Supplemental Materials for
"Spectral Compressed Sensing via Structured Matrix Completion"

May 1, 2013

Abstract

This supplemental document presents details concerning analytical derivations that support the theorems made in the main text "Spectral Compressed Sensing via Structured Matrix Completion", accepted to the 30th International Conference on Machine Learning (ICML 2013). One can find here the detailed proof of Theorems 1- 3.

1 A Summary of Notation

Let the singular value decomposition (SVD) of $\boldsymbol{X}_{\mathrm{e}}$ be $\boldsymbol{X}_{\mathrm{e}} = \boldsymbol{U}\Lambda\boldsymbol{V}^*$. Denote by

$$T := \left\{ \boldsymbol{U}\boldsymbol{M}^* + \tilde{\boldsymbol{M}}\boldsymbol{V}^* : \boldsymbol{M} \in \mathbb{C}^{(n_1-k_1+1)(n_2-k_1+1)\times r}; \tilde{\boldsymbol{M}} \in \mathbb{C}^{k_1 k_2 \times r} \right\}$$

the tangent space with respect to $\boldsymbol{X}_{\mathrm{e}}$, and $T^{\perp}$ the orthogonal complement of T. Denote by $\mathcal{P}_U$ (resp. $\mathcal{P}_V$, $\mathcal{P}_T$) the orthogonal projections onto the column (resp. row, tangent) space of $\boldsymbol{X}_{\mathrm{e}}$, i.e. for any $\boldsymbol{M}$

$$\mathcal{P}_U\boldsymbol{M} = \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{M}; \quad \mathcal{P}_V\boldsymbol{M} = \boldsymbol{M}\boldsymbol{V}\boldsymbol{V}^*; \quad \mathcal{P}_T = \mathcal{P}_U + \mathcal{P}_V - \mathcal{P}_U\mathcal{P}_V.$$

We let $\mathcal{P}_{T^{\perp}} = \mathcal{I} - \mathcal{P}_T$ be the orthogonal complement of $\mathcal{P}_T$, where $\mathcal{I}$ denotes the identity operator.

We denote by $\|\boldsymbol{M}\|_*$, $\|\boldsymbol{M}\|_{\mathrm{F}}$, $\|\boldsymbol{M}\|$ the nuclear norm, the Frobenious norm, and the spectral norm (or operator norm) of $\boldsymbol{M}$, respectively. The inner product between two matrices is defined as

$$\langle \boldsymbol{B}, \boldsymbol{C} \rangle = \mathrm{trace}\left(\boldsymbol{B}^*\boldsymbol{C}\right).$$

Besides, we denote by $\Omega_{\mathrm{e}}(i,l)$ the set of locations of the enhanced matrix $\boldsymbol{X}_{\mathrm{e}}$ containing copies of $x_{i,l}$. Due to the Hankel and block-Hankel structures, one can easily verify the following: for any $\Omega_{\mathrm{e}}(i,l)$, there exists at most one index lying in any given row of the enhanced form, and at most one index coming from any given column. For each $(i,l) \in [n_1] \times [n_2]$, we use $\boldsymbol{A}_{(i,l)}$ to denote a basis matrix that extracts the average of all entries in $\Omega_{\mathrm{e}}(i,l)$. Specifically,

$$\left(\boldsymbol{A}_{(i,l)}\right)_{\alpha,\beta} := \begin{cases} \frac{1}{\sqrt{|\Omega_{\mathrm{e}}(i,l)|}}, & \text{if } (\alpha,\beta) \in \Omega_{\mathrm{e}}(i,l), \\ 0, & \text{else.} \end{cases} \tag{1}$$

We will use $\omega_{i,l} := |\Omega_{\mathrm{e}}(i,l)|$ as a short hand notation.

2 A List of Main Theorems

For convenience of presentation, we restate our main theorems in this section, which are the subjects to prove in this manuscript.

Definition 1. *[**Incoherence**]Let $\boldsymbol{X}_e$ denote the enhanced matrix associated with $\boldsymbol{X}$, and suppose the SVD of $\boldsymbol{X}_e$ is given by $\boldsymbol{X}_e = \boldsymbol{U}\Lambda\boldsymbol{V}^*$. Then $\boldsymbol{X}$ is said to have incoherence (μ_1, μ_2, μ_3) if they are respectively the smallest values obeying*

$$\sigma_{\min}\left(\boldsymbol{G}_L\right) \geq \frac{1}{\mu_1}, \quad \sigma_{\min}\left(\boldsymbol{G}_R\right) \geq \frac{1}{\mu_1}; \tag{2}$$

$$\max_{(i,l)\in[n_1]\times[n_2]} \frac{1}{\left|\Omega_e(i,l)\right|^2} \left| \sum_{(\alpha,\beta)\in\Omega_e(i,l)} \left(\boldsymbol{U}\boldsymbol{V}^*\right)_{\alpha,\beta} \right|^2 \leq \frac{\mu_2 r}{n_1^2 n_2^2}; \tag{3}$$

$$\forall (k,l) \in [n_1]\times[n_2] : \sum_{(\alpha,\beta)\in[n_1]\times[n_2]} \omega_{\alpha,\beta} \left| \left\langle \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{(k,l)} \boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{(\alpha,\beta)} \right\rangle \right|^2 \leq \frac{\mu_3 r}{n_1 n_2}\omega_{k,l}. \tag{4}$$

Theorem 1. *Let $\boldsymbol{X}$ be a $n_1 \times n_2$ data matrix, and Ω the random location set of size m. Define $c_s := \max\left(\frac{n_1 n_2}{k_1 k_2}, \frac{n_1 n_2}{(n_1-k_1+1)(n_2-k_2+1)}\right)$. If all measurements are noiseless, then there exists a constant $c_1 > 0$ such that under either of the following conditions:*

i) Condition (2), (3) and (4) hold and

$$m > c_1 \max\left(\mu_1 c_s, \mu_3 c_s, \mu_2\right) r \log^2\left(n_1 n_2\right); \tag{5}$$

ii) Condition (2) holds and

$$m > c_1 \mu_1^2 c_s^2 r^2 \log^2(n_1 n_2); \tag{6}$$

then $\boldsymbol{X}$ is the unique solution of EMaC with probability exceeding $1 - \frac{1}{n_1^2 n_2^2}$.

The performance in the presence of noise is states as follows.

Theorem 2. *Consider a 2-fold Hankel matrix $\boldsymbol{X}_\mathrm{e}$ of rank r, and suppose that the total power of the noise is δ. Let $\hat{\boldsymbol{X}}$ be the optimizer of EMaC-Noisy. Under the conditions of Theorem 1, one can bound*

$$\left\|\boldsymbol{X}_\mathrm{e} - \hat{\boldsymbol{X}}_\mathrm{e}\right\|_\mathrm{F} \leq \left\{2\sqrt{n_1 n_2} + 8 n_1 n_2 + \frac{8\sqrt{2} n_1^2 n_2^2}{m}\right\}\delta$$

with probability exceeding $1 - \frac{1}{n_1^2 n_2^2}$.

Their counterpart for the Hankel matrix completion problem is stated in the following theorem.

Theorem 3. *Consider a 2-fold Hankel matrix $\boldsymbol{X}_e$ of rank r. The bounds in Theorem 1 and 2 continue to hold, if the incoherence μ_1 is measured as the smallest number that satisfies*

$$\forall\,(i,l) \in [n_1]\times[n_2], \quad \left\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{(i,l)}\right\|_F^2 \leq \frac{\mu_1 c_s r}{n_1 n_2}, \text{ and } \left\|\boldsymbol{A}_{(i,l)}\boldsymbol{V}\boldsymbol{V}^*\right\|_F^2 \leq \frac{\mu_1 c_s r}{n_1 n_2}. \tag{7}$$

The proof in the noiseless setting (i.e. Theorem 1 and the noiseless part of Theorem 3) is provided in Section 3. The analyses of the noisy counterparts (i.e. Theorem 2 and the noisy part of Theorem 3) are built upon the noiseless situation, which is deferred to Appendix G.

3 Main Proof for Exact Recovery

The algorithm EMaC has similar spirit as the well-known matrix completion algorithms [1, 2] except that we impose Hankel and block-Hankel structures on the matrices. While [2] has derived a general sufficient condition for exact recovery under any basis (see [2, Theorem 3]), the basis in our case does not exhibit a good coherence property required in [2], and hence these results cannot yield useful estimates in our framework. Nevertheless, the beautiful golfing scheme introduced in [2] lays the foundation of our analysis in the sequel.

For concreteness, the analysis in this paper focuses on recovering harmonically sparse signals as stated in Theorem 1, since proving Theorem 1 is slightly more involved than proving Theorem 3. We note, however, that our analysis already entails all reasoning required for Theorem 3.

Before proceeding to the proof, we would first like to stress that the incoherence measure (μ_1, μ_2, μ_3) are not independent. In addition to them, we define another measure μ_4 as the smallest number that satisfies

$$\forall \boldsymbol{b} \in [n_1] \times [n_2] : \qquad \sum_{\boldsymbol{a} \in [n_1] \times [n_2]} \omega_{\boldsymbol{a}} \left| \langle \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}} \rangle \right|^2 \le \frac{\mu_4 r}{n_1 n_2} \omega_{\boldsymbol{b}}, \tag{8}$$

Some of their mutual connections are listed as follows.

Lemma 1. *Suppose that $\boldsymbol{X}_{\mathrm{e}}$ has incoherence $(\mu_1, \mu_2, \mu_3, \mu_4)$. We have the following.*

1. $\boldsymbol{G}_{\mathrm{L}} = \boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}$, *and* $\boldsymbol{G}_{\mathrm{R}} = \left(\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*\right)^T$;
2. *For any $\boldsymbol{a}, \boldsymbol{b} \in [n_1] \times [n_2]$, one has*

$$\left| \langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}} \rangle \right| \le \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2}; \tag{9}$$

3. *The incoherence measure satisfies*

$$\mu_2 \le \mu_1^2 c_{\mathrm{s}}^2 r, \quad \mu_3 \le \mu_1^2 c_s^2 r, \tag{10}$$

and

$$\mu_4 \le 9 \mu_1^2 c_{\mathrm{s}}^2 r; \tag{11}$$

4. *The measure μ_4 can be bounded by μ_1 and μ_3 as follows*

$$\mu_4 \le 6 \mu_1 c_{\mathrm{s}} + 3 \mu_3 c_{\mathrm{s}}.$$

Proof. See Appendix A. □

Note that the above lemma indicates that our new incoherence measure μ_4 can be bounded by the sum of μ_1 and μ_3 up to some multiplicative constant. In fact, we will prove instead the following theorem based on (μ_1, μ_2, μ_4), which is slightly more general than Theorem 1.

Theorem 4. *Suppose that $\boldsymbol{X}$ has incoherence measure $(\mu_1, \mu_2, \mu_3, \mu_4)$. If*

$$m > c_0 \max\left(\mu_1 c_s, \mu_2, \mu_4\right) r \log^2\left(n_1 n_2\right), \tag{12}$$

then $\boldsymbol{X}$ is the unique solution of EMaC with probability exceeding $1 - \frac{1}{n_1^2 n_2^2}$

Note that Theorem 1 can be delivered as an immediate consequence of Theorem 4 by exploiting the relations among $(\mu_1, \mu_2, \mu_3, \mu_4)$ given in Lemma 1.

3.1 Dual Certification

Denote by $\mathcal{A}_{(k,l)}(\boldsymbol{M})$ the projection of $\boldsymbol{M}$ onto the subspace spanned by $\boldsymbol{A}_{(k,l)}$, and define the projection operator onto the space spanned by all $\boldsymbol{A}_{(k,l)}$ and its orthogonal complement as

$$\mathcal{A} := \sum_{(k,l) \in [n_1] \times [n_2]} \mathcal{A}_{(k,l)}, \quad \text{and} \quad \mathcal{A}^{\perp} = \mathcal{I} - \mathcal{A}. \tag{13}$$

Here, $\left\{\mathcal{A}^{\perp}(\boldsymbol{M})\right\}$ spans a $\left[k_1 k_2 \left(n_1 - k_1 + 1\right)\left(n_2 - k_2 + 1\right) - n_1 n_2\right]$ dimensional subspace.

There are two common ways to describe the randomness of Ω: one corresponds to sampling *without* replacement, and another concerns sampling *with* replacement (i.e. Ω contains m indices $\{\boldsymbol{a}_i \in [n_1] \times [n_2] : 1 \le i \le m\}$ that are i.i.d. generated). As discussed in [2, Section II.A], while both situations result in the same order-wide bounds, the latter situation admits simpler analysis due to independence. Therefore, we will assume that Ω is a multiset (possibly with repeated elements) and a_i's are independently and uniformly distributed throughout the proofs of this paper, and define the associated operators as

$$\mathcal{A}_{\Omega} := \sum_{i=1}^{m} \mathcal{A}_{\boldsymbol{a}_i}. \tag{14}$$

We also define another projection operator $\mathcal{A}'_{\Omega}$ similar to (14), but with the sum extending only over *distinct* samples. Its complement operator is defined as $\mathcal{A}'_{\Omega^{\perp}} := \mathcal{A} - \mathcal{A}'_{\Omega}$. Note that $\mathcal{A}_{\Omega}(\boldsymbol{M}) = 0$ is equivalent to $\mathcal{A}'_{\Omega}(\boldsymbol{M}) = 0$.

With these definitions, EMaC can be rewritten as the following general matrix completion problem:

$$\begin{aligned} \underset{\boldsymbol{M}}{\text{minimize}} \quad & \|\boldsymbol{M}\|_* \\ \text{subject to} \quad & \mathcal{A}'_{\Omega}(\boldsymbol{M}) = \mathcal{A}'_{\Omega}(\boldsymbol{X}_{\mathrm{e}}), \\ & \mathcal{A}^{\perp}(\boldsymbol{M}) = \mathcal{A}^{\perp}(\boldsymbol{X}_{\mathrm{e}}) = 0. \end{aligned} \tag{15}$$

To prove exact recovery of convex optimization, it suffices to produce an appropriate dual certificate, as stated in the following lemma.

Lemma 2. *For a location set Ω that contains m random indices. Suppose that the sampling operator $\mathcal{A}_{\Omega}$ obeys*

$$\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_{\Omega}\mathcal{P}_T\right\| \leq \frac{1}{2}. \tag{16}$$

If there exists a matrix $\boldsymbol{W}$ that obeys

$$\mathcal{A}'_{\Omega^{\perp}}(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W}) = 0, \tag{17}$$

$$\|\mathcal{P}_T(\boldsymbol{W})\|_{\mathrm{F}} \leq \frac{1}{2n_1^2 n_2^2}, \tag{18}$$

and

$$\|\mathcal{P}_{T^{\perp}}(\boldsymbol{W})\| \leq \frac{1}{2}. \tag{19}$$

Then $\boldsymbol{X}_{\mathrm{e}}$ is the unique optimizer of (15) or, equivalently, $\boldsymbol{X}$ is the unique minimizer of EMaC.

Proof. See Appendix B. □

Condition (16) will be analyzed in Section 3.2, while a valid certificate $\boldsymbol{W}$ will be constructed in Section 3.3. These are the objectives of the remaining part of the section.

3.2 Deviation of $\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_{\Omega}\mathcal{P}_T\right\|$

Lemma 2 requires that $\mathcal{A}_{\Omega}$ is sufficiently incoherent with respect to T. The following lemma quantifies the projection of each $\boldsymbol{A}_{(k,l)}$ onto the tangent space T.

Lemma 3. *Suppose that* (2) *holds, then*

$$\left\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2 \leq \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}, \quad \left\|\boldsymbol{A}_{(k,l)}\boldsymbol{V}\boldsymbol{V}^*\right\|_{\mathrm{F}}^2 \leq \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}, \quad \left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2 \leq \frac{2\mu_1 c_{\mathrm{s}} r}{n_1 n_2} \tag{20}$$

for all $(k,l) \in [n_1] \times [n_2]$.

Proof. See Appendix C. □

As long as (20) holds, the deviation of $\mathcal{P}_T\mathcal{A}_{\Omega}\mathcal{P}_T$ can be bounded reasonably well in the following lemma. This establishes Condition (16) required by Lemma 2.

Lemma 4. *Suppose that*

$$\left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2 \leq \frac{2\mu_1 c_{\mathrm{s}} r}{n_1 n_2},$$

for $(k,l) \in [n_1] \times [n_2]$. *Then for any small constant* $\delta \leq 2$, *one has*

$$\left\|\frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \leq \delta \tag{21}$$

with probability exceeding $1 - 2n_1 n_2 \exp\left(-\frac{\delta^2 m}{16\mu_1 c_{\mathrm{s}} r}\right)$.

Proof. See Appendix D. □

The above two lemmas taken collectively lead to the following fact: for any given constant $\epsilon < e^{-1} < \frac{1}{2}$, $\left\|\frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_\Omega\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \leq \epsilon$ holds with probability exceeding $1 - (n_1 n_2)^{-4}$, provided that $m > c_1\mu_1 c_{\mathrm{s}} r \log(n_1 n_2)$ for some constant $c_1 > 0$.

3.3 Construction of Dual Certificate

Now we are in a position to construct the dual certificate, for which we will employ the golfing scheme introduced in [2]. Suppose that we generate j_0 independent random location multisets Ω_i ($1 \leq i \leq j_0$), each containing $\frac{m}{j_0}$ i.i.d. samples. This way the distribution of Ω is the same as $\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_{j_0}$. Note that Ω_i's correspond to sampling *with* replacement. Let $\rho := \frac{m}{n_1 n_2}$ and $q := \frac{\rho}{j_0}$ denote the undersampling factors of Ω and Ω_i, respectively.

Consider a small constant $\epsilon < \frac{1}{e}$, and choose $j_0 := 3\log_{\frac{1}{\epsilon}} n_1 n_2$. The construction of the dual then proceeds as follows:

Construction of a dual certificate $\boldsymbol{W}$ via the golfing scheme.
1. Set $\boldsymbol{B}_0 = 0$, and $j_0 := 3\log_{\frac{1}{\epsilon}}(n_1 n_2)$.
2. For all i ($1 \leq i \leq j_0$), let $\boldsymbol{B}_i = \boldsymbol{B}_{i-1} + \left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\left(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{i-1}\right)$.
3. Set $\boldsymbol{W} := -\left(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{j_0}\right)$.

We will establish that $\boldsymbol{W}$ is a valid dual certificate if we can show that $\boldsymbol{W}$ satisfies the conditions stated in Lemma 2, which we will verify step by step.

First, by construction, we have the identities

$$\left(\mathcal{A}'_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{B}_i) = \boldsymbol{B}_i,$$

for all $1 \leq i \leq j_0$. Since $\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W} = \boldsymbol{B}_{j_0}$, this validates that $\mathcal{A}'_{\Omega^\perp}(\boldsymbol{U}\boldsymbol{V}^* + \boldsymbol{W}) = 0$, as required in (17).

Secondly, if one defines the deviation of $\mathcal{P}_T\boldsymbol{B}_i$ from $\boldsymbol{U}\boldsymbol{V}^*$ as

$$\boldsymbol{F}_i := \boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_i,$$

and hence $\boldsymbol{W} = \boldsymbol{F}_{j_0}$, then one can verify that

$$\begin{aligned}\mathcal{P}_T(\boldsymbol{F}_i) &= \mathcal{P}_T(\boldsymbol{U}\boldsymbol{V}^*) - \mathcal{P}_T\left(\boldsymbol{B}_{i-1} + \left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\left(\boldsymbol{U}\boldsymbol{V}^* - \boldsymbol{B}_{i-1}\right)\right) \\ &= \left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)(\boldsymbol{F}_{i-1}).\end{aligned}$$

Lemma 4 asserts the following: if $qn_1 n_2 \geq c_1\mu_1 c_{\mathrm{s}} r \log(n_1 n_2)$ or, equivalently, $m \geq \tilde{c}_1\mu_1 c_{\mathrm{s}} r \log^2(n_1 n_2)$, then with overwhelming probability one has

$$\left\|\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right\| = \left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_i}\mathcal{P}_T\right\| \leq \epsilon < \frac{1}{2}.$$

This allows us to bound $\left\|\mathcal{P}_T\left(\boldsymbol{F}_i\right)\right\|_{\mathrm{F}}$ as follows

$$\left\|\mathcal{P}_T\left(\boldsymbol{F}_i\right)\right\|_{\mathrm{F}} \leq \epsilon^i\left\|\mathcal{P}_T\left(\boldsymbol{F}_0\right)\right\|_{\mathrm{F}} \leq \epsilon^i\left\|\boldsymbol{U}\boldsymbol{V}^*\right\|_{\mathrm{F}} = \epsilon^i\sqrt{r},$$

which immediately validates Condition (18):

$$\left\|\mathcal{P}_T\left(\boldsymbol{W}\right)\right\|_{\mathrm{F}} = \left\|\mathcal{P}_T\left(\boldsymbol{F}_{j_0}\right)\right\|_{\mathrm{F}} \leq \epsilon^{j_0}\sqrt{r} < \frac{1}{2n_1^2n_2^2}.$$

Finally, it remains to show that $\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\| \leq \frac{1}{2}$. For any $\boldsymbol{F} \in T$, define the following homogeneity measure

$$\nu\left(\boldsymbol{F}\right) = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{F}\right\rangle\right|^2, \tag{22}$$

which largely relies on the average per-entry energy in each skew diagonal. We would like to show that $\nu\left(\left(\mathcal{I} - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\right)\boldsymbol{F}\right) \leq \frac{1}{4}\nu\left(\boldsymbol{F}\right)$ with high probability. This is supplied in the following lemma.

Lemma 5. *Consider any given $\boldsymbol{F} \in T$, and suppose that* (2) *and* (8) *hold. If the following bound holds,*

$$m > c_7 \max\left\{\mu_4, \mu_1 c_{\mathrm{s}}\right\} r \log^2\left(n_1n_2\right),$$

then one has

$$\nu\left(\left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)\boldsymbol{F}\right) \leq \frac{1}{4}\nu\left(\boldsymbol{F}\right) \tag{23}$$

for all $1 \leq i \leq j_0$ with probability exceeding $1 - \left(n_1n_2\right)^{-3}$.

Proof. See Appendix E. □

This lemma basically indicates that a homogeneous $\boldsymbol{F}$ with respect to the observation basis typically results in a homogeneous $\left(\mathcal{P}_T - \mathcal{P}_T\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\mathcal{P}_T\right)\left(\boldsymbol{F}\right)$, and hence we can hope that the homogeneity condition (3) of $\boldsymbol{F}_0$ can carry over to every $\mathcal{P}_T\left(\boldsymbol{F}_i\right)$ $(1 \leq i \leq j_0)$.

Observe that Condition (3) is equivalent to saying

$$\nu\left(\boldsymbol{F}_0\right) = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}}\left|\left\langle \boldsymbol{A}_{(k,l)}, \boldsymbol{U}\boldsymbol{V}^*\right\rangle\right|^2 = \max_{(k,l)\in[n_1]\times[n_2]} \frac{1}{\omega_{k,l}^2}\left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(k,l)}\left(\boldsymbol{U}\boldsymbol{V}^*\right)_{\alpha,\beta}\right|^2 \leq \frac{\mu_2 r}{\left(n_1n_2\right)^2}.$$

One can then verify that for every i $(0 \leq i \leq j_0)$,

$$\nu\left(\mathcal{P}_T\left(\boldsymbol{F}_i\right)\right) \leq \frac{1}{4}\nu\left(\mathcal{P}_T\left(\boldsymbol{F}_{i-1}\right)\right) \leq \left(\frac{1}{4}\right)^i\nu\left(\boldsymbol{F}_0\right) \leq \left(\frac{1}{4}\right)^i\frac{\mu_2 r}{\left(n_1n_2\right)^2}$$

holds with high probability if $m > c_7\max\left\{\mu_4, \mu_1 c_{\mathrm{s}}\right\} r\log^2\left(n_1n_2\right)$ for some constant $c_7 > 0$.

The following lemma then relates the homogeneity measure with $\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\left(\boldsymbol{F}_i\right)\right\|$.

Lemma 6. *For any given $\boldsymbol{F} \in T$ such that $\nu(\boldsymbol{F})$. Then there exist positive constants c_8 and c_9 such that for any $t \leq \sqrt{\nu(\boldsymbol{F})}n_1n_2$,*

$$\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_i} + \mathcal{A}^\perp\right)\left(\boldsymbol{F}\right)\right\| > t$$

holds with probability at most $c_8\exp\left(-\frac{c_9 q t^2}{\nu(\boldsymbol{F})n_1n_2}\right)$.

Proof. See Appendix F. □

Since $\nu(\boldsymbol{F}_i) \leq \left(\frac{1}{4}\right)^i \frac{\mu_2 r}{n_1^2 n_2^2}$ for all $1 \leq i \leq j_0$ with high probability, then one can bound

$$\frac{\sqrt{\nu(\boldsymbol{F}_i)} n_1 n_2}{\sqrt{16\mu_2 r}} \leq \left(\frac{1}{2}\right)^{i+2}.$$

Lemma 6 immediately yields that for all i $(0 \leq i \leq j_0)$

$$\begin{aligned}\mathbb{P}\left\{\forall i: \left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{F}_i)\right\| \leq \left(\frac{1}{2}\right)^{i+2}\right\} &\geq \mathbb{P}\left\{\forall i: \left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{F}_i)\right\| \leq \frac{\sqrt{\nu(\boldsymbol{F}_i)} n_1 n_2}{\sqrt{16\mu_2 r}}\right\} \\ &\geq 1 - c_8 n_1 n_2 \exp\left(-\frac{c_9 q n_1 n_2}{16\mu_2 r}\right) \\ &\geq 1 - c_8 \left(n_1 n_2\right)^{-4},\end{aligned}$$

holds if $q n_1 n_2 > c_{12} \max\left(\mu_1 c_{\mathrm{s}}, \mu_4, \mu_2\right) r \log\left(n_1 n_2\right)$ for some constant $c_{12} > 0$. This is also equivalent to

$$m > c_{13} \max\left(\mu_1 c_{\mathrm{s}}, \mu_4, \mu_2\right) r \log^2\left(n_1 n_2\right)$$

for some constant $c_{13} > 0$. Under this condition, we can conclude

$$\begin{aligned}\left\|\mathcal{P}_{T^\perp}\left(\boldsymbol{W}\right)\right\| &\leq \sum_{i=0}^{j_0}\left\|\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{F}_i)\right\| \\ &\leq \sum_{i=0}^{j_0}\left(\frac{1}{2}\right)^{i+2} < \frac{1}{2}.\end{aligned}$$

So far, we have successfully established that with high probability, $\boldsymbol{W}$ is a valid dual certificate, and hence EMaC admits perfect reconstruction of $\boldsymbol{X}$.

A Proof of Lemma 1

(1) We first show that $\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}$ and $\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*$ coincide with the matrices $\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{G}_R^T$. Since $\boldsymbol{Y}_{\mathrm{d}}$ is a diagonal matrix, one can verify the identities

$$\left(\boldsymbol{Y}_{\mathrm{d}}^{l*} \boldsymbol{Z}_{\mathrm{L}}^* \boldsymbol{Z}_{\mathrm{L}} \boldsymbol{Y}_{\mathrm{d}}^l\right)_{i_1,i_2} = \left(y_{i_1}^* y_{i_2}\right)^l \left(\boldsymbol{Z}_{\mathrm{L}}^* \boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2},$$

and

$$\left(\boldsymbol{Z}_{\mathrm{L}}^* \boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2} = \sum_{k=0}^{k_2-1}\left(z_{i_1}^* z_{i_2}\right)^k = \begin{cases}\frac{1-\left(z_{i_1}^* z_{i_2}\right)^{k_2}}{1-z_{i_1}^* z_{i_2}}, & \text{if } i_1 \neq i_2, \\ k_2, & \text{if } i_1 = i_2,\end{cases}$$

which immediately give

$$\begin{aligned}\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}} &= \frac{1}{k_1 k_2}\left[\boldsymbol{Z}_{\mathrm{L}}^*, \boldsymbol{Y}_{\mathrm{d}}^* \boldsymbol{Z}_{\mathrm{L}}^*, \cdots, \left(\boldsymbol{Y}_{\mathrm{d}}^*\right)^{k_1-1} \boldsymbol{Z}_{\mathrm{L}}^*\right]\begin{bmatrix}\boldsymbol{Z}_L \\ \boldsymbol{Z}_L \boldsymbol{Y}_{\mathrm{d}} \\ \vdots \\ \boldsymbol{Z}_L \boldsymbol{Y}_{\mathrm{d}}^{k_1-1}\end{bmatrix} = \frac{1}{k_1 k_2}\sum_{l=0}^{k_1-1} \boldsymbol{Y}_{\mathrm{d}}^{l*} \boldsymbol{Z}_{\mathrm{L}}^* \boldsymbol{Z}_{\mathrm{L}} \boldsymbol{Y}_{\mathrm{d}}^l \\ &= \frac{1}{k_1 k_2}\left(\left(\sum_{l=0}^{k_1-1}\left(y_{i_1}^* y_{i_2}\right)^l\right)\left(\boldsymbol{Z}_{\mathrm{L}}^* \boldsymbol{Z}_{\mathrm{L}}\right)_{i_1,i_2}\right)_{1 \leq i_1, i_2 \leq r} \\ &= \frac{1}{k_1 k_2}\left(\frac{1-\left(y_{i_1}^* y_{i_2}\right)^{k_1}}{1-y_{i_1}^* y_{i_2}} \frac{1-\left(z_{i_1}^* z_{i_2}\right)^{k_2}}{1-z_{i_1}^* z_{i_2}}\right)_{1 \leq i_1, i_2 \leq r}\end{aligned}$$

with the convention that $\frac{1-\left(y_{i_1}^* y_{i_1}\right)^{k_1}}{1-y_{i_1}^* y_{i_1}} = k_1$ and $\frac{1-\left(z_{i_1}^* z_{i_1}\right)^{k_2}}{1-z_{i_1}^* z_{i_1}} = k_2$. That said, all diagonal entries satisfy $\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)_{i_1,i_1} = 1$, and the magnitude of off-diagonal entries can be calculated as

$$\left|\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)_{i_1,i_2}\right| = \left|\frac{\sin\left[\pi k_1\left(f_{1i_1} - f_{1i_2}\right)\right]}{k_1 \sin\left[\pi\left(f_{1i_1} - f_{1i_2}\right)\right]} \frac{\sin\left[\pi k_2\left(f_{2i_1} - f_{2i_2}\right)\right]}{k_2 \sin\left[\pi\left(f_{2i_1} - f_{2i_2}\right)\right]}\right|.$$

Recall that this exactly coincides with the definition of $\boldsymbol{G}_{\mathrm{L}}$. Similarly, $\boldsymbol{G}_{\mathrm{R}} = \left(\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*\right)^T$. These findings immediately yield

$$\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right) \geq \frac{1}{\mu_1}, \quad \text{and} \quad \sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*\right) \geq \frac{1}{\mu_1}. \tag{24}$$

(2) Consider the case in which we only know $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{L}}\right) \geq \frac{1}{\mu_1}$ and $\sigma_{\min}\left(\boldsymbol{G}_{\mathrm{R}}\right) \geq \frac{1}{\mu_1}$. In fact, since $\left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}}\rangle\right| = \left|\langle \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\rangle\right|$, we only need to examine the situation where $\omega_{\boldsymbol{b}} < \omega_{\boldsymbol{a}}$.

Observe that

$$\left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}}\rangle\right| \leq \left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}}\rangle\right| + \left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{V}\boldsymbol{V}^*\rangle\right| + \left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{V}\boldsymbol{V}^*\rangle\right|.$$

Owing to the multi-fold Hankel structure of $\boldsymbol{A}_{\boldsymbol{a}}$, the matrix $\boldsymbol{U}\boldsymbol{U}^* \sqrt{\omega_{\boldsymbol{a}}} \boldsymbol{A}_{\boldsymbol{a}}$ consists of $\omega_{\boldsymbol{a}}$ columns of $\boldsymbol{U}\boldsymbol{U}^*$. Since there are only $\omega_{\boldsymbol{b}}$ nonzero entries in $\boldsymbol{A}_{\boldsymbol{b}}$ each of magnitude $\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}$, we can derive

$$\begin{aligned}
\left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}}\rangle\right| &\leq \left\|\boldsymbol{A}_{\boldsymbol{b}}\right\|_1 \left\|\boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}}\right\|_\infty = \omega_{\boldsymbol{b}} \cdot \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}} \cdot \max_{\alpha,\beta} \left|\left(\boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}}\right)_{\alpha,\beta}\right| \\
&\leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \max_{\alpha,\beta} \left|\left(\boldsymbol{U}\boldsymbol{U}^*\right)_{\alpha,\beta}\right|.
\end{aligned}$$

Denote by $\boldsymbol{M}_{*k}$ and $\boldsymbol{M}_{k*}$ the kth column and kth row of $\boldsymbol{M}$, respectively, then it can be observed that each entry of $\boldsymbol{U}\boldsymbol{U}^*$ is bounded in magnitude by

$$\begin{aligned}
\left|\left(\boldsymbol{U}\boldsymbol{U}^*\right)_{k,l}\right| &= \left|\left(\boldsymbol{E}_{\mathrm{L}} \left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)^{-1} \boldsymbol{E}_{\mathrm{L}}^*\right)_{k,l}\right| = \left|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*} \left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)^{-1} \left(\left(\boldsymbol{E}_{\mathrm{L}}\right)_{l*}\right)^*\right| \\
&\leq \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\right\|_{\mathrm{F}} \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{l*}\right\|_{\mathrm{F}} \left\|\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)^{-1}\right\| \\
&\leq \frac{r}{k_1 k_2} \frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)} \leq \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2},
\end{aligned} \tag{25}$$

which immediately implies that

$$\left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}}\rangle\right| \leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}. \tag{26}$$

Similarly, one can derive

$$\left|\langle \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{V}\boldsymbol{V}^*\rangle\right| \leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}. \tag{27}$$

We still need to bound the magnitude of $\langle \boldsymbol{U}\boldsymbol{U}^* \boldsymbol{A}_{\boldsymbol{a}} \boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle$. One can observe that for any $1 \leq k \leq k_1 k_2$:

$$\begin{aligned}
\left\|\left(\boldsymbol{U}\boldsymbol{U}^*\right)_{k*}\right\|_{\mathrm{F}} &\leq \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*} \left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)^{-1} \boldsymbol{E}_{\mathrm{L}}^*\right\|_{\mathrm{F}} \\
&\leq \left\|\left(\boldsymbol{E}_{\mathrm{L}}\right)_{k*}\right\|_{\mathrm{F}} \left\|\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)^{-1} \boldsymbol{E}_{\mathrm{L}}^*\right\| \leq \sqrt{\frac{r}{k_1 k_2}} \cdot \frac{1}{\sqrt{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)}} \\
&\leq \sqrt{\frac{c_{\mathrm{s}} r}{n_1 n_2 \sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}}\right)}}.
\end{aligned}$$

Similarly, for any $1 \leq l \leq (n_1 - k_1 + 1)(n_2 - k_2 + 1)$, one has $\|(\boldsymbol{V}\boldsymbol{V}^*)_{*l}\|_{\mathrm{F}} \leq \sqrt{\frac{c_{\mathrm{s}} r}{n_1 n_2 \sigma_{\min}(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}})}}$. The magnitude of all entries of $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*$ can now be bounded by

$$\begin{aligned}\max_{k,l} \left|(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*)_{k,l}\right| &\leq \|\boldsymbol{A}_{\boldsymbol{a}}\| \max_k \|(\boldsymbol{U}\boldsymbol{U}^*)_{k*}\|_{\mathrm{F}} \max_l \|(\boldsymbol{V}\boldsymbol{V}^*)_{*l}\|_{\mathrm{F}} \\ &\leq \frac{1}{\sqrt{\omega_{\boldsymbol{a}}}} \frac{c_{\mathrm{s}} r}{n_1 n_2 \sigma_{\min}(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}})} \\ &\leq \frac{1}{\sqrt{\omega_{\boldsymbol{a}}}} \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}.\end{aligned}$$

Since $\boldsymbol{A}_{\boldsymbol{b}}$ has only $\omega_{\boldsymbol{b}}$ nonzero entries each has magnitude $\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}$, one can verify that

$$|\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle| \leq \left(\max_{k,l} \left|(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*)_{k,l}\right|\right) \cdot \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}} \omega_{\boldsymbol{b}} = \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}. \tag{28}$$

The above bounds (26), (27) and (28) taken together lead to

$$\begin{aligned}|\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}}\rangle| &\leq |\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{A}_{\boldsymbol{b}}\rangle| + |\langle \boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle| + |\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle| \\ &\leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2}. \end{aligned} \tag{29}$$

(3) On the other hand, the bound on $|\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{a}}\rangle|$ immediately leads the following upper bounds on $\sum_{\boldsymbol{a}} |\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle|^2 \omega_{\boldsymbol{a}}$ and $\sum_{\boldsymbol{a}} |\langle \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\rangle|^2 \omega_{\boldsymbol{a}}$:

$$\begin{aligned}&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} |\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{a}}\boldsymbol{V}\boldsymbol{V}^*, \boldsymbol{A}_{\boldsymbol{b}}\rangle|^2 \omega_{\boldsymbol{a}} \\ \leq &\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \left(\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}\right)^2 \omega_{\boldsymbol{a}} = \omega_{\boldsymbol{b}} \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \left(\frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}\right)^2 \\ = &\omega_{\boldsymbol{b}} \frac{\mu_1^2 c_{\mathrm{s}}^2 r^2}{n_1 n_2}\end{aligned}$$

which simply come from the inequality (28), and

$$\begin{aligned}&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} |\langle \mathcal{P}_T \boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\rangle|^2 \omega_{\boldsymbol{a}} \\ \leq &\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \left(\sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}} \frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2}\right)^2 \omega_{\boldsymbol{a}} = \omega_{\boldsymbol{b}} \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \left(\frac{3\mu_1 c_{\mathrm{s}} r}{n_1 n_2}\right)^2 \\ = &\omega_{\boldsymbol{b}} \frac{9\mu_1^2 c_{\mathrm{s}}^2 r^2}{n_1 n_2},\end{aligned}$$

which is an immediate consequence of (29). These bounds indicate that $\mu_3 \leq \mu_1^2 c_{\mathrm{s}}^2 r$ and $\mu_4 \leq 9\mu_1^2 c_{\mathrm{s}}^2 r$.

We can also obtain an upper bound on μ_2 through μ_1 as follows. Observe that there exists a unitary matrix $\boldsymbol{B}$ such that

$$\boldsymbol{U}\boldsymbol{V}^* = \boldsymbol{E}_{\mathrm{L}} (\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}})^{-\frac{1}{2}} \boldsymbol{B} (\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*)^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{R}}.$$

For any $(k,l) \in [n_1] \times [n_2]$, we can then bound

$$\begin{aligned}\left|(\boldsymbol{U}\boldsymbol{V}^*)_{k,l}\right| &= \left|\left(\boldsymbol{E}_{\mathrm{L}} (\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}})^{-\frac{1}{2}} \boldsymbol{B} (\boldsymbol{E}_{\mathrm{R}} \boldsymbol{E}_{\mathrm{R}}^*)^{-\frac{1}{2}} \boldsymbol{E}_{\mathrm{R}}\right)_{k,l}\right| \\ &\leq \|(\boldsymbol{E}_{\mathrm{L}})_{k*}\|_{\mathrm{F}} \left\|(\boldsymbol{E}_{\mathrm{L}}^* \boldsymbol{E}_{\mathrm{L}})^{-\frac{1}{2}}\right\| \|\boldsymbol{B}\| \left\|(\boldsymbol{E}_{\mathrm{R}}^* \boldsymbol{E}_{\mathrm{R}})^{-\frac{1}{2}}\right\| \|(\boldsymbol{E}_{\mathrm{R}})_{*l}\|_{\mathrm{F}} \\ &\leq \sqrt{\frac{r}{k_1 k_2}} \mu_1 \sqrt{\frac{r}{(n_1 - k_1 + 1)(n_2 - k_2 + 1)}} \\ &\leq \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}.\end{aligned}$$

Since $\boldsymbol{A}_{(k,l)}$ has only $\omega_{k,l}$ nonzero entries each of magnitude $\frac{1}{\sqrt{\omega_{k,l}}}$, this leads to

$$\begin{aligned}\frac{1}{\omega_{k,l}^2}\left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(k,l)}\left(\boldsymbol{U}\boldsymbol{V}^*\right)_{\alpha,\beta}\right|^2 &= \frac{1}{\omega_{k,l}}\left|\left\langle \boldsymbol{U}\boldsymbol{V}^*, \boldsymbol{A}_{(k,l)}\right\rangle\right|^2 \\ &\leq \frac{1}{\omega_{k,l}}\left\{\left(\max_{k,l}\left|\left(\boldsymbol{U}\boldsymbol{V}^*\right)_{k,l}\right|\right)\frac{1}{\sqrt{\omega_{k,l}}}\cdot\omega_{k,l}\right\}^2 \\ &\leq \left(\max_{k,l}\left|\left(\boldsymbol{U}\boldsymbol{V}^*\right)_{k,l}\right|\right)^2 \leq \mu_1^2 c_{\mathrm{s}}^2 r\frac{r}{n_1^2 n_2^2},\end{aligned}$$

which indicates that $\mu_2 \leq \mu_1^2 c_{\mathrm{s}}^2 r$.

(4) Finally, we split $\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2$ as follows

$$\begin{aligned}&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle\left(\mathcal{P}_U+\mathcal{P}_V-\mathcal{P}_U\mathcal{P}_V\right)\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 \\ &\leq 3\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left\{\left|\langle \mathcal{P}_U\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2+\left|\langle \mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2+\left|\langle \mathcal{P}_U\mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\right\}\end{aligned}$$

Now look at $\sum_{\boldsymbol{a}}\left|\left\langle \mathcal{P}_U\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2 = \sum_{\boldsymbol{a}}\left|\left\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2$. We know that

$$\left\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\right\|_{\mathrm{F}}^2 \leq \frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}, \tag{30}$$

and that $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$ has $\omega_{\boldsymbol{b}}$ non-zero columns, or,

$$\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}} \overset{\text{column permutation}}{=} \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left[\underbrace{\overline{\boldsymbol{U}}_{\boldsymbol{b}}}_{\omega_{\boldsymbol{b}}\text{ columns}}, \boldsymbol{0}\right], \tag{31}$$

and hence $\left\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\right\rangle$ is simply the sum of all entries of $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$ lying in the set $\Omega_{\mathrm{e}}(\boldsymbol{a})$. Since there are at most $\omega_{\boldsymbol{b}}$ nonzero entries (due to the above structure of $\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}$) in each sum, we can bound

$$\left|\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 = \left|\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}\left(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\right)_{\alpha,\beta}\right|^2 \leq \omega_{\boldsymbol{b}}\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}\left|\left(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\right)_{\alpha,\beta}\right|^2$$

using the inequality $\left(\sum_{i=1}^{\omega_{\boldsymbol{b}}} x_i\right)^2 \leq \omega_{\boldsymbol{b}}\sum_{i=1}^{\omega_{\boldsymbol{b}}} x_i^2$. This then gives

$$\begin{aligned}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 &\leq \omega_{\boldsymbol{b}}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\sum_{(\alpha,\beta)\in\Omega_{\mathrm{e}}(\boldsymbol{a})}\left|\left(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_{\boldsymbol{b}}\right)_{\alpha,\beta}\right|^2 \\ &\leq \omega_{\boldsymbol{b}}\left\|\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{A}_b\right\|_{\mathrm{F}}^2 \leq \omega_{\boldsymbol{b}}\frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2},\end{aligned}$$

where the last inequality follows from Lemma 3. Similarly, one has

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \boldsymbol{A}_{\boldsymbol{b}}\boldsymbol{V}\boldsymbol{V}^*, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 \leq \omega_{\boldsymbol{b}}\frac{\mu_1 c_{\mathrm{s}} r}{n_1 n_2}.$$

To summarize,

$$\begin{aligned}&\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2 \\ &\leq 3\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left\{\left|\langle \mathcal{P}_U\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2+\left|\langle \mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2+\left|\langle \mathcal{P}_U\mathcal{P}_V\boldsymbol{A}_{\boldsymbol{b}}, \sqrt{\omega_{\boldsymbol{a}}}\boldsymbol{A}_{\boldsymbol{a}}\rangle\right|^2\right\} \\ &\leq \frac{6\mu_1 c_{\mathrm{s}}\omega_{\boldsymbol{b}} r}{n_1 n_2}+\frac{3\mu_3 c_{\mathrm{s}}\omega_{\boldsymbol{b}} r}{n_1 n_2}.\end{aligned}$$

B Proof of Lemma 2

Consider any valid perturbation $\boldsymbol{H}$ obeying $\mathcal{P}_{\Omega}(\boldsymbol{X}+\boldsymbol{H})=\mathcal{P}_{\Omega}(\boldsymbol{X})$, and denote by $\boldsymbol{H}_{\mathrm{e}}$ the enhanced form of $\boldsymbol{H}$. We note that the constraint requires $\mathcal{A}'_{\Omega}(\boldsymbol{H}_{\mathrm{e}})=0$ (or $\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})=0$) and $\mathcal{A}^{\perp}(\boldsymbol{H}_{\mathrm{e}})=0$. In addition, set $\boldsymbol{W}_0=\mathcal{P}_{T^{\perp}}(\boldsymbol{B})$ for any $\boldsymbol{B}$ that satisfies $\langle\boldsymbol{B},\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\rangle=\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_*$ and $\|\boldsymbol{B}\|\leq 1$. Therefore, $\boldsymbol{W}_0\in T^{\perp}$ and $\|\boldsymbol{W}_0\|\leq 1$, and hence $\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0$ is a subgradient of the nuclear norm at $\boldsymbol{X}_{\mathrm{e}}$. We will establish this lemma by considering two scenarios separately.

(1) Consider first the case in which $\boldsymbol{H}_{\mathrm{e}}$ satisfies

$$\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}\leq\frac{n_1^2n_2^2}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}. \tag{32}$$

Since $\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0$ is a subgradient of the nuclear norm at $\boldsymbol{X}_{\mathrm{e}}$, it follows that

$$\begin{aligned}
\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\|_* &\geq \|\boldsymbol{X}_{\mathrm{e}}\|_*+\langle\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle\\
&=\|\boldsymbol{X}_{\mathrm{e}}\|_*+\langle\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle+\langle\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle\\
&=\|\boldsymbol{X}_{\mathrm{e}}\|_*+\left\langle\left(\mathcal{A}'_{\Omega}+\mathcal{A}^{\perp}\right)(\boldsymbol{U}\boldsymbol{V}^*+\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\right\rangle+\langle\boldsymbol{W}_0,\boldsymbol{H}_{\mathrm{e}}\rangle-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle & (33)\\
&\geq\|\boldsymbol{X}_{\mathrm{e}}\|_*+\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_*-\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle & (34)
\end{aligned}$$

where (33) holds from (17), and (34) follows from the property of $\boldsymbol{W}_0$ and the fact that $\left(\mathcal{A}'_{\Omega}+\mathcal{A}^{\perp}\right)(\boldsymbol{H}_{\mathrm{e}})=0$. The last term of (34) can be bounded as

$$\begin{aligned}
\langle\boldsymbol{W},\boldsymbol{H}_{\mathrm{e}}\rangle &= \langle\mathcal{P}_T(\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\rangle+\langle\mathcal{P}_{T^{\perp}}(\boldsymbol{W}),\boldsymbol{H}_{\mathrm{e}}\rangle\\
&\leq\|\mathcal{P}_T(\boldsymbol{W})\|_{\mathrm{F}}\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}+\|\mathcal{P}_{T^{\perp}}(\boldsymbol{W})\|\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_*\\
&\leq\frac{1}{2n_1^2n_2^2}\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}+\frac{1}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_*,
\end{aligned}$$

where the last inequality follows from the assumptions (18) and (19). Plugging this into (34) yields

$$\begin{aligned}
\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\|_* &\geq\|\boldsymbol{X}_{\mathrm{e}}\|_*-\frac{1}{2n_1^2n_2^2}\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}+\frac{1}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_*\\
&\geq\|\boldsymbol{X}_{\mathrm{e}}\|_*-\frac{1}{4}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}+\frac{1}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}} & (35)\\
&\geq\|\boldsymbol{X}_{\mathrm{e}}\|_*+\frac{1}{4}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}
\end{aligned}$$

where (35) follows from the inequality $\|\boldsymbol{M}\|_*\geq\|\boldsymbol{M}\|_{\mathrm{F}}$ and (32). Therefore, $\boldsymbol{X}_{\mathrm{e}}$ is the minimizer of EMaC.

We still need to prove the uniqueness of the minimizer. The inequality (35) implies that $\|\boldsymbol{X}_{\mathrm{e}}+\boldsymbol{H}_{\mathrm{e}}\|_*=\|\boldsymbol{X}_{\mathrm{e}}\|_*$ only when $\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}=0$. If $\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}=0$, then $\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}\leq\frac{n_1^2n_2^2}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}=0$, and hence $\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})=\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})=0$, which only occurs when $\boldsymbol{H}_{\mathrm{e}}=0$. Hence, $\boldsymbol{X}_{\mathrm{e}}$ is the unique minimizer in this situation.

(2) On the other hand, consider the complement scenario where the following holds

$$\|\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}\geq\frac{n_1^2n_2^2}{2}\|\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\|_{\mathrm{F}}. \tag{36}$$

We would first like to bound $\left\|\left(\frac{n_1n_2}{m}\mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right)\mathcal{P}_T(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}$ and $\left\|\left(\frac{n_1n_2}{m}\mathcal{A}_{\Omega}+\mathcal{A}^{\perp}\right)\mathcal{P}_{T^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}$. The former term can be lower bounded by

$$
\begin{aligned}
& \left\| \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}^2 \\
= & \left\langle \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle \\
= & \left\langle \frac{n_1 n_2}{m} \mathcal{A}_\Omega \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \frac{n_1 n_2}{m} \mathcal{A}_\Omega \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle + \left\langle \mathcal{A}^\perp \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \mathcal{A}^\perp \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle \\
\geq & \left\langle \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \frac{n_1 n_2}{m} \mathcal{A}_\Omega \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle + \left\langle \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \mathcal{A}^\perp \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle \qquad (37) \\
= & \left\langle \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \mathcal{P}_T \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle \\
= & \left\langle \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle + \left\langle \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right), \left(\frac{n_1 n_2}{m} \mathcal{P}_T \mathcal{A}_\Omega \mathcal{P}_T - \mathcal{P}_T \mathcal{A} \mathcal{P}_T \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\rangle \\
\geq & \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}^2 - \left\| \mathcal{P}_T \mathcal{A} \mathcal{P}_T - \frac{n_1 n_2}{m} \mathcal{P}_T \mathcal{A}_\Omega \mathcal{P}_T \right\| \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}^2 \qquad (38) \\
\geq & \left(1 - \left\| \mathcal{P}_T \mathcal{A} \mathcal{P}_T - \frac{n_1 n_2}{m} \mathcal{P}_T \mathcal{A}_\Omega \mathcal{P}_T \right\| \right) \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}^2 \\
\geq & \frac{1}{2} \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}^2 . \qquad (39)
\end{aligned}
$$

On the other hand, since the operator norm of any projection operator is bounded above by 1, one can verify that

$$
\left\| \frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right\| \leq \frac{n_1 n_2}{m} \left(\left\| \mathcal{A}_{a_1} + \mathcal{A}^\perp \right\| + \sum_{i=2}^{m} \left\| \mathcal{A}_{a_i} \right\| \right) \leq n_1 n_2,
$$

where a_i $(1 \leq i \leq m)$ are m uniform random indices that form Ω. This implies the following bound:

$$
\left\| \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_{T^\perp} \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \leq n_1 n_2 \left\| \mathcal{P}_{T^\perp} \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \leq \frac{2}{n_1 n_2} \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F}, \qquad (40)
$$

where the last inequality arises from our assumption. Combining this with the above two bounds yields

$$
\begin{aligned}
0 = \left\| \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} & \geq \left\| \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} - \left\| \left(\frac{n_1 n_2}{m} \mathcal{A}_\Omega + \mathcal{A}^\perp \right) \mathcal{P}_{T^\perp} \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \\
& \geq \sqrt{\frac{1}{2}} \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} - \frac{2}{n_1 n_2} \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \\
& \geq \frac{1}{2} \left\| \mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \geq \frac{n_1^2 n_2^2}{4} \left\| \mathcal{P}_{T^\perp} \left(\boldsymbol{H}_\mathrm{e} \right) \right\|_\mathrm{F} \geq 0,
\end{aligned}
$$

which immediately indicates $\mathcal{P}_{T^\perp} \left(\boldsymbol{H}_\mathrm{e} \right) = 0$ and $\mathcal{P}_T \left(\boldsymbol{H}_\mathrm{e} \right) = 0$. Hence, (36) can only hold when $\boldsymbol{H}_\mathrm{e} = 0$.

C Proof of Lemma 3

By definition, we have the identities

$$
\begin{aligned}
\left\| \mathcal{P}_T \left(\boldsymbol{A}_{(k,l)} \right) \right\|_\mathrm{F}^2 & = \left\langle \mathcal{P}_T \left(\boldsymbol{A}_{(k,l)} \right), \boldsymbol{A}_{(k,l)} \right\rangle \\
& = \left\langle \mathcal{P}_U \left(\boldsymbol{A}_{(k,l)} \right) + \mathcal{P}_V \left(\boldsymbol{A}_{(k,l)} \right) - \mathcal{P}_U \mathcal{P}_V \left(\boldsymbol{A}_{(k,l)} \right), \boldsymbol{A}_{(k,l)} \right\rangle \\
& = \left\| \mathcal{P}_U \left(\boldsymbol{A}_{(k,l)} \right) \right\|_\mathrm{F}^2 + \left\| \mathcal{P}_V \left(\boldsymbol{A}_{(k,l)} \right) \right\|_\mathrm{F}^2 - \left\| \mathcal{P}_U \mathcal{P}_V \left(\boldsymbol{A}_{(k,l)} \right) \right\|_\mathrm{F}^2 .
\end{aligned}
$$

Since $\boldsymbol{U}$ (resp. $\boldsymbol{V}$) and $\boldsymbol{E}_\mathrm{L}$ (resp. $\boldsymbol{E}_\mathrm{R}$) determine the same column (resp. row) space, we can write

$$
\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{E}_\mathrm{L} \left(\boldsymbol{E}_\mathrm{L}^* \boldsymbol{E}_\mathrm{L} \right)^{-1} \boldsymbol{E}_\mathrm{L}^* \quad \text{and} \quad \boldsymbol{V}\boldsymbol{V}^* = \boldsymbol{E}_\mathrm{R}^* \left(\boldsymbol{E}_\mathrm{R} \boldsymbol{E}_\mathrm{R}^* \right)^{-1} \boldsymbol{E}_\mathrm{R},
$$

and thus

$$\begin{aligned}\left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2 &\le \left\|\mathcal{P}_U\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2+\left\|\mathcal{P}_V\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2\\ &\le \left\|\boldsymbol{E}_{\mathrm{L}}\left(\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}\right)^{-1}\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2+\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^*\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*\right)^{-1}\boldsymbol{E}_{\mathrm{R}}\right\|_{\mathrm{F}}^2\\ &\le \frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}\right)}\left\|\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2+\frac{1}{\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*\right)}\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^*\right\|_{\mathrm{F}}^2.\end{aligned}$$

Note that $\sqrt{\omega_{k,l}}\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{A}_{(k,l)}$ consists of $\omega_{k,l}$ columns of $\boldsymbol{E}_{\mathrm{L}}^*$ (and hence it contains $r\omega_{k,l}$ nonzero entries in total). Owing to the fact that each entry of $\boldsymbol{E}_{\mathrm{L}}^*$ has magnitude $\frac{1}{\sqrt{k_2k_2}}$, one can derive

$$\left\|\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2=\frac{1}{\omega_{k,l}}\cdot r\omega_{k,l}\cdot\frac{1}{k_1k_2}=\frac{r}{k_1k_2}\le\frac{rc_{\mathrm{s}}}{n_1n_2}.$$

A similar argument yields

$$\left\|\boldsymbol{A}_{(k,l)}\boldsymbol{E}_{\mathrm{R}}^*\right\|_{\mathrm{F}}^2\le\frac{c_{\mathrm{s}}r}{n_1n_2}.$$

We know from Lemma 1 that $\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}=\boldsymbol{G}_{\mathrm{L}}$ and $\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*=\boldsymbol{G}_{\mathrm{L}}^T$, and hence $\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{L}}^*\boldsymbol{E}_{\mathrm{L}}\right)\ge\frac{1}{\mu_1}$ and $\sigma_{\min}\left(\boldsymbol{E}_{\mathrm{R}}\boldsymbol{E}_{\mathrm{R}}^*\right)\ge\frac{1}{\mu_1}$. One can, therefore, conclude that for every $(k,l)\in[n_1]\times[n_2]$,

$$\left\|\mathcal{P}_U\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2\le\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2},\quad\left\|\mathcal{P}_V\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2\le\frac{\mu_1c_{\mathrm{s}}r}{n_1n_2},\quad\text{and}\quad\left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2\le\frac{2\mu_1c_{\mathrm{s}}r}{n_1n_2}.\tag{41}$$

D Proof of Lemma 4

Define a family of operators

$$\forall(k,l)\in[n_1]\times[n_2]:\quad\mathcal{Z}_{(k,l)}:=\frac{n_1n_2}{m}\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T-\frac{1}{m}\mathcal{P}_T\mathcal{A}\mathcal{P}_T.$$

We can also compute

$$\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T(\boldsymbol{M})=\mathcal{P}_T\left\{\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_T\boldsymbol{M}\right\rangle\boldsymbol{A}_{(k,l)}\right\}=\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\left\langle\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle,\tag{42}$$

and hence

$$\begin{aligned}\left(\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right)^2(\boldsymbol{M}) &= \left[\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\left\{\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\}\right]\left\langle\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle\\ &= \mathcal{P}_T\left\{\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\boldsymbol{A}_{(k,l)}\right\}\left\langle\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle\\ &= \left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\left\langle\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right),\boldsymbol{M}\right\rangle.\end{aligned}\tag{43}$$

Comparing (42) and (43) gives

$$\left(\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right)^2=\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\le\frac{2\mu_1c_{\mathrm{s}}r}{n_1n_2}\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T,\tag{44}$$

where the inequality follows from our assumption that

$$\left\langle\boldsymbol{A}_{(k,l)},\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\rangle=\left\|\mathcal{P}_T\left(\boldsymbol{A}_{(k,l)}\right)\right\|_{\mathrm{F}}^2\le\frac{2\mu_1c_{\mathrm{s}}r}{n_1n_2}.$$

Let $\boldsymbol{a}_i$ $(1\le i\le m)$ be m independent random pairs uniformly chosen from $[n_1]\times[n_2]$, then we have $\mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_i}\right)=0$. This further gives

$$\begin{aligned}\mathbb{E}\mathcal{Z}_{\boldsymbol{a}_i}^2 &= \mathbb{E}\left(\frac{n_1n_2}{m}\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2-\left(\mathbb{E}\left(\frac{n_1n_2}{m}\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)\right)^2\\ &= \frac{n_1^2n_2^2}{m^2}\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2-\frac{1}{m^2}\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right)^2,\end{aligned}$$

We can then bound the operator norm as

$$\begin{aligned}\left\|\mathbb{E}\left(\mathcal{Z}_{\boldsymbol{a}_i}^2\right)\right\| &\leq \left\|\frac{n_1^2n_2^2}{m^2}\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2\right\| + \frac{1}{m^2}\left\|\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right)^2\right\| \\ &\leq \frac{n_1^2n_2^2}{m^2}\left\|\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)^2\right\| + \frac{1}{m^2} \\ &\leq \frac{n_1^2n_2^2}{m^2}\frac{2\mu_1 c_{\mathrm{s}} r}{n_1 n_2}\left\|\mathbb{E}\left(\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right)\right\| + \frac{1}{m^2} \qquad (45) \\ &= \frac{2\mu_1 c_{\mathrm{s}} r n_1 n_2}{m^2}\frac{1}{n_1 n_2}\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| + \frac{1}{m^2} \\ &\leq \frac{4\mu_1 c_{\mathrm{s}} r}{m^2} := V_0, \qquad (46)\end{aligned}$$

where (45) uses the fact that $\mathcal{P}_T\mathcal{A}_{\boldsymbol{a}_i}\mathcal{P}_T \succeq 0$. Besides, the first equality of (44) gives $\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\|^2 \leq \left\|\mathcal{P}_T\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\|$ and hence $\left\|\mathcal{P}_T\mathcal{A}_{(k,l)}\mathcal{P}_T\right\| \leq \left\|\mathcal{P}_T\boldsymbol{A}_{(k,l)}\right\|_{\mathrm{F}}^2$, which immediately yields

$$\left\|\mathcal{Z}_{\boldsymbol{a}_i}\right\| \leq \frac{n_1 n_2}{m}\left\|\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}_i}\mathcal{P}_T\right\| + \frac{1}{m}\left\|\mathcal{P}_T\mathcal{A}\mathcal{P}_T\right\| \leq \frac{n_1 n_2}{m}\left\|\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}_i}\right\|_{\mathrm{F}}^2 + \frac{1}{m} < \frac{4\mu_1 c_{\mathrm{s}} r}{m}.$$

This together with (46) gives

$$\frac{2mV_0}{\left\|\mathcal{Z}_{\boldsymbol{a}_i}\right\|} \geq 2.$$

Applying the Operator Bernstein Inequality [2, Theorem 6] yields that for any $t \leq 2$, we have

$$\mathbb{P}\left(\left\|\sum_{i=1}^{m}\mathcal{Z}_{\boldsymbol{a}_i}\right\| > t\right) \leq 2n_1 n_2 \exp\left(-\frac{t^2}{16\frac{\mu_1 c_{\mathrm{s}} r}{m}}\right).$$

Finally, one can observe that $\sum_{i=1}^{m}\mathcal{Z}_{\boldsymbol{a}_i}$ is equivalent to $\frac{n_1 n_2}{m}\mathcal{P}_T\mathcal{A}_{\Omega}\mathcal{P}_T - \mathcal{P}_T\mathcal{A}\mathcal{P}_T$ in distribution, which completes the proof.

E Proof of Lemma 5

Fix any $\boldsymbol{b} \in [n_1] \times [n_2]$. For any $\boldsymbol{a} \in [n_1] \times [n_2]$, define

$$z_{\boldsymbol{a}} = \frac{1}{qn_1 n_2}\left\langle \boldsymbol{A}_b, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \left\langle \boldsymbol{A}_b, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle.$$

Then for any i.i.d. α_i's chosen uniformly at random from $[n_1] \times [n_2]$, we can easily check that $\mathbb{E}\left(z_{\alpha_i}\right) = 0$. Define a multiset $\Omega_l := \{\alpha_i \mid 1 \leq i \leq qn_1 n_2\}$, then the decomposition

$$\mathcal{A}_{\Omega_l}\boldsymbol{F} = \sum_{i=1}^{qn_1 n_2}\boldsymbol{A}_{\alpha_i}\left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle$$

allows us to derive

$$\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}_{\Omega_l}\boldsymbol{F}\right\rangle = \left\langle \boldsymbol{A}_{\boldsymbol{b}}, \sum_{i=1}^{qn_1 n_2}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right\rangle,$$

and thus

$$\begin{aligned}\sum_{i=1}^{qn_1 n_2} z_{\alpha_i} &= \left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \sum_{i=1}^{qn_1 n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle\left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle \\ &= \left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle - \frac{1}{q}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}_{\Omega_l}\boldsymbol{F}\right\rangle \\ &= \left\langle \boldsymbol{A}_{\boldsymbol{b}}, \left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_l}\mathcal{P}_T\right)\boldsymbol{F}\right\rangle.\end{aligned}$$

Owing to the fact that $\mathbb{E}z_{\alpha_i}=0$, we can bound the variance of each term as follows

$$\begin{aligned}
\mathbb{E}\left|z_{\alpha_i}\right|^2 &= \mathbf{Var}\left(\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right) \\
&\leq \mathbb{E}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right|^2 \\
&= \frac{1}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right|^2 \\
&\leq \frac{1}{q^2}\frac{\nu\left(\boldsymbol{F}\right)}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2\omega_{\boldsymbol{a}} \\
&\leq \frac{\mu_4 r\nu(\boldsymbol{F})}{\left(qn_1n_2\right)^2}\omega_{\boldsymbol{b}},
\end{aligned}$$

where the last inequality arises from the definition of μ_4, i.e. for every $\boldsymbol{b}\in[n_1]\times[n_2]$,

$$\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\left\langle \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{b}}, \boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right|^2\omega_{\boldsymbol{a}} \leq \frac{\mu_4 r}{n_1n_2}\omega_{\boldsymbol{b}}. \tag{47}$$

This immediately gives

$$\frac{1}{\omega_{\boldsymbol{b}}}\mathbb{E}\left(\sum_{i=1}^{qn_1n_2}\left|z_{\alpha_i}\right|^2\right) \leq \frac{\mu_4 r\nu(\boldsymbol{F})}{qn_1n_2} \leq \frac{\max\left\{\mu_4, 3\mu_1 c_{\mathrm{s}}\right\} r\nu(\boldsymbol{F})}{qn_1n_2} := V.$$

On the other hand, Lemma 1 shows the inequality

$$\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right| \leq \sqrt{\frac{\omega_{\boldsymbol{b}}}{\omega_{\boldsymbol{a}}}}\frac{3\mu_1 c_{\mathrm{s}} r}{n_1n_2}, \tag{48}$$

which further leads to

$$\begin{aligned}
\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right| &\leq \sqrt{\omega_{\boldsymbol{a}}\nu\left(\boldsymbol{F}\right)}\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}q}\left|\left\langle \boldsymbol{A}_b, \mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle\right| \\
&\leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{3\mu_1 c_{\mathrm{s}} r}{n_1n_2}.
\end{aligned}$$

Since $\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle = \mathbb{E}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle\left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle$, one has as well

$$\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle\right| = \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\mathbb{E}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\boldsymbol{a}}\right\rangle \left\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\right\rangle\right| \leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{3\mu_1 c_{\mathrm{s}} r}{n_1n_2},$$

which immediately leads to

$$\begin{aligned}
\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|z_{\alpha_i}\right| &\leq \frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|\frac{1}{qn_1n_2}\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \mathcal{P}_T\mathcal{A}\boldsymbol{F}\right\rangle\right| + \frac{1}{\sqrt{\omega_b}}\left|\left\langle \boldsymbol{A}_{\boldsymbol{b}}, \frac{1}{q}\mathcal{P}_T\boldsymbol{A}_{\alpha_i}\right\rangle \left\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\right\rangle\right| \\
&\leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}\frac{6\mu_1 c_{\mathrm{s}} r}{n_1n_2}
\end{aligned}$$

The above bounds indicate that

$$\frac{2V}{\frac{1}{\sqrt{\omega_{\boldsymbol{b}}}}\left|z_{\boldsymbol{a}}\right|} \geq \sqrt{\nu(\boldsymbol{F})}.$$

Applying the operator Bernstein inequality [2, Theorem 6] yields for any $t<\nu\left(\boldsymbol{F}\right)$,

$$\mathbb{P}\left(\frac{1}{\omega_{\boldsymbol{b}}}\left|\sum_{i=1}^{qn_1n_2} z_{\alpha_i}\right|^2 > t\right) \leq c_6\exp\left(-\frac{tqn_1n_2}{4\max\left\{\mu_4, 3\mu_1 c_{\mathrm{s}}\right\} r\nu(\boldsymbol{F})}\right).$$

Thus, there are some constants $c_7, \tilde{c}_7 > 0$ such that whenever $qn_1n_2 > \tilde{c}_7 \max\{\mu_4, 3\mu_1 c_{\rm s}\} r \log(n_1n_2)$ or, equivalently, $m > c_7 \max\{\mu_4, 3\mu_1 c_{\rm s}\} r \log^2(n_1n_2)$, we have

$$\mathbb{P}\left(\frac{\left|\sum_{i=1}^{qn_1n_2} z_{\alpha_i}\right|^2}{\omega_{\boldsymbol{b}}} > \frac{1}{4}\nu(\boldsymbol{F})\right) \leq \tilde{c}_6 \exp\left(-\frac{qn_1n_2}{16\max\{\mu_4, 3\mu_1 c_{\rm s}\} r\nu(\boldsymbol{F})}\right) \leq \frac{1}{(n_1n_2)^4}.$$

Finally, we observe that in distribution,

$$v\left(\left(\mathcal{P}_T\mathcal{A}\mathcal{P}_T - \frac{1}{q}\mathcal{P}_T\mathcal{A}_{\Omega_l}\mathcal{P}_T\right)\boldsymbol{F}\right) = \max_{\boldsymbol{b}\in[n_1]\times[n_2]} \frac{\left|\sum_{i=1}^{qn_1n_2} z_{\alpha_i}\right|^2}{\omega_{\boldsymbol{b}}}.$$

Applying a simple union bound over all $\boldsymbol{b} \in [n_1] \times [n_2]$ allows us to derive (23).

F Proof of Lemma 6

For any $\boldsymbol{a} \in [n_1] \times [n_2]$, define

$$\boldsymbol{H}_{\boldsymbol{a}} = \frac{1}{q}\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle + \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}).$$

Let α_i ($1 \leq i \leq qn_1n_2$) be independently and uniformly drawn from $[n_1] \times [n_2]$ which forms Ω_l. Observing that

$$\mathcal{A}\boldsymbol{F} = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \boldsymbol{A}_{\boldsymbol{a}}\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle,$$

we can write

$$\mathcal{P}_{T^\perp}\mathcal{A}\boldsymbol{F} = \sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle.$$

This immediately gives

$$\begin{aligned}
\mathbb{E}\boldsymbol{H}_{\alpha_i} &= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}) + \frac{1}{qn_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle \\
&= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp(\boldsymbol{F}) + \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}(\boldsymbol{F}) \\
&= \frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}(\boldsymbol{F}) = 0.
\end{aligned}$$

Moreover, we have, in distribution, the following identity

$$\mathcal{P}_{T^\perp}\left(\frac{1}{q}\mathcal{A}_{\Omega_l} + \mathcal{A}^\perp\right)(\boldsymbol{F}) = \sum_{i=1}^{qn_1n_2} \boldsymbol{H}_{\alpha_i}.$$

On the other hand, since $\mathbb{E}\boldsymbol{H}_{\alpha_i} = 0$, if we denote $\mathcal{Y}_i = \frac{1}{q}\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\alpha_i})\langle \boldsymbol{A}_{\alpha_i}, \boldsymbol{F}\rangle$, then $\boldsymbol{H}_{\alpha_i} = \mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i$, and hence

$$\mathbb{E}\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^* = \mathbb{E}\left\{(\mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i)(\mathcal{Y}_i - \mathbb{E}\mathcal{Y}_i)^*\right\} \leq \mathbb{E}\mathcal{Y}_i\mathcal{Y}_i^* = \frac{1}{q^2n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]} |\langle \boldsymbol{A}_{\boldsymbol{a}}, \boldsymbol{F}\rangle|^2 \mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}})(\mathcal{P}_{T^\perp}(\boldsymbol{A}_{\boldsymbol{a}}))^*.$$

The definition of the spectral norm $\|\boldsymbol{M}\| := \max_{\psi:\|\psi\|_2=1} \langle \psi, \boldsymbol{M}\psi\rangle$ allows us to bound

$$\begin{aligned}
\left\|\mathbb{E}\left(\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^*\right)\right\| &\leq \frac{1}{q^2}\max_{\psi:\|\psi\|_2=1}\left\{\frac{1}{n_1n_2}\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\left|\langle \boldsymbol{A}_{\boldsymbol{a}},\boldsymbol{F}\rangle\right|^2\left\langle\psi,\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\left(\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\right)^*\psi\right\rangle\right\} \\
&\leq \frac{1}{q^2n_1n_2}\nu(\boldsymbol{F})\max_{\psi:\|\psi\|_2=1}\left\langle\psi,\left(\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\omega_{\boldsymbol{a}}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\left(\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\right)^*\right)\psi\right\rangle \\
&\leq \frac{1}{q^2n_1n_2}\nu(\boldsymbol{F})\left(\sum_{\boldsymbol{a}\in[n_1]\times[n_2]}\omega_{\boldsymbol{a}}\left\|\boldsymbol{A}_{\boldsymbol{a}}\right\|^2\right)\max_{\psi:\|\psi\|_2=1}\langle\psi,\psi\rangle \\
&\leq \frac{\nu(\boldsymbol{F})}{q^2},
\end{aligned}$$

where the last inequality uses the fact that $\|\boldsymbol{A}_{\boldsymbol{a}}\|^2 = \frac{1}{\omega_{\boldsymbol{a}}}$. Therefore,

$$\left\|\mathbb{E}\left(\sum_{i=1}^{qn_1n_2}\boldsymbol{H}_{\alpha_i}\boldsymbol{H}_{\alpha_i}^*\right)\right\| \leq \nu(\boldsymbol{F})n_1n_2\frac{1}{q} := V.$$

.

Besides, the definition (22) of $\nu(\boldsymbol{F})$ allows us to bound

$$\left\|\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\boldsymbol{a}}\right)\langle\boldsymbol{A}_{\boldsymbol{a}},\boldsymbol{F}\rangle\right\| \leq \sqrt{\nu\left(\boldsymbol{F}\right)\omega_{\boldsymbol{a}}}\frac{1}{q}\left\|\boldsymbol{A}_{\boldsymbol{a}}\right\| = \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q}.$$

The fact that $\mathbb{E}\boldsymbol{H}_{\alpha_i} = 0$ yields

$$\left\|\frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp\left(\boldsymbol{F}\right)\right\| = \left\|\mathbb{E}\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\alpha_i}\right)\langle\boldsymbol{A}_{\alpha_i},\boldsymbol{F}\rangle\right\| \leq \sqrt{\nu\left(\boldsymbol{F}\right)}\frac{1}{q},$$

and hence

$$\|\boldsymbol{H}_{\alpha_i}\| \leq \left\|\frac{1}{qn_1n_2}\mathcal{P}_{T^\perp}\mathcal{A}^\perp\left(\boldsymbol{F}\right)\right\| + \left\|\frac{1}{q}\mathcal{P}_{T^\perp}\left(\boldsymbol{A}_{\alpha_i}\right)\langle\boldsymbol{A}_{\alpha_i},\boldsymbol{F}\rangle\right\| \leq \frac{2\sqrt{\nu(\boldsymbol{F})}}{q}.$$

Applying the Operator Bernstein inequality [2, Theorem 6] yields that for any $t \leq \sqrt{\nu(\boldsymbol{F})}n_1n_2$, we have

$$\left\|\mathcal{P}_{T^\perp}\left(\frac{n_1n_2}{m}\mathcal{A}_\Omega + \mathcal{A}^\perp\right)(\boldsymbol{F})\right\| > t$$

with probability at most $c_8\exp\left(-\frac{c_9qt^2}{\nu(\boldsymbol{F})n_1n_2}\right)$ for some positive constants c_8 and c_9.

G Proof of Theorem 2

We prove this theorem under the conditions of Lemma 2, i.e. (16)–(19). Note that these conditions are satisfied with high probability, as we have shown in the proof of Theorem 1.

Denote the solution of Noisy-EMaC as $\hat{\boldsymbol{X}}_{\rm e} = \boldsymbol{X}_{\rm e} + \boldsymbol{H}_{\rm e}$. Since $\boldsymbol{H}_{\rm e}$ is a two-fold Hankel matrix, i.e. $\boldsymbol{H}_{\rm e} = \mathcal{A}_\Omega\left(\boldsymbol{H}_{\rm e}\right) + \mathcal{A}_{\Omega^\perp}\left(\boldsymbol{H}_{\rm e}\right)$, we can obtain

$$\|\boldsymbol{X}_{\rm e}\|_* \geq \|\hat{\boldsymbol{X}}_{\rm e}\|_* = \|\boldsymbol{X}_{\rm e} + \boldsymbol{H}_{\rm e}\|_* \geq \|\boldsymbol{X}_{\rm e} + \mathcal{A}_{\Omega^\perp}(\boldsymbol{H}_{\rm e})\|_* - \|\mathcal{A}_\Omega(\boldsymbol{H}_{\rm e})\|_*. \tag{49}$$

The second term can be bounded using the triangle inequality as

$$\|\mathcal{A}_\Omega\left(\boldsymbol{H}_{\rm e}\right)\|_{\rm F} \leq \left\|\mathcal{A}_\Omega\left(\hat{\boldsymbol{X}}_{\rm e} - \boldsymbol{X}_{\rm e}^{\rm o}\right)\right\|_{\rm F} + \left\|\mathcal{A}_\Omega\left(\boldsymbol{X}_{\rm e} - \boldsymbol{X}_{\rm e}^{\rm o}\right)\right\|_{\rm F}. \tag{50}$$

Since the constraint of Noisy-EMaC requires $\left\|\mathcal{P}_{\Omega}\left(\hat{\boldsymbol{X}}-\boldsymbol{X}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \leq \delta$ and $\left\|\mathcal{P}_{\Omega}\left(\boldsymbol{X}-\boldsymbol{X}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \leq \delta$, the Hankel structure of the enhanced form allows us to bound $\left\|\mathcal{A}_{\Omega}\left(\hat{\boldsymbol{X}}_{\mathrm{e}}-\boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \leq \sqrt{n_1 n_2}\delta$ and $\left\|\mathcal{A}_{\Omega}\left(\boldsymbol{X}_{\mathrm{e}}-\boldsymbol{X}_{\mathrm{e}}^{\mathrm{o}}\right)\right\|_{\mathrm{F}} \leq \sqrt{n_1 n_2}\delta$, which immediately leads to

$$\left\|\mathcal{A}_{\Omega}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \leq 2\sqrt{n_1 n_2}\delta.$$

Using the same analysis as for (35) allows us to bound the perturbation $\mathcal{A}_{\Omega^{\perp}}(\boldsymbol{H}_{\mathrm{e}})$ as follows

$$\left\|\boldsymbol{X}_{\mathrm{e}}+\mathcal{A}_{\Omega^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{*} \geq \left\|\boldsymbol{X}_{\mathrm{e}}\right\|_{*}+\frac{1}{4}\left\|\mathcal{P}_{T^{\perp}}\mathcal{A}_{\Omega^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}.$$

Combining this with (49), we have

$$\left\|\mathcal{P}_{T^{\perp}}\mathcal{A}_{\Omega^{\perp}}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}} \leq 4\left\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\right\|_{*} \leq 4\sqrt{n_1 n_2}\left\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}} \leq 8n_1 n_2\delta.$$

Further from Lemma 2, we know that

$$\left\|\mathcal{P}_{T}\mathcal{A}_{\Omega^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \leq \frac{n_1 n_2}{m}\sqrt{2}\left\|\mathcal{P}_{T^{\perp}}\mathcal{A}_{\Omega^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}. \tag{51}$$

Therefore, combining all the above results give

$$\begin{aligned}\left\|\boldsymbol{H}_{\mathrm{e}}\right\|_{\mathrm{F}} &\leq \left\|\mathcal{A}_{\Omega}(\boldsymbol{H}_{\mathrm{e}})\right\|_{\mathrm{F}}+\left\|\mathcal{P}_{T}\mathcal{A}_{\Omega^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}}+\left\|\mathcal{P}_{T^{\perp}}\mathcal{A}_{\Omega^{\perp}}\left(\boldsymbol{H}_{\mathrm{e}}\right)\right\|_{\mathrm{F}} \\ &\leq \left\{2\sqrt{n_1 n_2}+8n_1 n_2+\frac{8\sqrt{2}n_1^2 n_2^2}{m}\right\}\delta.\end{aligned}$$

H Proof of Theorem 3

In order to extend the results to structured Hankel matrix completion, from the proof of Theorem 1 it is sufficient to have the first two conditions in (20) to hold for general Hankel matrices. The proof is done by recognizing these two conditions are equivalent to (7).

References

[1] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, April 2009.

[2] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, March 2011.