$. Then it follows that,

$$\begin{aligned}
\|\mathbf{M} + \mathbf{E}\|_* &\geq \langle \mathbf{W} + \mathbf{W}_\perp, \mathbf{M} + \mathbf{E}\rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} + \mathbf{W}_\perp, \mathbf{E}\rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} + \mathbf{W}_\perp - \mathbf{F}, \mathbf{E}\rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} + \mathbf{W}_\perp - \mathbf{F}, \mathcal{P}_T(\mathbf{E}) + \mathcal{P}_{T^\perp}(\mathbf{E}) + \mathcal{P}_{S^\perp}(\mathbf{E})\rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} - \mathcal{P}_T(\mathbf{F}), \mathcal{P}_T(\mathbf{E})\rangle + \langle \mathbf{W}_\perp - \mathcal{P}_{T^\perp}(\mathbf{F}), \mathcal{P}_{T^\perp}(\mathbf{E})\rangle \\
&= \|\mathbf{M}\|_* - \langle \mathcal{P}_T(\mathbf{F}) - \mathbf{W}, \mathcal{P}_T(\mathbf{E})\rangle + \langle \mathbf{W}_\perp, \mathcal{P}_{T^\perp}(\mathbf{E})\rangle - \langle \mathcal{P}_{T^\perp}(\mathbf{F}), \mathcal{P}_{T^\perp}(\mathbf{E})\rangle \\
&\geq \|\mathbf{M}\|_* - \|\mathcal{P}_T(\mathbf{F}) - \mathbf{W}\|_F \|\mathcal{P}_T(\mathbf{E})\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* - \|\mathcal{P}_{T^\perp}(\mathbf{F})\| \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \\
&> \|\mathbf{M}\|_* - \sqrt{\frac{r}{2n_3}}\|\mathcal{P}_T(\mathbf{E})\|_F + \frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \\
&\geq \|\mathbf{M}\|_* = \sqrt{n_3}\|\mathbf{A}\|_* + \sqrt{n_1}\|\mathbf{B}\|_* + \sqrt{n_2}\|\mathbf{C}\|_*,
\end{aligned}$$

where the first inequality follows from the variational characterization of nuclear norm $\|\mathbf{M} + \mathbf{E}\|_* = \sup_{\|\mathbf{Q}\|=1}\langle \mathbf{Q}, \mathbf{M} + \mathbf{E}\rangle$. We have also used the fact that $\mathbf{E} \in S$ and therefore $\mathcal{P}_{S^\perp}(\mathbf{E}) = \mathbf{0}$. Therefore, if there exists any $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ obeying $\mathcal{R}_\Omega(\mathrm{diag}(\sqrt{n_3}(\mathbf{X} - \mathbf{A}), \sqrt{n_1}(\mathbf{Y} - \mathbf{B}), \sqrt{n_2}(\mathbf{Z} - \mathbf{C})) = \mathbf{0}$, or equivalently $\frac{1}{\sqrt{n_3}}\sqrt{n_3}(X_{ij} - A_{ij}) + \frac{1}{\sqrt{n_1}}\sqrt{n_1}(Y_{jk} - B_{jk}) + \frac{1}{\sqrt{n_2}}\sqrt{n_2}(Z_{ki} - C_{ki}) = 0$ for all $(i,j,k) \in \Omega$, we would have $\|\mathbf{X}\|_* + \|\mathbf{Y}\|_* + \|\mathbf{Z}\|_* > \|\mathbf{A}\|_* + \|\mathbf{B}\|_* + \|\mathbf{C}\|_*$. In other words, if $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ statisifes $X_{ij} + Y_{jk} + Z_{ki} = A_{ij} + B_{jk} + C_{ki}$ for any $(i,j,k) \in \Omega$, the weighted sum of the nuclear norm of $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ would be strictly larger than that of $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$. Therefore, $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ is the unique minimizer of program Eq. (1). ☐

Therefore, we remain to show that such a dual certificate $\mathbf{F}$ exists with high probability. The proof relies on a series of applications of noncommutative Bernstein inequality and the clever golfing scheme proposed by [4]. We begin with an elementary bound on $\|\mathcal{P}_T(\sigma_{abc})\|_F$.

Proposition 2. *Suppose that $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are (μ_0, μ_1)-incoherent. Then, for any $(a,b,c) \in [n_1] \times [n_2] \times [n_3]$, the length of orthogonal projection of σ_{abc} onto space T is bounded by*

$$\|\mathcal{P}_T(\sigma_{abc})\|_F^2 \leq \frac{28\mu_0 r}{n_1 n_2},$$

where $\sigma_{abc} \triangleq \frac{1}{\sqrt{n_3}}\delta_{ab}^{(A)} + \frac{1}{\sqrt{n_1}}\delta_{bc}^{(B)} + \frac{1}{\sqrt{n_2}}\delta_{ca}^{(C)}$.

Proof. By definition, we have $\|\mathcal{P}_T(\sigma_{abc})\|_F^2 = \frac{1}{n_3}\|\mathcal{P}_{T_A}(\delta_{ab})\|_F^2 + \frac{1}{n_1}\|\mathcal{P}_{T_B}(\delta_{bc})\|_F^2 + \frac{1}{n_2}\|\mathcal{P}_{T_C}(\delta_{ca})\|_F^2$. Therefore, it suffices to bound these terms individually.

We first deal with $\|\mathcal{P}_{T_B}(\delta_{bc})\|_F$. It can be decomposed as

$$\begin{aligned}\|\mathcal{P}_{T_B}(\delta_{bc})\|_F^2 &= \langle \mathcal{P}_{T_B}(\delta_{bc}), \mathcal{P}_{T_B}(\delta_{bc})\rangle \\ &= \langle \mathcal{P}_{T_B}(\delta_{bc}), \delta_{bc}\rangle \\ &= \|\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc})\|_F^2 + \|\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2 - \|\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2 \\ &\le \|\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc})\|_F^2 + \|\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2 .\end{aligned}$$

Now it suffices to bound both terms $\|\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc})\|_F^2$ and $\|\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2$. We have

$$\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc}) = \mathbf{P}_{U_B}\delta_{bc} - \frac{1}{n_3}\mathbf{P}_{U_B}\mathbf{1}\mathbf{1}^T\delta_{bc} = \mathbf{P}_{U_B}\delta_{bc},$$

where the second equality holds since $\mathbf{B} \in S_B$ and therefore $\mathbf{P}_{U_B}\mathbf{1} = \mathbf{0}$. Combining with the incoherence property **A0**, we have

$$\|\mathbf{P}_{U_B}\mathcal{P}_{S_B}(\delta_{bc})\|_F^2 = \|\mathbf{P}_{U_B}\delta_{bc}\|_F^2 = \|\mathbf{P}_{U_B}\mathbf{e}_b\|_F^2 \le \frac{\mu_0 r}{n_2}.$$

Next, we need to bound $\|\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2$. We have

$$\begin{aligned}\|\mathcal{P}_{S_B}(\delta_{bc})\mathbf{P}_{V_B}\|_F^2 &= \left\|\delta_{bc}\mathbf{P}_{V_B} - \frac{1}{n_2}\mathbf{1}\mathbf{1}^T\delta_{bc}\mathbf{P}_{V_B}\right\|_F^2 \\ &\le 2\,\|\delta_{bc}\mathbf{P}_{V_B}\|_F^2 + \frac{2}{n_2^2}\left\|\mathbf{1}\mathbf{1}^T\delta_{bc}\mathbf{P}_{V_B}\right\|_F^2 \\ &= 2\,\|\mathbf{P}_{V_B}\mathbf{e}_c\|_F^2 + \frac{2}{n_2^2}\left\|\mathbf{1}\mathbf{e}_c^T\mathbf{P}_{V_B}\right\|_F^2 \\ &= 2\,\|\mathbf{P}_{V_B}\mathbf{e}_c\|_F^2 + \frac{2}{n_2^2}\left\|\sum_{b'}\mathbf{e}_{b'}\mathbf{e}_c^T\mathbf{P}_{V_B}\right\|_F^2 \\ &= 2\,\|\mathbf{P}_{V_B}\mathbf{e}_c\|_F^2 + \frac{2}{n_2^2}\,\|n_2\mathbf{P}_{V_B}\mathbf{e}_c\|_F^2 \\ &= 4\,\|\mathbf{P}_{V_B}\mathbf{e}_c\|_F^2 \\ &\le \frac{4\mu_0 r}{n_3},\end{aligned}$$

where the first inequality is the Cauchy-Schwartz inequality and the final inequality is due to incoherence property **A0**. Therefore,

$$\|\mathcal{P}_{T_B}(\delta_{bc})\|_F^2 \le \frac{\mu_0 r}{n_2} + \frac{4\mu_0 r}{n_3} \le \frac{5\mu_0 r}{n_2}.$$

In addition, we can bound $\|\mathcal{P}_{T_C}(\delta_{ca})\|_F$ using the same method.

It remains to bound $\|\mathcal{P}_{T_A}(\delta_{ab})\|_F^2$ which is no greater than $\|\mathbf{P}_{U_A}\mathcal{P}_{S_A}(\delta_{ab})\|_F^2 + \|\mathcal{P}_{S_A}(\delta_{ab})\mathbf{P}_{V_A}\|_F^2$ following a similar analysis. Again, we start with bounding $\|\mathbf{P}_{U_A}\mathcal{P}_{S_A}(\delta_{ab})\|_F^2$. We have

$$\begin{aligned}\|\mathbf{P}_{U_A}\mathcal{P}_{S_A}(\delta_{ab})\|_F^2 &= \left\|\mathbf{P}_{U_A}\delta_{ab} - \frac{1}{n_1}\mathbf{P}_{U_A}\mathbf{1}\mathbf{e}_b^T + \frac{1}{n_1 n_2}\mathbf{P}_{U_A}\mathbf{1}\mathbf{1}^T\right\|_F^2 \\ &\le 3\,\|\mathbf{P}_{U_A}\delta_{ab}\|_F^2 + \frac{3}{n_1^2}\left\|\mathbf{P}_{U_A}\mathbf{1}\mathbf{e}_b^T\right\|_F^2 + \frac{3}{n_1^2 n_2^2}\left\|\mathbf{P}_{U_A}\mathbf{1}\mathbf{1}^T\right\|_F^2 \\ &= 3\,\|\mathbf{P}_{U_A}\mathbf{e}_a\|_F^2 + \frac{3}{n_1^2}\left\|\sum_{a'}\mathbf{P}_{U_A}\mathbf{e}_{a'}\right\|_F^2 + \frac{3}{n_1^2 n_2^2}\left\|\sum_{a'b'}\mathbf{P}_{U_A}\mathbf{e}_{a'}\right\|_F^2 \\ &\le 3\,\|\mathbf{P}_{U_A}\mathbf{e}_a\|_F^2 + \frac{3}{n_1}\sum_{a'}\|\mathbf{P}_{U_A}\mathbf{e}_{a'}\|_F^2 + \frac{3}{n_1 n_2}\sum_{a'b'}\|\mathbf{P}_{U_A}\mathbf{e}_{a'}\|_F^2 \\ &\le \frac{9\mu_0 r}{n_1},\end{aligned}$$

where we have repeatedly applied Cauchy-Schwartz inequality and assumption A1. We can also bound $\|\mathbf{P}_{S_A}(\delta_{ab})\mathbf{P}_{V_A}\|_F^2$ using the same method as

$$\|\mathbf{P}_{S_A}(\delta_{ab})\mathbf{P}_{V_A}\|_F^2 \leq \frac{9\mu_0 r}{n_2}.$$

Therefore, we have

$$\|\mathcal{P}_{T_A}(\delta_{ab})\|_F^2 \leq \frac{9\mu_0 r}{n_1} + \frac{9\mu_0 r}{n_2} \leq \frac{18\mu_0 r}{n_1}.$$

Finally, combining the above inequalities, we have

$$\begin{aligned}
\|\mathcal{P}_T(\sigma_{abc})\|_F^2 = \frac{1}{n_3}\|\mathcal{P}_{T_A}(\delta_{ab})\|_F^2 + \frac{1}{n_1}\|\mathcal{P}_{T_B}(\delta_{bc})\|_F^2 + \frac{1}{n_2}\|\mathcal{P}_{T_C}(\delta_{ca})\|_F^2 \leq \frac{26\mu_0 r}{n_1 n_2} \\
&\leq \frac{18\mu_0 r}{n_1 n_3} + \frac{5\mu_0 r}{n_1 n_2} + \frac{5\mu_0 r}{n_1 n_2} \\
&\leq \frac{28\mu_0 r}{n_1 n_2}.
\end{aligned}$$

□

The next proposition shows that, in expectation, $\frac{n_1 n_2 n_3}{m}\mathcal{R}_\Omega^*\mathcal{R}_\Omega$ is an isometric operator on S. Therefore, the observation operator $\mathcal{R}_\Omega$ can be regarded as an orthogonal projection operator in expectation on subspace S.

Proposition 3. *Suppose Ω is a set of entries of size m which is sampled independent and uniformly with replacement. Then for any block diagonal matrix $\mathbf{E} = \operatorname{diag}(\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C)$ satisfying that $\mathbf{E} \in S$, denoting $\mathcal{O}(\mathbf{E}) \triangleq \frac{n_1 n_2 n_3}{m}\mathbb{E}[\mathcal{R}_\Omega^*\mathcal{R}_\Omega(\mathbf{E})]$, we have*

$$\mathcal{P}_S(\mathcal{O}(\mathbf{E})) = \mathbf{E}.$$

Proof. We can calculate $\mathcal{O}(\mathbf{E})$ as follows,

$$\begin{aligned}
\mathcal{O}(\mathbf{E}) &= \sum_{abc}\langle \mathbf{E}, \sigma_{abc}\rangle \sigma_{abc} \\
&= \sum_{abc}\left[\left\langle \mathbf{E}, \left(\frac{1}{\sqrt{n_3}}\delta_{ab}^{(A)} + \frac{1}{\sqrt{n_1}}\delta_{bc}^{(B)} + \frac{1}{\sqrt{n_2}}\delta_{ca}^{(C)}\right)\right\rangle\right]\left(\frac{1}{\sqrt{n_3}}\delta_{ab}^{(A)} + \frac{1}{\sqrt{n_1}}\delta_{bc}^{(B)} + \frac{1}{\sqrt{n_2}}\delta_{ca}^{(C)}\right) \\
&= \Bigg(\mathbf{E}_A + \frac{1}{\sqrt{n_1 n_3}}\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T\mathbf{E}_B^T + \frac{1}{\sqrt{n_2 n_3}}\mathbf{E}_C^T\mathbf{1}_{n_3}\mathbf{1}_{n_2}^T, \\
&\qquad \mathbf{E}_B + \frac{1}{\sqrt{n_1 n_2}}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T\mathbf{E}_C^T + \frac{1}{\sqrt{n_1 n_3}}\mathbf{E}_A^T\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T, \\
&\qquad \mathbf{E}_C + \frac{1}{\sqrt{n_2 n_3}}\mathbf{1}_{n_3}\mathbf{1}_{n_2}^T\mathbf{E}_A^T + \frac{1}{\sqrt{n_1 n_2}}\mathbf{E}_B^T\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T\Bigg) \\
&= \left(\mathbf{E}_A + \frac{1}{\sqrt{n_1 n_3}}\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T\mathbf{E}_B^T, \mathbf{E}_B + \frac{1}{\sqrt{n_1 n_2}}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T\mathbf{E}_C^T + \frac{1}{\sqrt{n_1 n_3}}\mathbf{E}_A^T\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T, \mathbf{E}_C + \frac{1}{\sqrt{n_2 n_3}}\mathbf{1}_{n_3}\mathbf{1}_{n_2}^T\mathbf{E}_A^T\right),
\end{aligned}$$

where the third equality follows since $\mathbf{1}_{n_2}^T\mathbf{E}_B = \mathbf{0}_{n_3}^T$ and $\mathbf{1}_{n_3}^T\mathbf{E}_C = \mathbf{0}_{n_1}^T$.

Now, by the definition of S_A and $S_A^\perp$, since $\mathbf{1}_{n_3}^T\mathbf{E}_B^T\mathbf{1}_{n_2} = 0$, we have $\mathbf{1}_{n_2}\mathbf{1}_{n_3}^T\mathbf{E}_B^T \in S_A^\perp$ and therefore $\mathcal{P}_{S_A}(\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T\mathbf{E}_B^T) = \mathbf{0}_{n_1\times n_2}$. We also have $\mathcal{P}_{S_B}(\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T\mathbf{E}_C^T) = \mathbf{0}_{n_2\times n_3}$ and $\mathcal{P}_{S_C}(\mathbf{1}_{n_3}\mathbf{1}_{n_2}^T\mathbf{E}_A^T) = \mathbf{0}_{n_3\times n_1}$. In addition, we have $\mathbf{E}_A^T\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T \in S_B^\perp$ and hence $\mathcal{P}_{S_B}(\mathbf{E}_A^T\mathbf{1}_{n_1}\mathbf{1}_{n_3}^T) = \mathbf{0}_{n_3\times n_1}$. Combining these facts, we have

$$\mathcal{P}_S(\mathcal{O}(\mathbf{E})) = \mathbf{E}.$$

□

Next, we show that, with high probability, $\mathcal{R}_\Omega^*\mathcal{R}_\Omega$ is very close to an isometry on subspace T if the number of observations $|\Omega|$ is sufficient by appealing to the non-commutative Bernstein inequality.

Lemma 2. *Suppose Ω is a set of entries of size m which is sampled independently and uniformly from $[n_1] \times [n_2] \times [n_3]$ with replacement. Then for all $\beta > 1$,*

$$\frac{n_1 n_2 n_3}{m} \left\| \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T - \frac{m}{n_1 n_2 n_3} \mathcal{P}_T \right\| \leq \sqrt{\frac{16 p \mu_0 r n_3 \beta \log(n_3)}{3m}}$$

with probability at least $1 - 2n_3^{2-2\beta}$ if $m > \frac{448}{3} \mu_0 r n_3 \beta \log(n_3)$.

Proof. By Proposition 3, for any $\mathbf{E} \in T$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T\right] &= \mathcal{P}_T \mathbb{E}\left[\mathcal{R}_\Omega^* \mathcal{R}_\Omega\right] \mathcal{P}_T \\
&= \mathcal{P}_T \left(\frac{m}{n_1 n_2 n_3} \mathcal{O} \right) \mathcal{P}_T \\
&= \frac{m}{n_1 n_2 n_3} \mathcal{P}_T.
\end{aligned}$$

Now we use noncommunicative Bernstein inequality to bound the deviation of the operator $\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T$ from its expected value $\frac{m}{n_1 n_2 n_3} \mathcal{P}_T$ in spectral norm.

Consider any block diagonal matrix $\mathbf{E} = \operatorname{diag}(\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C)$, we can decompose $\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E})$ as follows,

$$\begin{aligned}
\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E}) &= \sum_{k=1}^{m} \left\langle \mathcal{P}_T(\mathbf{E}), \sigma_{a_k b_k c_k} \right\rangle \mathcal{P}_T\left(\sigma_{a_k b_k c_k}\right) \\
&= \sum_{k=1}^{m} \left\langle \mathbf{E}, \mathcal{P}_T(\sigma_{a_k b_k c_k}) \right\rangle \mathcal{P}_T(\sigma_{a_k b_k c_k})
\end{aligned}$$

Define the operator τ_{abc} which maps $\mathbf{E}$ to $\left\langle \mathbf{E}, \mathcal{P}_T(\sigma_{abc}) \right\rangle \mathcal{P}_T(\sigma_{abc})$. Clearly, we have $\mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T = \sum_{k=1}^{m} \tau_{a_k b_k c_k}$ and $\mathbb{E}[\tau_{a_k b_k c_k}] = \frac{1}{n_1 n_2 n_3} \mathcal{P}_T$. We can bound the operator norm $\|\tau_{abc}\|$ using Proposition 3 as follows

$$\begin{aligned}
\|\tau_{abc}\| &= \sup_{\|\mathbf{E}\|_F = 1} \|\tau_{abc}(\mathbf{E})\|_F \\
&= \|\mathcal{P}_T(\sigma_{abc})\|_F^2 \\
&\leq \frac{28 \mu_0 r}{n_1 n_2}.
\end{aligned}$$

Now we can compute the bound,

$$\left\| \tau_{a_k b_k c_k} - \frac{1}{n_1 n_2 n_3} \mathcal{P}_T \right\| \leq \max\left\{ \frac{28 \mu_0 r}{n_1 n_2}, \frac{1}{n_1 n_2 n_3} \right\} \leq \frac{28 \mu_0 r}{n_1 n_2},$$

where we have utilized the fact that $\|\mathbf{A} - \mathbf{B}\| \leq \max\{\|\mathbf{A}\|, \|\mathbf{B}\|\}$ for positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$. We also have

$$\begin{aligned}
\left\|\mathbb{E}[\tau_{a_k b_k c_k}^2]\right\| &= \left\| \mathbb{E}\left[\|\mathcal{P}_T(\sigma_{a_k b_k c_k}))\|_F^2 \, \tau_{a_k b_k c_k} \right] \right\| \\
&\leq \frac{28 \mu_0 r}{n_1 n_2} \left\| \mathbb{E}\left[\tau_{a_k b_k c_k}\right] \right\| \\
&= \frac{28 \mu_0 r}{n_1^2 n_2^2 n_3}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\| \mathbb{E}\left[\left(\tau_{a_k b_k c_k} - \frac{1}{n_1 n_2 n_3} \mathcal{P}_T \right)^2 \right] \right\| &= \left\| \mathbb{E}[\tau_{a_k b_k c_k}^2] - \frac{1}{n_1^2 n_2^2 n_3^2} \mathcal{P}_T \right\| \\
&\leq \max\left\{ \left\|\mathbb{E}[\tau_{a_k b_k c_k}^2]\right\|, \frac{1}{n_1^2 n_2^2 n_3^2} \right\} \\
&\leq \max\left\{ \frac{28 \mu_0 r}{n_1^2 n_2^2 n_3}, \frac{1}{n_1^2 n_2^2 n_3^2} \right\} \\
&\leq \frac{28 \mu_0 r}{n_1^2 n_2^2 n_3}.
\end{aligned}$$

The lemma follows by applying the noncommuntative Bernstein inequality. □

The next lemma asserts that, for a fixed matrix $\mathbf{E}$, $\mathcal{R}_\Omega^*\mathcal{R}_\Omega(\mathbf{E})$ is close to $\mathcal{O}(\mathbf{E})$ in spectral norm.

Lemma 3. *Suppose Ω is a set of entries of size m which is sampled independent and uniformly with replacement. Then, for any $\beta > 1$ and any $\mathbf{E} \in S$,*

$$\left\| \frac{n_1 n_2 n_3}{m} \mathcal{P}_S \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E} \right\| \le \sqrt{\frac{72\beta n_2 n_3^2 \log(n_1+n_2+n_3)}{m}} \|\mathbf{E}\|_\infty ,$$

holds with probability at least $1-2(n_1+n_2+n_3)^{1-\beta}$ provided that $m > \frac{98}{9}\beta n_2 \log(n_1+n_2+n_3)$.

Proof. Define the operator γ_{abc} which maps $\mathbf{E}$ to $n_1n_2n_3 \langle \mathbf{E}, \sigma_{abc}\rangle \sigma_{abc}$. We can decompose $\frac{n_1n_2n_3}{m}\mathcal{P}_S\mathcal{R}_\Omega^*\mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E}$ as

$$\frac{n_1 n_2 n_3}{m} \mathcal{P}_S \mathcal{R}_\Omega^* \mathcal{R}_\Omega - \mathbf{E} = \frac{1}{m}\left[\sum_k \left(\mathcal{P}_S \gamma_{a_k b_k c_k}(\mathbf{E}) - \mathbf{E}\right)\right].$$

We can bound $\|\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E})\|$ as,

$$\begin{aligned} \|\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E})\| &\le \|\gamma_{a_kb_kc_k}(\mathbf{E})\| \\ &= n_1n_2n_3 \left\|\langle \mathbf{E}, \sigma_{a_kb_kc_k}\rangle \sigma_{a_kb_kc_k}\right\| \\ &\le 3n_2n_3 \|\mathbf{E}\|_\infty \end{aligned}$$

Therefore, we have

$$\|\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}) - \mathbf{E}\| \le \|\mathcal{P}_S\gamma_{a_ksb_kc_k}(\mathbf{E})\| + \|\mathbf{E}\| \le 3n_2n_3\|\mathbf{E}\|_\infty + n_3\|\mathcal{O}(\mathbf{E})\|_\infty \le \frac{7}{2}n_2n_3\|\mathbf{E}\|_\infty ,$$

where we used the fact that $\|\mathbf{E}\| \le n_3\|\mathbf{E}\|_\infty$ and $n\|\mathbf{E}\|_\infty \le \frac{1}{2}n_2n_3\|\mathbf{E}\|_\infty$ for $n_2 \ge 2$. We also have

$$\begin{aligned} \left\|\mathbb{E}\left[(\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}))^* (\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}))\right]\right\| &\le \left\|\mathbb{E}\left[(\gamma_{a_kb_kc_k}(\mathbf{E}))^* (\gamma_{a_kb_kc_k}(\mathbf{E}))\right]\right\| \\ &= \left\|\mathbb{E}\left[n_1^2n_2^2n_3^2 \langle \mathbf{E}, \sigma_{a_kb_kc_k}\rangle^2 \sigma_{a_kb_kc_k}^*\sigma_{a_kb_kc_k}\right]\right\| \\ &= \left\|n_1n_2n_3 \sum_{abc} \langle \mathbf{E}, \sigma_{abc}\rangle^2 \sigma_{abc}^*\sigma_{abc}\right\| \\ &\le \left\|9n_2n_3 \sum_{abc} \|\mathbf{E}\|_\infty^2 \sigma_{abc}^*\sigma_{abc}\right\| \\ &= 9n_2n_3\|\mathbf{E}\|_\infty^2 \left\|\sum_{ab}\delta_{bb}^{(A)} + \sum_{bc}\delta_{bc}^{(B)} + \sum_{ca}\delta_{ca}^{(C)}\right\| \\ &\le 9n_2n_3^2\|\mathbf{E}\|_\infty^2 \end{aligned}$$

We now obtain,

$$\begin{aligned} &\left\|\mathbb{E}\left[(\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}) - \mathbf{E})^* (\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}) - \mathbf{E})\right]\right\| \\ \le\ & \max\left\{\|\mathbf{E}[(\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}))^*(\mathcal{P}_S\gamma_{a_kb_kc_k}(\mathbf{E}))]\| , \|\mathbf{E}^*\mathbf{E}\|\right\} \\ \le\ & 9n_2n_3^2\|\mathbf{E}\|_\infty^2 . \end{aligned}$$

Then the lemma follows by the noncommunative Bernstein Inequality. □

The next concentration result is the final piece for constructing the dual certificate.

Lemma 4. *Suppose Ω is a set of entries of size m sampled independently with replacement. Then for any $\mathbf{E} \in \mathcal{T}$ and any $\beta > 2$, we have*

$$\left\| \frac{n_1 n_2 n_3}{m} \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E} \right\|_\infty \le \sqrt{\frac{864\beta\mu_0 r n_3 \log n_3}{m}} \|\mathbf{E}\|_\infty$$

with probability at least $1 - 6n_3^{2-\beta}$ if $m > 50\beta\mu_0 r n_3 \log(n_3)$.

Proof. Without loss of generality, for each a, b, we define the random variable

$$\chi_{ab^{(A)}} = \left\langle \delta_{ab}^{(A)}, n_1 n_2 n_3 \left\langle \mathbf{E}, \sigma_{a'b'c'} \right\rangle \mathcal{P}_T(\sigma_{a'b'c'}) - \mathbf{E} \right\rangle,$$

where a', b', c' is sampled uniformly random from $[n_1] \times [n_2] \times [n_3]$. We also define $\chi_{bc^{(B)}}$ and $\chi_{ca^{(C)}}$ similarly. We now bound each of $\chi_{ab^{(A)}}$, $\chi_{bc^{(B)}}$ and $\chi_{ca^{(C)}}$ using standard Berstein inequality. By definition, we have $\mathbb{E}[\chi_{ab^{(A)}}] = 0$ and

$$\begin{aligned}
|\chi_{ab^{(A)}}| &\le \left| \left\langle \delta_{ab}^{(A)}, n_1 n_2 n_3 \left\langle \mathbf{E}, \sigma_{a'b'c'} \right\rangle \mathcal{P}_T(\sigma_{a'b'c'}) \right\rangle \right| + \left| \left\langle \delta_{ab}^{(A)}, \mathbf{E} \right\rangle \right| \\
&= n_1 n_2 n_3 \left| \left\langle \mathbf{E}, \sigma_{a'b'c'} \right\rangle \right| \left| \left\langle \delta_{ab}^{(A)}, \mathcal{P}_T(\sigma_{a'b'c'}) \right\rangle \right| + \left| \left\langle \delta_{ab}^{(A)}, \mathbf{E} \right\rangle \right| \\
&\le 3\sqrt{n_1} n_2 n_3 \left\| \mathbf{E} \right\|_\infty \left\| \mathcal{P}_T(\delta_{ab}^{(A)}) \right\|_F \left\| \mathcal{P}_T(\sigma_{a'b'c'}) \right\|_F + \left\| \mathbf{E} \right\|_\infty \\
&\le 90 n_3 \mu_0 r \left\| \mathbf{E} \right\|_\infty
\end{aligned}$$

Now we can also bound $\mathbb{E}[\chi_{ab^{(A)}}^2]$ as follows,

$$\begin{aligned}
\mathbb{E}[\chi_{ab^{(A)}}^2] &= \frac{1}{n_1 n_2 n_3} \sum_{a'b'c'} \left\langle \delta_{ab}^{(A)}, n_1 n_2 n_3 \left\langle \mathbf{E}, \sigma_{a'b'c'} \mathcal{P}_T(\sigma_{a'b'c'}) \right\rangle - \mathbf{E} \right\rangle^2 \\
&= n_1 n_2 n_3 \sum_{a'b'c'} \left\langle \mathbf{E}, \sigma_{a'b'c'} \right\rangle^2 \left\langle \delta_{ab}^{(A)}, \mathcal{P}_T(\sigma_{a'b'c'}) \right\rangle^2 - \left\langle \mathbf{E}, \delta_{ab}^{(A)} \right\rangle^2 \\
&\le n_1 n_2 \sum_{a'b'c'} \left\langle \mathbf{E}, \sigma_{a'b'c'} \right\rangle^2 \left\langle \delta_{ab}^{(A)}, \mathcal{P}_T(\delta_{a'b'}^{(A)}) \right\rangle^2 \\
&\le 9 n_1 n_2 n_3 \left\| \mathbf{E} \right\|_\infty^2 \sum_{a'b'} \left\langle \delta_{ab}^{(A)}, \mathcal{P}_T(\sigma_{a'b'}) \right\rangle^2 \\
&\le 9 n_1 n_2 n_3 \left\| \mathbf{E} \right\|_\infty^2 \left\| \mathcal{P}_T(\delta_{ab}^{(A)}) \right\|_F^2 \\
&\le 162 \mu_0 r n_3 \left\| \mathbf{E} \right\|_\infty^2 .
\end{aligned}$$

Clearly the entry $\left\langle \delta_{ab}^{(A)}, \frac{n_1 n_2 n_3}{m} \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E} \right\rangle$ is the mean value of m i.i.d copies of $\chi_{ab^{(A)}}$. Apply the Bernstein's Inequality, we have

$$\Pr\left[\left| \left\langle \delta_{ab}^{(A)}, \frac{n_1 n_2 n_3}{m} \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E} \right\rangle \right| > \sqrt{\frac{864 \beta \mu_0 r n_3 \log(n_3)}{m}} \left\| \mathbf{E} \right\|_\infty \right] \le 2 n_3^{-\beta}.$$

By union bound, we have

$$\Pr\left[\left\| \frac{n_1 n_2 n_3}{m} \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{E}) - \mathbf{E} \right\|_\infty > \sqrt{\frac{864 \beta \mu_0 r n_3 \log(n_3)}{m}} \left\| \mathbf{E} \right\|_\infty \right] \le 6 n_3^{2-\beta}.$$

□

Finally, we adapt the "golfling scheme" proposed by [4] to construct the dual certificate $\mathbf{F}$.

Lemma 5. *Suppose Ω is a set of entries of size sample independently with replacement for $m > 3600 \max\{\mu_1^2, \mu_0\} r n_3 \beta \log^2(6n_3)$. There exists $\mathbf{F} \in \operatorname{range}(\mathcal{R}_\Omega^*)$ satisfies*

$$\left\| \mathcal{P}_T(\mathbf{F}) - \mathbf{W} \right\|_F \le \sqrt{\frac{r}{2n_3}}, \quad \left\| \mathcal{P}_{T^\perp}(\mathbf{F}) \right\| < \frac{1}{2},$$

with probability at least $1 - 3\log(6n_3)(3n_3)^{2-\beta}$ for all $\beta > 2$.

Proof. Partition m entries of Ω into p partitions of size q, where

$$q \ge 3600 \max\{\mu_0, \mu_1^2\} r n_3 \beta \log(6n_3), \quad p \ge \log(6n_3).$$

Denote Ω_j be the j th partition. By Lemma 2 and union bound, we have

$$\Pr\left[\frac{n_1 n_2 n_3}{q} \left\| \mathcal{P}_T \mathcal{R}_{\Omega_j}^* \mathcal{R}_{\Omega_j} \mathcal{P}_T - \frac{q}{n_1 n_2 n_3} \mathcal{P}_T \right\| \le \frac{1}{2} \quad \text{for all } j \in [p] \right] \ge 1 - \log(6n_3) 2 n_3^{2-2\beta}.$$

Now suppose the above event happens. Define $\mathbf{F}_0 = 0$, $\mathbf{G}_0 = \mathbf{W}$ and

$$\mathbf{F}_j = \mathbf{F}_{j-1} + \frac{n_1 n_2 n_3}{q} \mathcal{R}^*_{\Omega_{j-1}} \mathcal{R}_{\Omega_{j-1}}(\mathbf{G}_{j-1}), \mathbf{G}_j = \mathbf{W} - \mathcal{P}_T(\mathbf{F}_j)$$

for $j \in [p]$. We can now bound $\|\mathbf{G}_j\|_F$ as follows,

$$\begin{aligned}
\|\mathbf{G}_j\|_F &= \|\mathbf{W} - \mathcal{P}_T(\mathbf{F}_j)\|_F \\
&= \left\|\mathbf{W} - \mathcal{P}_T(\mathbf{F}_{j-1}) - \frac{n_1 n_2 n_3}{q} \mathcal{P}_T \mathcal{R}^*_{\Omega_{j-1}} \mathcal{R}_{\Omega_{j-1}}(\mathbf{G}_{j-1})\right\|_F \\
&= \left\|\mathbf{G}_{j-1} - \frac{n_1 n_2 n_3}{q} \mathcal{P}_T \mathcal{R}^*_{\Omega_{j-1}} \mathcal{R}_{\Omega_{j-1}}(\mathbf{G}_{j-1})\right\|_F \\
&\leq \frac{1}{2} \|\mathbf{G}_{j-1}\|_F .
\end{aligned}$$

It follows that $\|\mathbf{G}_p\|_F \leq 2^{-p} \|\mathbf{G}_0\|_F = 2^{-p}\sqrt{3r} \leq \frac{r}{2n_3}$, since $p \geq \log(2n_3) \geq \log_2 \sqrt{2n_3}$. Now choose $\mathbf{F} = \mathbf{F}_p$, it is easy to check that

$$\|\mathcal{P}_T(\mathbf{F}) - \mathbf{W}\|_F \leq \sqrt{\frac{r}{2n_3}}$$

with at least probability $1 - \log(6n_3) 2 n_3^{2-\beta}$.

We now argue that $\mathbf{F}_p$ also satisfies the second inequality in this lemma with high probability. Apply Lemma 3 and Lemma 4, we have

$$\Pr\left[\left\|\frac{n_1 n_2 n_3}{q} \mathcal{P}_S \mathcal{R}^*_{\Omega_j} \mathcal{R}_{\Omega_j}(\mathbf{G}_{j-1}) - \mathbf{G}_{j-1}\right\| \leq \sqrt{\frac{72 n_1 n_2^2 \beta \log(n_1 + n_2 + n_3)}{q}} \|\mathbf{G}_{j-1}\|_\infty\right] \geq 1 - 2(n_1 + n_2 + n_3)^{1-\beta},$$

$$\Pr\left[\left\|\mathbf{G}_{j-1} - \frac{n_1 n_2 n_3}{q} \mathcal{P}_T \mathcal{R}^*_{\Omega_j} \mathcal{R}_{\Omega_j}(\mathbf{G}_{j-1})\right\| \leq \frac{1}{2} \|\mathbf{G}_{j-1}\|_\infty\right] \geq 1 - 6 n_3^{2-\beta}.$$

By union bound, the above random events holds for all $j = 1, \ldots, p$ with probability at least $1 - 2\log(6n_3)(3n_3)^{1-\beta}$. Suppose these random event happens, we can bound $\mathcal{P}_{T^\perp}(\mathbf{F}_p)$ as follows.

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}(\mathbf{F}_p)\| &\leq \sum_{j=1}^{p} \left\|\frac{n_1 n_2 n_3}{q} \mathcal{P}_{T^\perp} \mathcal{R}^*_{\Omega_j} \mathcal{R}_{\Omega_j}(\mathbf{G}_{j-1})\right\| \\
&= \sum_{j=1}^{p} \left\|\mathcal{P}_{T^\perp}\left(\frac{n_1 n_2 n_3}{q} \mathcal{P}_S \mathcal{R}^*_{\Omega_j} \mathcal{R}_{\Omega_j}(\mathbf{G}_{j-1}) - \mathbf{G}_{j-1}\right)\right\| \\
&\leq \sum_{j=1}^{p} \left\|\frac{n_1 n_2 n_3}{q} \mathcal{P}_S \mathcal{R}^*_{\Omega_j} \mathcal{R}_{\Omega_j}(\mathbf{G}_{j-1}) - \mathbf{G}_{j-1}\right\| \\
&\leq \sum_{j=1}^{p} \sqrt{\frac{72 n_2 n_3^2 \beta \log(n_1 + n_2 + n_3)}{q}} \|\mathbf{G}_{j-1}\|_\infty \\
&= 2\sum_{j=1}^{p} 2^{-j} \sqrt{\frac{72 n_2 n_3^2 \beta \log(n_1 + n_2 + n_3)}{q}} \|\mathbf{W}\|_\infty \\
&< \sqrt{\frac{288 \mu_1^2 r n_3 \beta \log(3n_3)}{q}} \\
&< \frac{1}{2},
\end{aligned}$$

where the second equality holds because $\mathcal{P}_{T^\perp}(\mathbf{G}_{j-1}) = \mathbf{0}$ for all j and the last inequality follows since $q > 3600\mu_1^2 r n_3 \beta \log(6n_3)$. Finally, by union bound, the probability that all above random events happen is at least $1 - 3\log(6n_3)(3n_3)^{2-\beta}$. □

Remark 1. *By golfing scheme construction, the dual certificate* $\mathbf{F}$ *can be decomposed by*

$$\mathbf{F} = \sum_{i=1}^{p} \frac{n_1 n_2 n_3}{q} \mathcal{R}^*_{\Omega_i} \mathcal{R}_{\Omega_i}(\mathbf{G}_i),$$

for some $\mathbf{G}_1, \ldots, \mathbf{G}_p$.

We will use this property later for proving the stable recovery result in the presence of noise.

The next lemma is an elementary Chernoff bound which shows that maximum duplication of any entry in Ω when sampling with replacement is bounded by $\frac{8}{3}\beta\log(n_1)$. This gives us the upper bound of the spectral norm of $\mathcal{R}_\Omega$.

Lemma 6. *Suppose* Ω *is a set of entries of size sample independently with replacement for* $m > 3600\max\{\mu_1^2, \mu_0\} r n_3 \beta \log^2(6n_3)$. *We have*

$$\|\mathcal{R}_\Omega\| \leq \sqrt{\frac{8\beta\log(n_3)}{n_1}}$$

for $n_3 \geq 1$ *and* $\beta \geq 1$ *with probability at least* $1 - 3n_3^{2-2\beta}$.

Proof. Given a set of entries $\Omega = \{(a_k, b_k, c_k)\}_{k\in[m]}$ sampled uniformly with replacement, denote the number of reptitions as $\eta^{(A)}_{ab} = |\{k|a_k = a, b_k = b\}|$, $\eta^{(B)}_{bc} = |\{k|b_k = b, c_k = c\}|$ and $\eta^{(C)}_{ca} = |\{k|c_k = c, a_k = a\}|$.

$$\begin{aligned}
\|\mathcal{R}_\Omega\| &= \sup_{\|\mathbf{E}\|_F=1} \|\mathcal{R}_\Omega(\mathbf{E})\|_F \\
&\leq \frac{1}{\sqrt{n_1}} \sup_{\|\mathbf{E}\|_F=1} \sqrt{\sum_{k=1}^{m} \left\langle \mathbf{E}, \delta^{(A)}_{a_k b_k} + \delta^{(B)}_{b_k c_k} + \delta^{(C)}_{c_k a_k} \right\rangle^2} \\
&\leq \frac{1}{\sqrt{n_1}} \sup_{\|\mathbf{E}\|_F=1} \sqrt{3\left[\sum_{k=1}^{m} \left\langle \mathbf{E}, \delta^{(A)}_{a_k b_k}\right\rangle^2 + \sum_{k=1}^{m} \left\langle \mathbf{E}, \delta^{(B)}_{b_k c_k}\right\rangle^2 + \sum_{k=1}^{m} \left\langle \mathbf{E}, \delta^{(C)}_{c_k a_k}\right\rangle^2\right]} \\
&= \frac{1}{\sqrt{n_1}} \sup_{\|\mathbf{E}\|_F=1} \sqrt{3\left[\sum_{ab} \left\langle \mathbf{E}, \delta^{(A)}_{ab}\right\rangle^2 \eta^{(A)}_{ab} + \sum_{bc} \left\langle \mathbf{E}, \delta^{(B)}_{bc}\right\rangle^2 \eta^{(B)}_{bc} + \sum_{ca} \left\langle \mathbf{E}, \delta^{(C)}_{ca}\right\rangle^2 \eta^{(C)}_{ca}\right]} \\
&= \frac{1}{\sqrt{n_1}} \sqrt{3\max\{\max_{ab} \eta^{(A)}_{ab}, \max_{bc} \eta^{(B)}_{bc}, \max_{ca} \eta^{(C)}_{ca}\}}.
\end{aligned}$$

Therefore, it suffices to bound the maximum number of repetitions of any entry in Ω. To this end, we can apply Chernoff bound for the Bernoulli distribution. The probability of an entry a, b be sampled for more than t times can be bounded by Chernoff bound.

$$\Pr\left[\eta^{(A)}_{ab} \geq \frac{8}{3}\beta\log(n_1)\right] \leq \left(\frac{8}{3}\beta\log(n_1)\right)^{-\frac{8}{3}\beta\log(n_1)} \exp\left(\frac{8}{3}\beta\log(n_1)\right) \leq n_1^{-2\beta},$$

when $n_1 \geq 9$. We can also bound $\eta^{(B)}_{bc}$ and $\eta^{(C)}_{ca}$ similarly. By union bound, we have

$$\max\{\max_{ab} \eta^{(A)}_{ab}, \max_{bc} \eta^{(B)}_{bc}, \max_{ca} \eta^{(C)}_{ca}\} \leq \frac{8}{3}\beta\log(n)$$

hold with probability at least $1 - 3n_3^{2-2\beta}$ by union bound. □

Finally, to apply Theorem 4, we require the following bound relating $\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_*$ and $\|\mathcal{P}_T(\mathbf{E})\|_F$ for any fixed matrix $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$.

Lemma 7. *Suppose* Ω *is a set of entries of size sample independently with replacement for* $m > 3600\max\{\mu_1^2, \mu_0\} r n_3 \beta \log^2(2n_3)$. *Then, for any* $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$, *we have*

$$\frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* < \sqrt{\frac{r}{2n_3}} \|\mathcal{P}_T(\mathbf{E})\|_F,$$

with probability at least $1 - 3n_3^{2-\beta}$.

Proof. Since $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$, we have

$$0 = \|\mathcal{R}_\Omega(\mathbf{E})\|_F \geq \|\mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E})\|_F - \|\mathcal{R}_\Omega \mathcal{P}_{T^\perp}(\mathbf{E})\|_F .$$

Apply Lemma 2,

$$\frac{n_1 n_2 n_3}{m} \left\| \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T - \frac{m}{n_1 n_2 n_3} \mathcal{P}_T \right\| \leq \frac{1}{2} \tag{9}$$

holds with probability at least $1 - 3n_3^{2-\beta}$. Suppose Eq. (9) holds, we can bound $\|\mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E})\|_F$ as follows

$$\begin{aligned} \|\mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E})\|_F^2 &= \langle \mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E}), \mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E}) \rangle \\ &= \langle \mathbf{E}, \mathcal{P}_T \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E}) \rangle \\ &\geq \frac{m}{2 n_1 n_2 n_3} \|\mathcal{P}_T(\mathbf{E})\|_F^2 . \end{aligned}$$

On the other hand, we need to bound $\|\mathcal{R}_\Omega \mathcal{P}_{T^\perp}(\mathbf{E})\|_F$. Suppose $\|\mathcal{R}_\Omega\| \leq \sqrt{\frac{8\beta \log(n_3)}{n_1}} \leq \sqrt{\frac{8\beta}{n_1}} \log(n_3)$ which holds with probability at least $1 - n_3^{2-\beta}$ by Lemma 6, we have

$$\begin{aligned} \|\mathcal{R}_\Omega \mathcal{P}_{T^\perp}(\mathbf{E})\|_F &\leq \|\mathcal{R}_\Omega\| \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_F \\ &\leq \sqrt{\frac{8\beta}{n_1}} \log(n_3) \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_F . \end{aligned}$$

Therefore,

$$\begin{aligned} \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* &\geq \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_F \\ &\geq \sqrt{\frac{n_1}{8\beta \log^2(n_3)}} \|\mathcal{R}_\Omega \mathcal{P}_{T^\perp}(\mathbf{E})\|_F \\ &\geq \sqrt{\frac{n_1}{8\beta \log^2(n_3)}} \|\mathcal{R}_\Omega \mathcal{P}_T(\mathbf{E})\|_F \\ &\geq \sqrt{\frac{m n_1}{8 n_1 n_2 n_3 \beta \log^2(n_3)}} \|\mathcal{P}_T(\mathbf{E})\|_F \\ &\geq \sqrt{\frac{3600 n_3 n_1 r \mu_0 \beta \log^2(6n_3)}{8 n_1 n_2 n_3 \beta \log^2(6n)}} \|\mathcal{P}_T(\mathbf{E})\|_F \\ &> \sqrt{\frac{2r}{n_3}} \|\mathcal{P}_T(\mathbf{E})\|_F . \end{aligned}$$

□

We are now ready to prove the exact recovery result Theorem 1.

Proof. (Theorem 1) By Lemma 5, there exists $\mathbf{F} \in \mathrm{range}(\mathcal{R}_\Omega^*)$ such that

$$\|\mathcal{P}_T(\mathbf{F}) - \mathbf{W}\|_F \leq \sqrt{\frac{r}{2n_3}}, \quad \|\mathcal{P}_{T^\perp}(\mathbf{F})\| < \frac{1}{2},$$

with probability at least $1 - 3\log(6n_3)(3n_3)^{2-\beta}$ for all $\beta > 2$.

On the other hand, Lemma 7 shows that for any $\mathbf{E} \in \ker(\mathcal{R}_\Omega)$,

$$\frac{1}{2} \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* > \sqrt{\frac{r}{2n_3}} \|\mathcal{P}_T(\mathbf{E})\|_F ,$$

holds with probability at least $1 - 3n_3^{2-\beta}$.

By union bound, the above random events happen simultaneously with probability at least $1 - 3\log(6n_3)(3n_3)^{2-\beta} - 3n_3^{2-\beta}$. Finally, in the case that both random events holds, by Theorem. 4, the solution to Eq. (1) is unique and exactly recovers $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and therefore the pairwise interaction tensor $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$. □

3 Proof of Theorem 2

In this section, we generalize the proof of Theorem 2 for the noisy cases.

Proof. (Theorem 2) First, define operator $\mathcal{Q} = \gamma^2 \mathcal{R}_\Omega^* \mathcal{R}_\Omega$, where $\gamma = \|\mathcal{R}_\Omega\|^{-1}$, as the normalized version of $\mathcal{R}_\Omega^* \mathcal{R}_\Omega$. Clearly, we have $\|\mathcal{Q}\| = 1$. By Lemma 6, we can bound γ by $\gamma \geq \sqrt{\frac{n_1}{8\beta \log(n_3)}}$.

We can decompose the optimal solution $\hat{\mathbf{M}}$ of the convex progam Eq. (3) into the sum of the true matrix $\mathbf{M}$ and the error matrix $\mathbf{E}$, namely, $\hat{\mathbf{M}} = \mathbf{M} + \mathbf{E}$. To prove the theorem, we need to bound the error term $\mathbf{E}$ in its nuclear norm $\|\mathbf{E}\|_*$. To do this, we start with bounding $\|\mathcal{Q}(\mathbf{E})\|_F$. Denote the noisy observations as an m-dimensional vector $\mathbf{y}$, where $y_i = T_{a_i b_i c_i}$. We have

$$
\begin{aligned}
\|\mathcal{Q}(\mathbf{E})\|_F &\leq \left\|\mathcal{Q}(\hat{\mathbf{M}}) - \gamma^2 \mathcal{R}_\Omega^*(\mathbf{y})\right\|_F + \left\|\gamma^2 \mathcal{R}_\Omega^*(\mathbf{y}) - \mathcal{Q}(\mathbf{M})\right\|_F \\
&= \gamma^2 \left\|\mathcal{R}_\Omega^* \mathcal{R}_\Omega(\hat{\mathbf{M}}) - \mathcal{R}_\Omega^*(\mathbf{y})\right\|_F + \gamma^2 \left\|\mathcal{R}_\Omega^*(\mathbf{y}) - \mathcal{R}_\Omega^* \mathcal{R}_\Omega(\mathbf{M})\right\|_F \\
&\leq \gamma^2 \left\|\mathcal{R}_\Omega^*\right\| \left\|\mathcal{R}_\Omega(\hat{\mathbf{M}}) - \mathbf{y}\right\|_F + \gamma^2 \left\|\mathcal{R}_\Omega^*\right\| \left\|\mathbf{y} - \mathcal{R}_\Omega(\mathbf{M})\right\|_F \\
&\leq \gamma\epsilon_1 + \gamma\epsilon_2 \triangleq \delta. \qquad (10)
\end{aligned}
$$

In the last inequality, the first term $\left\|\mathcal{R}_\Omega(\hat{\mathbf{M}}) - \mathbf{y}\right\|_F$ is no greater than ϵ_2 due to constraint of optimization problem and the second term $\|\mathbf{y} - \mathcal{R}_\Omega(\mathbf{M})\|_F \leq \epsilon_1$ is the assumption on the observation noise.

On the other hand, we can bound $\|\mathcal{Q}(\mathbf{E})\|_F$ by following

$$
\|\mathcal{Q}(\mathbf{E})\|_F \geq \|\mathcal{Q}\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F - \|\mathcal{Q}\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_F .
$$

For the second term, we have $\|\mathcal{Q}\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_F \leq \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_F$. Now, we focus on the first term, we have

$$
\begin{aligned}
\|\mathcal{Q}\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F &= \gamma^2 \|\mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_\mathcal{T}(\mathbf{E})\|_F \\
&\geq \gamma^2 \|\mathcal{P}_\mathcal{T} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_\mathcal{T}(\mathbf{E})\|_F \\
&\geq \gamma^2 \frac{m}{n_1 n_2 n_3} \left\|\frac{n_1 n_2 n_3}{m} \mathcal{P}_\mathcal{T} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_\mathcal{T}(\mathbf{E})\right\|_F \\
&\geq \gamma^2 \frac{m}{n_1 n_2 n_3} \left[\|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F - \left\|\frac{n_1 n_2 n_3}{m} \mathcal{P}_\mathcal{T} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_\mathcal{T}(\mathbf{E}) - \mathcal{P}_\mathcal{T}(\mathbf{E})\right\|_F\right] \\
&\geq \gamma^2 \frac{m}{n_1 n_2 n_3} \left[\|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F - \left\|\frac{n_1 n_2 n_3}{m} \mathcal{P}_\mathcal{T} \mathcal{R}_\Omega^* \mathcal{R}_\Omega \mathcal{P}_\mathcal{T} - \mathcal{P}_\mathcal{T}\right\| \|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F\right] \\
&\geq \gamma^2 \frac{m}{n_1 n_2 n_3} \frac{1}{2} \|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F \geq \frac{m}{16\beta \log(n_3) n_2 n_3} \|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F
\end{aligned}
$$

Therefore, we have

$$
\|\mathcal{Q}(\mathbf{E})\|_F \geq \frac{m}{16\beta n_2 n_3 \log(n_3)} \|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F - \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_F . \qquad (11)
$$

Now, combine Eq. (10) and Eq. (11), we have

$$
\|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F \leq \frac{16\beta n_2 n_3 \log(n_3)}{m} \left(\delta + \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_F\right) \leq \frac{16\beta n_2 n_3 \log(n_3)}{m} \left(\delta + \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_*\right).
$$

Next, we proceed to bound $\|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_*$. We can use a similar subgradient argument as in the proof of Theorem 4. Let $\mathbf{F} \in \mathrm{range}(\mathcal{R}_\Omega^*)$ be the dual certificate as described in Theorem 4, we have

$$
\begin{aligned}
\left\|\hat{\mathbf{M}}\right\|_* &= \|\mathbf{M} + \mathbf{E}\|_* \\
&\geq \langle \mathbf{W} + \mathbf{W}_\perp, \mathbf{M} + \mathbf{E} \rangle = \|\mathbf{M}\|_* + \langle \mathbf{W} + \mathbf{W}_\perp, \mathbf{E} \rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} + \mathbf{W}_\perp - \mathbf{F}, \mathbf{E} \rangle + \langle \mathbf{F}, \mathbf{E} \rangle \\
&= \|\mathbf{M}\|_* + \langle \mathbf{W} - \mathcal{P}_T(\mathbf{F}), \mathcal{P}_\mathcal{T}(\mathbf{E}) \rangle + \langle \mathbf{W}_\perp - \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{F}), \mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E}) \rangle + \langle \mathbf{F}, \mathbf{E} \rangle \\
&\geq \|\mathbf{M}\|_* - \frac{\sqrt{r}}{n_3^2} \|\mathcal{P}_\mathcal{T}(\mathbf{E})\|_F + \frac{1}{2} \|\mathcal{P}_{\mathcal{T}^\perp}(\mathbf{E})\|_* + \langle \mathbf{F}, \mathbf{E} \rangle .
\end{aligned}
$$

Recall that $\hat{\mathbf{M}}$ is the optimal solution to the convex program Eq. (3), we have $\left\|\hat{\mathbf{M}}\right\|_* \le \|\mathbf{M}\|_*$. Hence, we have

$$\frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \le \frac{\sqrt{r}}{n_3^2}\|\mathcal{P}_T(\mathbf{E})\|_F - \langle \mathbf{F}, \mathbf{E}\rangle .$$

Now we bound $\langle \mathbf{F}, \mathbf{E}\rangle$. By the golfling scheme construction of $\mathbf{F}$ in Lemma. 5, we have

$$\begin{aligned}\langle \mathbf{F}, \mathbf{E}\rangle &= \sum_{i=1}^{p}\left\langle \mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{G}_i), \mathbf{E}\right\rangle \\ &= \sum_{i=1}^{p}\left\langle \mathbf{G}_i, \mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{E})\right\rangle \\ &\ge -\sum_{i=1}^{p}\|\mathbf{G}_i\|_F\left\|\mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{E})\right\|_F .\end{aligned}$$

For each i, we can bound $\left\|\mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{E})\right\|_F$ by

$$\begin{aligned}\left\|\mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{E})\right\|_F &= \left\|\mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\hat{\mathbf{M}}) - \mathcal{R}^*_{\Omega_i}(\mathbf{y}_{\Omega_i})\right\|_F + \left\|\mathcal{R}^*_{\Omega_i}(\mathbf{y}_{\Omega_i}) - \mathcal{R}^*_{\Omega_i}\mathcal{R}_{\Omega_i}(\mathbf{M})\right\|_F \\ &\le \left\|\mathcal{R}^*_{\Omega_i}\right\|\left\|\mathcal{R}_{\Omega_i}(\hat{\mathbf{M}}) - \mathbf{y}_{\Omega_i}\right\|_F + \left\|\mathcal{R}^*_{\Omega_i}\right\|\|\mathbf{y}_{\Omega_i} - \mathcal{R}_{\Omega_i}(\mathbf{M})\|_F \\ &= \frac{\epsilon_1}{\gamma} + \frac{\epsilon_2}{\gamma},\end{aligned}$$

where $\mathbf{y}_{\Omega_i}$ is the restriction of $\mathbf{y}$ on Ω_i.

Therefore, we have

$$\begin{aligned}\langle \mathbf{F}, \mathbf{E}\rangle &\ge -\frac{\epsilon_1+\epsilon_2}{\gamma}\sum_{i=1}^{p}\|\mathbf{G}_i\|_F \\ &\ge -\frac{2(\epsilon_1+\epsilon_2)}{\gamma}\|\mathbf{G}_0\|_F \\ &= -\frac{2(\epsilon_1+\epsilon_2)}{\gamma}\|\mathbf{W}\|_F \ge -\frac{2(\epsilon_1+\epsilon_2)}{\gamma}\sqrt{r_1+r_2+r_3}.\end{aligned}$$

Consequently, for reasonable values of parameters, we have

$$\begin{aligned}\frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* &\le \frac{\sqrt{r}}{n_3^2}\|\mathcal{P}_T(\mathbf{E})\|_F + \frac{2(\epsilon_1+\epsilon_2)}{\gamma}\sqrt{r_1+r_2+r_3} \\ &\le \frac{16\beta\log(n_3)\sqrt{r}}{m}(\delta + \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_*) + \frac{2(\epsilon_1+\epsilon_2)}{\gamma}\sqrt{r_1+r_2+r_3} \\ &\le \frac{1}{16}(\delta + \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_*) + \frac{\epsilon_1+\epsilon_2}{16}.\end{aligned}$$

Hence, we have

$$\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \le \frac{16}{7}\left(\delta + \frac{\epsilon_1+\epsilon_2}{16}\right) \le 3\delta.$$

Last, combining the above inequalities and setting $\epsilon = \epsilon_1 + \epsilon_2$, we can finally bound the error $\mathbf{E}$ in terms of its nuclear norm as follows

$$\begin{aligned}\|\mathbf{E}\|_* &\le \sqrt{2r}\|\mathcal{P}_T(\mathbf{E})\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \\ &\le \sqrt{2r}n_2\delta + (\sqrt{2r}n_2 + 1)\|\mathcal{P}_{T^\perp}(\mathbf{E})\|_* \\ &\le \sqrt{2r}n_2\gamma\epsilon + 3(\sqrt{2r}n_2 + 1)\gamma\epsilon \\ &\le 5\sqrt{2r}n_2\gamma\epsilon \\ &\le 5\sqrt{\frac{2rn_1n_2^2}{8\beta\log(n_1)}}\epsilon.\end{aligned}$$

□

4 Proof of Proposition 1

We first review some terminologies. We call a matrix $\mathbf{A}$ a *doubly centered matrix*, if each column and each row of $\mathbf{A}$ sums up to zero, i.e. $\mathbf{1}^T\mathbf{A} = \mathbf{0}^T$ and $\mathbf{A}\mathbf{1} = \mathbf{0}$ hold simultaenously. We also call a vector $\mathbf{v}$ a *centered vector*, if the sum of its entries equals to zero, namely, $\mathbf{1}^T\mathbf{v} = 0$.

Lemma 8. *Given an arbitrary pariwise interaction tensor $\mathcal{T} = \mathrm{Pair}(\mathbf{A}, \mathbf{B}, \mathbf{C})$, there exists a unique 7-tuple $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ such that $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0$ are doubly centered matrices, $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are centered vectors and satisfies*

$$T_{ijk} = A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d, \quad \textit{for all } (i, j, k) \in [n_1] \times [n_2] \times [n_3]. \tag{12}$$

Remark 2. *We can interpret the quantities $\mathbf{a}, \mathbf{b}, \mathbf{c}, d$ in Lemma 8 as axis-aligned biases of tensor $\mathcal{T}$. For example, every entries of the form T_{1jk} are influenced by bias a_1; the entries of form T_{i1k} for all (i, k) are biased by b_1; the entries of form T_{ij1} for all (i, j) are biased by c_1. In addition, all entries of $\mathcal{T}$ is biased by d.*

Proof. (Lemma 8) In the following, we shall prove the existence and uniqueness separately.

Existence. Given any $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ of appropriate size, we now construct the 7-tuple $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ specified in the lemma.

We define the mean values of matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ by $\sigma_A = \frac{1}{n_1 n_2}\mathbf{1}^T\mathbf{A}\mathbf{1}$, $\sigma_B = \frac{1}{n_2 n_3}\mathbf{1}^T\mathbf{B}\mathbf{1}$ and $\sigma_C = \frac{1}{n_3 n_1}\mathbf{1}^T\mathbf{B}\mathbf{1}$. We also denote the mean vectors of columns of matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ by $\mathbf{a}_c = \frac{1}{n_1}\mathbf{A}^T\mathbf{1}$, $\mathbf{b}_c = \frac{1}{n_2}\mathbf{B}^T\mathbf{1}$ and $\mathbf{c}_c = \frac{1}{n_3}\mathbf{C}^T\mathbf{1}$. Similarly, we denote the mean vectors of rows of matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ by $\mathbf{a}_r = \frac{1}{n_2}\mathbf{A}\mathbf{1}$, $\mathbf{b}_r = \frac{1}{n_3}\mathbf{B}\mathbf{1}$ and $\mathbf{c}_r = \frac{1}{n_1}\mathbf{C}\mathbf{1}$.

Now, we construct the desired 7-tuple $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ by

$$\begin{gathered} A^0_{ij} = A_{ij} - a^c_j - a^r_i + \sigma_a, \quad B^0_{jk} = B_{jk} - b^c_j - b^r_k + \sigma_b, \quad C^0_{ki} = C_{ki} - c^c_i - c^r_k + \sigma_c \\ a_i = a^r_i + c^c_i - \sigma_a - \sigma_c, \quad b_j = b^r_j + a^c_j - \sigma_b - \sigma_a, \quad c_k = c^r_k + b^c_k - \sigma_c - \sigma_b \\ d = \sigma_a + \sigma_b + \sigma_c, \end{gathered}$$

where (i, j, k) ranges within $[n_1] \times [n_2] \times [n_3]$. It is easy to verify that $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0$ are doubly centered matrices and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are centered vectors and that $A_{ij} + B_{jk} + C_{ki} = A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d$.

Uniqueness. Suppose there exists two 7-tuples $(\mathbf{A}^0_1, \mathbf{B}^0_1, \mathbf{C}^0_1, \mathbf{a}_1, \mathbf{b}_1, \mathbf{c}_1, d_1)$ and $(\mathbf{A}^0_2, \mathbf{B}^0_2, \mathbf{C}^0_2, \mathbf{a}_2, \mathbf{b}_2, \mathbf{c}_2, d_2)$ that satisfy the centering property specified in the lemma. Consider their differences $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d) = (\mathbf{A}^0_1 - \mathbf{A}^0_2, \mathbf{B}^0_1 - \mathbf{B}^0_2, \mathbf{C}^0_1 - \mathbf{C}^0_2, \mathbf{a}_1 - \mathbf{a}_2, \mathbf{b}_1 - \mathbf{b}_2, \mathbf{c}_1 - \mathbf{c}_2, d_1 - d_2)$. Clearly, we remain to show that $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ are zeros.

It is clear $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0$ are doubly centered matrices and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are centered vectors. In addition, the following holds for all (i, j, k)

$$0 = A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d. \tag{13}$$

We first show $d = 0$. We can see this by summing over all i, j, k on both sides of Eq. (13),

$$0 = \sum_{ijk}[A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d] = n_1 n_2 n_3 d,$$

where the first inequality holds by the centering properties on $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}$.

Next, we show that $\mathbf{a} = \mathbf{0}$. This can be done by summing over all $(j, k) \in [n_2] \times [n_3]$ on both sides of Eq. (13) and for any i,

$$\begin{aligned} 0 &= \sum_{jk}[A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d] \\ &= \sum_{jk}[A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k] \\ &= n_2 n_3 a_i, \end{aligned}$$

where we have used the result that $d = 0$ and the centering properties. Similarly, we can show $\mathbf{b} = \mathbf{0}$ and $\mathbf{c} = \mathbf{0}$.

Finally, we remain to show $\mathbf{A} = \mathbf{0}$. Again, fix any i, j and sum over all $k \in [n_3]$, we have

$$0 = \sum_k [A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d] = n_3 A^0_{ij},$$

where we have used the facts that $a_i = b_j = c_k = d = 0$. We can prove $\mathbf{B} = \mathbf{0}$ and $\mathbf{C} = \mathbf{0}$ using similar arguments. □

Lemma 8 essentially states that the representation of a pairwise interaction tensor is unique if one separate out these bias components. We can immmediately obtain Proposition 1 by condensing the unique representation scheme $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ for pairwise interaction tensors identified by Lemma 8. In particular, we construct $\mathbf{A}' \in S_A, \mathbf{B}' \in S_B, \mathbf{C}' \in S_C$ by setting $A'_{ij} = A^0_{ij} + a_i + d$, $B'_{jk} = B^0_{jk} + b_j$ and $C'_{ij} = C^0_{ki} + c_k$. By the centering property of $\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0$ and $\mathbf{a}, \mathbf{b}, \mathbf{c}$, it is clear that each column of $\mathbf{A}'$ sums up to a same value $(n_1 d)$ and each column of $\mathbf{B}', \mathbf{C}'$ sums up to zero. Hence $\mathbf{A}', \mathbf{B}', \mathbf{C}'$ satisfy the constraints defined by S_A, S_B, S_C respectively. We can also easily show the uniqueness of $\mathbf{A}', \mathbf{B}', \mathbf{C}'$ under this constraints using the uniqueness of $(\mathbf{A}_0, \mathbf{B}_0, \mathbf{C}_0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$.

Proof. (Propostion 1) The existence follows immediately from Lemma 8. Specifically, we can set $A'_{ij} = A^0_{ij} + a_i + d$, $B'_{jk} = B^0_{jk} + b_j$ and $C'_{ki} = C^0_{ki} + c_k$. We can easily verify that $\mathbf{A}' \in S_A, \mathbf{B}' \in S_B$ and $\mathbf{C}' \in S_C$.

Now we prove the uniqueness. Suppose that we have $\mathrm{Pair}(\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_1) = \mathrm{Pair}(\mathbf{A}_2, \mathbf{B}_2, \mathbf{C}_2)$, where $\mathbf{A}_1 \in S_A$, $\mathbf{A}_2 \in S_A$, $\mathbf{B}_1 \in S_B$, $\mathbf{B}_2 \in S_B$, $\mathbf{C}_1 \in S_C$ and $\mathbf{C}_2 \in S_C$. Denote $\mathbf{A} = \mathbf{A}_1 - \mathbf{A}_2$, $\mathbf{B} = \mathbf{B}_1 - \mathbf{B}_2$ and $\mathbf{C} = \mathbf{C}_1 - \mathbf{C}_2$, we remain to show that the differences $\mathbf{A}$,$\mathbf{B}$ and $\mathbf{C}$ are zero.

Note that $\mathbf{A} \in S_A, \mathbf{B} \in S_B, \mathbf{C} \in S_C$. We can construct 7-tuple $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ similarly to the proof of Lemma 8. We define the mean values of matrix $\mathbf{A}$ by $\sigma_A = \frac{1}{n_1 n_2} \mathbf{1}^T \mathbf{A} \mathbf{1}$ (note that the mean value of $\mathbf{B}$ and $\mathbf{C}$ is zero). We denote the mean vectors of rows of matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ by $\mathbf{a}_r = \frac{1}{n_2} \mathbf{A} \mathbf{1}$, $\mathbf{b}_r = \frac{1}{n_3} \mathbf{B} \mathbf{1}$ and $\mathbf{c}_r = \frac{1}{n_1} \mathbf{C} \mathbf{1}$.

Now, we construct the desired 7-tuple $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ by

$$A^0_{ij} = A_{ij} - a^r_i + \sigma_a, \quad B^0_{jk} = B_{jk} - b^r_k, \quad C^0_{ki} = C_{ki} - c^r_k,$$
$$a_i = a^r_i - \sigma_a,$$

where (i, j, k) ranges within $[n_1] \times [n_2] \times [n_3]$. In addition, we set $\mathbf{b} = \mathbf{b}_r, \mathbf{c} = \mathbf{c}_r$ and $d = \sigma_a$. We can verify that $A^0_{ij} + B^0_{jk} + C^0_{ki} + a_i + b_j + c_k + d = A_{ij} + B_{jk} + C_{ki} = 0$. By Lemma 8, it follows immediately that $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{a}, \mathbf{b}, \mathbf{c}, d)$ are zeros. Therefore, we have $\mathbf{A} = \mathbf{0}, \mathbf{B} = \mathbf{0}$ and $\mathbf{C} = \mathbf{0}$.

□

5 Details of recovery algorithm

Algorithm 1 Exact Recovery of Pairwise Interaction Tensor

procedure EXACTRECOVER($\Omega = \{a_i b_i c_i\}_{i\in[m]}, \mathcal{P}_\Omega(\mathcal{T}) = \{T_{a_i b_i c_i}\}_{i\in[m]}, \tau, \delta, \epsilon$)
  $\mathbf{y} \leftarrow \mathbf{0}$
  for $k = 1, \dots, k_{\max}$ **do**
    $[\mathbf{X}, r_A] \leftarrow \text{shrink}_A(\mathcal{P}^*_{\Omega_A}(\mathbf{y}), \tau, r_A)$
    $[\mathbf{Y}, r_B] \leftarrow \text{shrink}_B(\mathcal{P}^*_{\Omega_B}(\mathbf{y}), \tau, r_B)$
    $[\mathbf{Z}, r_C] \leftarrow \text{shrink}_B(\mathcal{P}^*_{\Omega_C}(\mathbf{y}), \tau, r_C)$ ▷ shrink_C is algorithmically identical to shrink_B.
    $\mathbf{e} \leftarrow \mathcal{P}_\Omega(\mathcal{T}) - \mathcal{P}_\Omega(\text{Pair}(n_3^{-1/2}\mathbf{X}, n_1^{-1/2}\mathbf{Y}, n_2^{-1/2}\mathbf{Z})$
    if $\|\mathbf{e}\|_F \,/\, \|\mathcal{P}_\Omega(\mathcal{T})\|_F \le \epsilon$ **then**
      break
    end if
    $\mathbf{y} \leftarrow \mathbf{y} + \delta\mathbf{e}$
  end for
end procedure
return $[n_3^{-1/2}\mathbf{X}, n_1^{-1/2}\mathbf{Y}, n_2^{-1/2}\mathbf{Z}]$

Algorithm 2 Stable Recovery of Pairwise Interaction Tensor

procedure STABLERECOVER($\Omega = \{a_i b_i c_i\}_{i\in[m]}, \mathcal{P}_\Omega(\hat{\mathcal{T}}) = \{\hat{T}_{a_i b_i c_i}\}_{i\in[m]}, \tau, \delta, \epsilon, \epsilon_1$)
  $\mathbf{y} \leftarrow \mathbf{0}$
  $s \leftarrow 0$
  for $k = 1, \dots, k_{\max}$ **do**
    $[\mathbf{X}, r_A] \leftarrow \text{shrink}_A(\mathcal{P}^*_{\Omega_A}(\mathbf{y}), \tau, r_A)$
    $[\mathbf{Y}, r_B] \leftarrow \text{shrink}_B(\mathcal{P}^*_{\Omega_B}(\mathbf{y}), \tau, r_B)$
    $[\mathbf{Z}, r_C] \leftarrow \text{shrink}_B(\mathcal{P}^*_{\Omega_C}(\mathbf{y}), \tau, r_C)$
    $\mathbf{e} \leftarrow \mathcal{P}_\Omega(\hat{\mathcal{T}}) - \mathcal{P}_\Omega(\text{Pair}(n_3^{-1/2}\mathbf{X}, n_1^{-1/2}\mathbf{Y}, n_2^{-1/2}\mathbf{Z})$
    if $\|\mathbf{e}\|_F \,/\, \left\|\mathcal{P}_\Omega(\hat{\mathcal{T}})\right\|_F \le \epsilon$ **then**
      break
    end if
    $\mathbf{y} \leftarrow \mathbf{y} + \delta\mathbf{e}$
    $s \leftarrow s - \delta\epsilon_1$
    $[\mathbf{y}, s] \leftarrow \mathcal{P}_\mathcal{K}(\mathbf{y}, s)$
  end for
end procedure
return $[n_3^{-1/2}\mathbf{X}, n_1^{-1/2}\mathbf{Y}, n_2^{-1/2}\mathbf{Z}]$

Algorithm 3 Shrinkage operator

procedure SHRINK$_B$($\hat{\mathbf{X}}, \tau, r$)
  $s \leftarrow r + 1$
  repeat
    $[\mathbf{U}, \Sigma, \mathbf{V}] \leftarrow \text{svd}(\text{center}(\hat{\mathbf{X}}), s)$ ▷ $\text{svd}(\mathbf{M}, s)$: return top s singular vectors of $\mathbf{M}$
    $s \leftarrow s + 5$
  until $\sigma_{s-5} \le \tau$
  $r \leftarrow \max\{j : \sigma_j > \tau\}$
  $\mathbf{X} \leftarrow \sum_{i=1}^{r}(\sigma_j - \tau)\mathbf{u}_j\mathbf{v}_j^*$
  return $[\mathbf{X}, r]$
end procedure
procedure SHRINK$_A$($\hat{\mathbf{X}}, \tau, r$)
  $[\mathbf{X}, r] \leftarrow \text{shrink}_B(\hat{\mathbf{X}}, \tau, r)$
  $\delta \leftarrow \text{sum}(\hat{\mathbf{X}})$ ▷ $\text{sum}(\hat{\mathbf{X}})$: elemetwise sum of matrix $\hat{\mathbf{X}}$
  $\gamma \leftarrow \frac{1}{n_1 n_2}(\{\delta - \tau\}_+ + \{\delta + \tau\}_-)$
  return $[\mathbf{X} + \gamma\mathbf{1}\mathbf{1}^T, r]$
end procedure

References

[1] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on*, 48(3):569–579, 2002.

[2] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[3] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[4] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[5] Benjamin Recht. A simpler approach to matrix completion. *JMLR*, 2011.