Appendix A. Analysis of the active norm sampling algorithm

Proof of Lemma 1. This lemma is a direct corollary of Theorem 2 from [15]. First, let $P_i = \hat{c}_i/\hat{f}$ be the probability of selecting the i-th column of $\mathbf{M}$. By assumption, we have $P_i \geq \frac{1-\alpha}{1+\alpha}\|\boldsymbol{x}_i\|_2^2/\|\mathbf{M}\|_F^2$. Applying Theorem 2 [3] from [15] we have that with probability at least $1-\delta$, there exists an orthonormal set of vectors $\boldsymbol{y}^{(1)}, \cdots, \boldsymbol{y}^{(k)} \in \mathbb{R}^{n_1}$ in span($\mathbf{C}$) such that

$$\left\|\mathbf{M} - \left(\sum_{j=1}^{k} \boldsymbol{y}^{(j)}\boldsymbol{y}^{(j)^\top}\right)\mathbf{M}\right\|_F^2 \leq \|\mathbf{M}-\mathbf{M}_k\|_F^2 + \frac{(1+\alpha)k}{(1-\alpha)\delta s}\|\mathbf{M}\|_F^2. \tag{29}$$

Finally, to complete the proof, note that every column of $\left(\sum_{j=1}^{k} \boldsymbol{y}^{(j)}\boldsymbol{y}^{(j)^\top}\right)\mathbf{M}$ can be represented as a linear combination of columns in $\mathbf{C}$; furthermore,

$$\|\mathbf{M} - \mathcal{P}_C(\mathbf{M})\|_F = \min_{\mathbf{X}\in\mathbb{R}^{k\times n_2}} \|\mathbf{M}-\mathbf{C}\mathbf{X}\|_F \leq \left\|\mathbf{M} - \left(\sum_{j=1}^{k} \boldsymbol{y}^{(j)}\boldsymbol{y}^{(j)^\top}\right)\mathbf{M}\right\|_F. \tag{30}$$

□

Proof of Theorem 1. First, set $m_1 = \Omega(\mu_1 \log(n_2/\delta_1))$ we have that with probability $\geq 1-\delta_1$ the inequality

$$(1-\alpha)\|\boldsymbol{x}_i\|_2^2 \leq \hat{c}_i \leq (1+\alpha)\|\boldsymbol{x}_i\|_2^2$$

holds with $\alpha = 0.5$ for every column i, using Lemma 2. Next, putting $s \geq 6k/\delta_2\varepsilon^2$ and applying Lemma 1 we get

$$\|\mathbf{M} - \mathcal{P}_C(\mathbf{M})\|_F \leq \|\mathbf{M}-\mathbf{M}_k\|_F + \varepsilon\|\mathbf{M}\|_F \tag{31}$$

with probability at least $1-\delta_2$. Finally, note that when $\alpha \leq 1/2$ and $n_1 \leq n_2$ the bound in Lemma 3 is dominated by

$$\|\mathbf{M} - \widehat{\mathbf{M}}\|_2 \leq \|\mathbf{M}\|_F \cdot O\left(\sqrt{\frac{\mu_1}{m_2}}\log\left(\frac{n_1+n_2}{\delta}\right)\right). \tag{32}$$

Consequently, for any $\varepsilon' > 0$ if $m_2 = \Omega((\varepsilon')^{-2}\mu_1 \log^2((n_1+n_2)/\delta_3)$ we have with probability $\geq 1-\delta_3$

$$\|\mathbf{M} - \widehat{\mathbf{M}}\|_2 \leq \varepsilon'\|\mathbf{M}\|_F. \tag{33}$$

The proof is then completed by taking $\varepsilon' = \varepsilon/\sqrt{s}$:

$$\begin{aligned}
\|\mathbf{M}-\mathbf{C}\mathbf{X}\|_F &= \|\mathbf{M} - \mathcal{P}_C(\widehat{\mathbf{M}})\|_F \\
&\leq \|\mathbf{M} - \mathcal{P}_C(\mathbf{M})\|_F + \|\mathcal{P}_C(\mathbf{M}-\widehat{\mathbf{M}})\|_F \\
&\leq \|\mathbf{M}-\mathbf{M}_k\|_F + \varepsilon\|\mathbf{M}\|_F + \sqrt{s}\|\mathcal{P}_C(\mathbf{M}-\widehat{\mathbf{M}})\|_2 \\
&\leq \|\mathbf{M}-\mathbf{M}_k\|_F + \varepsilon\|\mathbf{M}\|_F + \sqrt{s}\cdot\varepsilon'\|\mathbf{M}\|_F \\
&\leq \|\mathbf{M}-\mathbf{M}_k\|_F + 2\varepsilon\|\mathbf{M}\|_F.
\end{aligned}$$

□

Appendix B. Analysis of the active volume sampling algorithm

Proof of Lemma 4. We first prove Eq. (15). Observe that $\dim(\mathcal{U}(C)) \leq s$. Let $\mathbf{R}_C = (\mathbf{R}^{(C(1))}, \cdots, \mathbf{R}^{(C(s))}) \in \mathbb{R}^{n_1\times s}$ denote the selected s columns in the noise matrix $\mathbf{R}$ and let $\mathcal{R}(C) = \text{span}(\mathbf{R}_C)$ denote the span of selected columns in $\mathbf{R}$. By definition, $\mathcal{U}(C) \subseteq \mathcal{U} \cup \mathcal{R}(C)$, where $\mathcal{U} = \text{span}(\mathbf{A})$ denotes the subspace spanned by columns in the deterministic matrix $\mathbf{A}$. Consequently, we have the following bound on $\|\mathcal{P}_{\mathcal{U}(C)}\boldsymbol{e}_i\|$ (assuming each entry in $\mathbf{R}$ follows a zero-mean Gaussian distribution with σ^2 variance):

$$\|\mathcal{P}_{\mathcal{U}(C)}\boldsymbol{e}_i\|_2^2 \leq \|\mathcal{P}_{\mathcal{U}}\boldsymbol{e}_i\|_2^2 + \|\mathcal{P}_{\mathcal{U}^\perp\cap\mathcal{R}(C)}\boldsymbol{e}_i\|_2^2$$

[3] The original theorem concerns random samples of rows; it is essentially the same for random samples of columns.

$$
\begin{aligned}
&\leq \|\mathcal{P}_{\mathcal{U}}\boldsymbol{e}_i\|_2^2 + \|\mathcal{P}_{\mathcal{R}(C)}\boldsymbol{e}_i\|_2^2\\
&\leq \frac{k\mu_0}{n_1} + \|\mathbf{R}_C\|_2^2\|(\mathbf{R}_C^\top\mathbf{R}_C)^{-1}\|_2^2\|\mathbf{R}_C^\top\boldsymbol{e}_i\|_2^2\\
&\leq \frac{k\mu_0}{n_1} + \frac{(\sqrt{n_1}+\sqrt{s}+\epsilon)^2\sigma^2}{(\sqrt{n_1}-\sqrt{s}-\epsilon)^4\sigma^4}\cdot\sigma^2(s+2\sqrt{s\log(2/\delta)}+2\log(2/\delta)).
\end{aligned}
$$

For the last inequality we apply Lemma 13 to bound the largest and smallest singular values of $\mathbf{R}_C$ and Lemma 11 to bound $\|\mathbf{R}_C^\top\boldsymbol{e}_i\|_2^2$, because $\mathbf{R}_C^\top\boldsymbol{e}_i$ follow i.i.d. Gaussian distributions with covariance $\sigma^2\mathbf{I}_{s\times s}$. If ϵ is set as $\epsilon=\sqrt{2\log(4/\delta)}$ then the last inequality holds with probability at least $1-\delta$. Furthermore, when $s\leq n_1/2$ and δ is not exponentially small (e.g., $\sqrt{2\log(4/\delta)}\leq\frac{\sqrt{n_1}}{4}$), the fraction $\frac{(\sqrt{n_1}+\sqrt{s}+\epsilon)^2}{(\sqrt{n_1}-\sqrt{s}-\epsilon)^4}$ is approximately $O(1/n_1)$. As a result, with probability $1-n_1\delta$ the following holds:

$$
\begin{aligned}
\mu(\mathcal{U}(C)) &= \frac{n_1}{s}\max_{1\leq i\leq n_1}\|\mathcal{P}_{\mathcal{U}(C)}\boldsymbol{e}_i\|_2^2\\
&\leq \frac{n_1}{s}\left(\frac{k\mu_0}{n_1}+O\left(\frac{s+\sqrt{s\log(1/\delta)}+\log(1/\delta)}{n_1}\right)\right) = O\left(\frac{k\mu_0+s+\sqrt{s\log(1/\delta)}+\log(1/\delta)}{s}\right). \quad (34)
\end{aligned}
$$

Finally, putting $\delta'=n_1/\delta$ we prove Eq. (15).

Next we try to prove Eq. (16). Let $\boldsymbol{x}$ be the i-th column of $\mathbf{M}$ and write $\boldsymbol{x}=\boldsymbol{a}+\boldsymbol{r}$, where $\boldsymbol{a}=\mathcal{P}_{\mathcal{U}}(\boldsymbol{x})$ and $\boldsymbol{r}=\mathcal{P}_{\mathcal{U}^\perp}(\boldsymbol{x})$. Since the deterministic component of $\boldsymbol{x}$ lives in $\mathcal{U}$ and the random component of $\boldsymbol{x}$ is a vector with each entry sampled from i.i.d. zero-mean Gaussian distributions, we know that $\boldsymbol{r}$ is also a zero-mean random Gaussian vector with i.i.d. sampled entries. Note that $\mathcal{U}(C)$ does not depend on the randomness over $\{\mathbf{M}^{(i)}: i\notin C\}$. Therefore, in the following analysis we will assume $\mathcal{U}(C)$ to be a fixed subspace $\widetilde{\mathcal{U}}$ with dimension at most s.

The projected vector $\boldsymbol{x}'=\mathcal{P}_{\widetilde{\mathcal{U}}^\perp}\boldsymbol{x}$ can be written as $\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{a}}+\tilde{\boldsymbol{r}}$, where $\tilde{\boldsymbol{a}}=\mathcal{P}_{\widetilde{\mathcal{U}}^\perp}\boldsymbol{a}$ and $\tilde{\boldsymbol{r}}=\mathcal{P}_{\widetilde{\mathcal{U}}^\perp}\boldsymbol{r}$. By definition, $\tilde{\boldsymbol{a}}$ lives in the subspace $\mathcal{U}\cap\widetilde{\mathcal{U}}^\perp$. So it satisfies the incoherence assumption

$$
\mu(\tilde{\boldsymbol{a}}) = \frac{n_1\|\tilde{\boldsymbol{a}}\|_\infty^2}{\|\tilde{\boldsymbol{a}}\|_2^2}\leq k\mu(\mathcal{U})\leq k\mu_0. \quad (35)
$$

On the other hand, because $\tilde{\boldsymbol{r}}$ is an orthogonal projection of some random Gaussian variable, $\tilde{\boldsymbol{r}}$ is still a Gaussian random vector, which lives in $\mathcal{U}^\perp\cap\widetilde{\mathcal{U}}^\perp$ with rank at least n_1-k-s. Subsequently, we have

$$
\begin{aligned}
\mu(\tilde{\boldsymbol{x}}) &= n_1\frac{\|\tilde{\boldsymbol{x}}\|_\infty^2}{\|\tilde{\boldsymbol{x}}\|_2^2}\leq 3n_1\frac{\|\tilde{\boldsymbol{a}}\|_\infty^2+\|\tilde{\boldsymbol{r}}\|_\infty^2}{\|\tilde{\boldsymbol{a}}\|_2^2+\|\tilde{\boldsymbol{r}}\|_2^2}\\
&\leq 3n_1\frac{\|\tilde{\boldsymbol{a}}\|_\infty^2}{\|\tilde{\boldsymbol{a}}\|_2^2}+3n_1\frac{\|\tilde{\boldsymbol{r}}\|_\infty^2}{\|\tilde{\boldsymbol{r}}\|_2^2}\\
&\leq 3k\mu_0+\frac{6\sigma^2 n_1\log(2n_1n_2/\delta)}{\sigma^2(n_1-k-s)-2\sigma^2\sqrt{(n_1-k-s)\log(n_2/\delta)}}.
\end{aligned}
$$

For the second inequality we use the fact that $\frac{\sum_i a_i}{\sum_i b_i}\leq\sum_i\frac{a_i}{b_i}$ whenever $a_i,b_i\geq 0$. For the last inequality we use Lemma 12 on the enumerator and Lemma 11 on the denominator. Finally, note that when $\max(s,k)\leq n_1/4$ and $\log(n_2/\delta)\leq n_1/64$ the denominator can be lower bounded by $\sigma^2n_1/4$; subsequently, we can bound $\mu(\tilde{\boldsymbol{x}})$ as

$$
\mu(\tilde{\boldsymbol{x}})\leq 3k\mu_0+\frac{24\sigma^2n_1\log(2n_1n_2/\delta)}{\sigma^2n_1}\leq 3k\mu_0+24\log(2n_1n_2/\delta). \quad (36)
$$

Taking a union bound over all n_2-s columns yields the result.

□

To prove the norm estimation consistency result in Lemma 5 we first cite a seminal theorem from [20] which provides a tight error bound on a subsampled projected vector in terms of the norm of the true projected vector.

Theorem 4. *Let $\mathcal{U}$ be a k-dimensional subspace of $\mathbb{R}^n$ and $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{x} \in \mathcal{U}$ and $\boldsymbol{v} \in \mathcal{U}^\perp$. Fix $\delta' > 0$, $m \geq \max\{\frac{8}{3}k\mu(\mathcal{U})\log\left(\frac{2k}{\delta'}\right), 4\mu(\boldsymbol{v})\log(1/\delta')\}$ and let Ω be an index set with entries sampled uniformly with replacement with probability m/n. Then with probability at least $1 - 4\delta'$:*

$$\frac{m(1-\alpha) - k\mu(\mathcal{U})\frac{\beta}{1-\gamma}}{n}\|\boldsymbol{v}\|_2^2 \leq \|\boldsymbol{y}_\Omega - \mathcal{P}_{U_\Omega}\boldsymbol{y}_\Omega\|_2^2 \leq (1+\alpha)\frac{m}{n}\|\boldsymbol{v}\|_2^2, \tag{37}$$

where $\alpha = \sqrt{2\frac{\mu(\boldsymbol{v})}{m}\log(1/\delta')} + 2\frac{\mu(\boldsymbol{v})}{3m}\log(1/\delta')$, $\beta = (1 + 2\sqrt{\log(1/\delta')})^2$ and $\gamma = \sqrt{\frac{8k\mu(\mathcal{U})}{3m}\log(2k/\delta')}$.

We are now ready to prove Lemma 5.

Proof of Lemma 5. By Algorithm 2, we know that $\dim(\mathcal{S}_t) = t$ with probability 1. Let $\boldsymbol{y} = \mathbf{M}^{(i)}$ denote the i-th column of $\mathbf{M}$ and let $\boldsymbol{v} = \mathcal{P}_{\mathcal{S}_t}\boldsymbol{y}$ be the projected vector. We can apply Theorem 4 to bound the estimation error between $\|\boldsymbol{v}\|$ and $\|\boldsymbol{y}_\Omega - \mathcal{P}_{\mathcal{S}_t(\Omega)}\boldsymbol{y}_\Omega\|$.

First, when m is set as in Eq. (19) it is clear that the conditions $m \geq \frac{8}{3}t\mu(\mathcal{U})\log\left(\frac{2t}{\delta'}\right) = \Omega(k\mu_0\log(n/\delta)\log(k/\delta'))$ and $m \geq 4\mu(\boldsymbol{v})\log(1/\delta') = \Omega(k\mu_0\log(n/\delta)\log(1/\delta'))$ are satisfied. We next turn to the analysis of α, β and γ. More specifically, we want $\alpha = O(1)$, $\gamma = O(1)$ and $\frac{t\mu(\mathcal{U})}{m}\beta = O(1)$.

For α, $\alpha = O(1)$ implies $m = \Omega(\mu(\boldsymbol{v})\log(1/\delta')) = \Omega(k\mu_0\log(n/\delta)\log(1/\delta'))$. Therefore, by carefullying selecting constants in $\Omega(\cdot)$ we can make $\alpha \leq 1/4$.

For γ, $\gamma = O(1)$ implies $m = \Omega(t\mu(\mathcal{U})\log(t/\delta')) = \Omega(k\mu_0\log(n/\delta)\log(k/\delta'))$. By carefully selecting constants in $\Omega(\cdot)$ we can make $\gamma \leq 0.2$.

For β, $\frac{t\mu(\mathcal{U})}{m}\beta = O(1)$ implies $m = O(t\mu(\mathcal{U})\beta) = O(k\mu_0\log(n/\delta)\log(1/\delta'))$. By carefully selecting constants we can have $\beta \leq 0.2$. Finllay, combining bounds on α, β and γ we prove the desired result. ☐

Before proving Lemma 6, we first cite a lemma from [9] that connects the volume of a simplex to the permutation sum of singular values.

Lemma 8 ([9])**.** *Fix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \leq n$. Suppose $\sigma_1, \cdots, \sigma_m$ are singular values of $\mathbf{A}$. Then*

$$\sum_{S\subseteq[n],|S|=k} vol(\Delta(S))^2 = \frac{1}{(k!)^2}\sum_{1\leq i_1<i_2<\cdots<i_k\leq m}\sigma_{i_1}^2\sigma_{i_2}^2\cdots\sigma_{i_k}^2. \tag{38}$$

Now we are ready to prove Lemma 6.

Proof of Lemma 6. Let $\mathbf{M}_k$ denote the best rank-k approximation of $\mathbf{M}$ and assume the singular values of $\mathbf{M}$ are $\{\sigma_i\}_{i=1}^{n_1}$. Let $C = \{i_1, \cdots, i_k\}$ be the selected columns. Let $\tau \in \Pi_k$, where Π_k denotes all permutations with k elements. By $\mathcal{H}_{\tau,t}$ we denote the linear subspace spanned by $\{\mathbf{M}^{(\tau(i_1))}, \cdots, \mathbf{M}^{(\tau(i_t))}\}$ and let $d(\mathbf{M}^{(i)}, \mathcal{H}_{\tau,t})$ denote the distance between column $\mathbf{M}^{(i)}$ and subspace $\mathcal{H}_{\tau,t}$. We then have

$$\begin{aligned}
\hat{p}_C &\leq \sum_{\tau\in\Pi_k}\left(\frac{5}{2}\right)^k \frac{\|\mathbf{M}^{(\tau(i_1))}\|_2^2}{\|\mathbf{M}\|_F^2}\frac{d(\mathbf{M}^{(\tau(i_2))}, \mathcal{H}_{\tau,1})^2}{\sum_{i=1}^{n_2} d(\mathbf{M}^{(i)}, \mathcal{H}_{\tau,1})^2}\cdots\frac{d(\mathbf{M}^{(\tau(i_k))}, \mathcal{H}_{\tau,k-1})^2}{\sum_{i=1}^{n_2} d(\mathbf{M}^{(i)}, \mathcal{H}_{\tau,k-1})^2} \\
&\leq 2.5^k \cdot \frac{\sum_{\tau\in\Pi_k}\|\mathbf{M}^{(\tau(i_1))}\|^2 d(\mathbf{M}^{(\tau(i_2))}, \mathcal{H}_{\tau,1})^2\cdots d(\mathbf{M}^{(\tau(i_k))}, \mathcal{H}_{\tau,k-1})^2}{\|\mathbf{M}\|_F^2\|\mathbf{M}-\mathbf{M}_1\|_F^2\cdots\|\mathbf{M}-\mathbf{M}_{k-1}\|_F^2} \\
&= 2.5^k \cdot \frac{\sum_{\tau\in\Pi_k}(k!)^2\mathrm{vol}(\Delta(C))^2}{\|\mathbf{M}\|_F^2\|\mathbf{M}-\mathbf{M}_1\|_F^2\cdots\|\mathbf{M}-\mathbf{M}_{k-1}\|_F^2} \\
&= 2.5^k \cdot \frac{(k!)^3\mathrm{vol}(\Delta(C))^2}{\sum_{i=1}^{n_1}\sigma_i^2\sum_{i=2}^{n_1}\sigma_i^2\cdots\sum_{i=k}^{n_1}\sigma_i^2} \\
&\leq 2.5^k \cdot \frac{(k!)^3\mathrm{vol}(\Delta(C))^2}{\sum_{1\leq i_1<i_2<\cdots<i_k\leq n_1}\sigma_{i_1}^2\sigma_{i_2}^2\cdots\sigma_{i_k}^2} \\
&= 2.5^k \cdot \frac{k!\mathrm{vol}(\Delta(C))^2}{\sum_{T:|T|=k}\mathrm{vol}(\Delta(T))^2} = 2.5^k k! p_C.
\end{aligned}$$

For the first inequality we apply Eq. (22) and for the second to last inequality we apply Lemma 8. □

To prove the approximation error bound in Lemma 7 we need the following two technical lemmas, cited from [19, 3].

Lemma 9 ([19])**.** *Suppose $\mathcal{U} \subseteq \mathbb{R}^n$ has dimension k and $\mathbf{U} \in \mathbb{R}^{n\times k}$ is the orthogonal matrix associated with $\mathcal{U}$. Let $\Omega \subseteq [n]$ be a subset of indices each sampled from i.i.d. Bernoulli distributions with probability m/n_1. Then for some vector $\boldsymbol{y} \in \mathbb{R}^n$, with probability at least $1-\delta$:*

$$\|\mathbf{U}_\Omega^\top \boldsymbol{y}_\Omega\|_2^2 \leq \beta \frac{m}{n_1}\frac{k\mu(\mathcal{U})}{n_1}\|\boldsymbol{y}\|_2^2, \tag{39}$$

where β is defined in Theorem 4.

Lemma 10 ([3])**.** *With the same notation in Lemma 9 and Theorem 4. With probability $\geq 1-\delta$ one has*

$$\|(\mathbf{U}_\Omega^\top \mathbf{U}_\Omega)^{-1}\| \leq \frac{n_1}{(1-\gamma)m}, \tag{40}$$

provided that $\gamma < 1$.

Now we can prove Lemma 7.

Proof of Lemma 7. Let $\mathcal{U} = \mathcal{U}(C)$ and $\mathbf{U} \in \mathbb{R}^{n_1\times k}$ be the orthogonal matrix associated with $\mathcal{U}$ (note that with probability one $\dim(\mathcal{U}) = k$). Fix a column i and let $\boldsymbol{x} = \mathbf{M}^{(i)} = \boldsymbol{a} + \boldsymbol{r}$, where $\boldsymbol{a} \in \mathcal{U}$ and $\boldsymbol{r} \in \mathcal{U}^\perp$. What we want is to bound $\|\boldsymbol{x} - \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top \boldsymbol{x}_\Omega\|_2^2$ in terms of $\|\boldsymbol{r}\|_2^2$.

Write $\boldsymbol{a} = \mathbf{U}\tilde{\boldsymbol{a}}$. By Lemma 10, if m satisfies the condition given in the Lemma then with probability over $1-\delta-\delta''$ we know $(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)$ is invertible and furthermore, $\|(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\|_2 \leq 2n_1/m$. Consequently,

$$\mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top\boldsymbol{a}_\Omega = \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top\mathbf{U}_\Omega\tilde{\boldsymbol{a}} = \mathbf{U}\tilde{\boldsymbol{a}} = \boldsymbol{a}. \tag{41}$$

That is, the subsampled projector preserves components of $\boldsymbol{x}$ in subspace $\mathcal{U}$.

Now let's consider the noise term $\boldsymbol{r}$. By Corollary 1 with probability $\geq 1-\delta$ we can bound the incoherence level of $\boldsymbol{y}$ as $\mu(\boldsymbol{y}) = O(k\mu_0\log(n/\delta))$. The incoherence of subspace $\mathcal{U}$ can also be bounded as $\mu(\mathcal{U}) = O(\mu_0\log(n/\delta))$. Subsequently, given $m = \Omega(k\mu_0\log(n/\delta)\log(n/\delta''))$ we have (with probability $\geq 1-\delta-2\delta''$)

$$\begin{aligned}
& \|\boldsymbol{x} - \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top(\boldsymbol{a}+\boldsymbol{r})\|_2^2 \\
= \; & \|\boldsymbol{a}+\boldsymbol{r} - \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top(\boldsymbol{a}+\boldsymbol{r})\|_2^2 \\
= \; & \|\boldsymbol{r} - \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top\boldsymbol{r}\|_2^2 \\
\leq \; & \|\boldsymbol{r}\|_2^2 + \|(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\|_2^2\|\mathbf{U}_\Omega^\top\boldsymbol{r}\|_2^2 \\
\leq \; & (1+O(1))\|\boldsymbol{r}\|_2^2.
\end{aligned}$$

For the second to last inequality we use the fact that $\boldsymbol{r} \in \mathcal{U}^\perp$. By carefully selecting constants in Eq. (21) we can make

$$\|\boldsymbol{x} - \mathbf{U}(\mathbf{U}_\Omega^\top\mathbf{U}_\Omega)^{-1}\mathbf{U}_\Omega^\top\boldsymbol{x}\|_2^2 \leq 2.5\|\mathcal{P}_{\mathcal{U}^\perp}\boldsymbol{x}\|_2^2. \tag{42}$$

Summing over all n_2 columns yields the desired result. □

Appendix C. Some concentration inequalities

Lemma 11 ([21])**.** *Let $X \sim \chi_d^2$. Then with probability $\geq 1-2\delta$ the following holds:*

$$-2\sqrt{d\log(1/\delta)} \leq X - d \leq 2\sqrt{d\log(1/\delta)} + 2\log(1/\delta). \tag{43}$$

Lemma 12. *Let $X_1, \cdots, X_n \sim \mathcal{N}(0,\sigma^2)$. Then with probability $\geq 1-\delta$ the following holds:*

$$\max_i |X_i| \leq \sigma\sqrt{2\log(2n/\delta)}. \tag{44}$$

Lemma 13 ([23])**.** *Let* $\mathbf{X}$ *be an* $n \times t$ *random matrix with i.i.d. standard Gaussian random entries. If* $t < n$ *then for every* $\epsilon \geq 0$ *with probability* $\geq 1 - 2\exp(-\epsilon^2/2)$ *the following holds:*

$$\sqrt{n} - \sqrt{t} - \epsilon \leq \sigma_{\min}(\mathbf{X}) \leq \sigma_{\max}(\mathbf{X}) \leq \sqrt{n} + \sqrt{t} + \epsilon. \tag{45}$$