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 $M$. By assumption, we have $P_i \geq \frac{1 - \alpha}{1 + \alpha} \|x_i\|_2^2 / \|M\|_F^2$. Applying Theorem 2 from [15] we have that with probability at least $1 - \delta$, there exists an orthonormal set of vectors $y^{(1)}, \cdots, y^{(k)} \in \mathbb{R}^{n_1}$ in span($C$) such that

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

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

$$\|M - \mathcal{P}_C(M)\|_F = \min_{x \in \mathbb{R}^{n_1 \times 2}} \|M - CX\|_F \leq \left\|M - \left(\sum_{j=1}^{k} y^{(j)} y^{(j)\top}\right)\right\|_F. \quad (30)$$

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

$$\Vert \epsilon \Vert F^2 \leq \hat{c}_i \leq (1 + \alpha) \Vert x_i \Vert^2_2$$

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

$$\|M - \mathcal{P}_C(M)\|_F \leq \|M - M_k\|_F + \epsilon\|M\|_F \quad (31)$$

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

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

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

$$\|M - \hat{M}\|_F \leq \epsilon'\|M\|_F. \quad (33)$$

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

$$\|M - CX\|_F = \|M - \mathcal{P}_C(\hat{M})\|_F \leq \|M - \mathcal{P}_C(M)\|_F + \|\mathcal{P}_C(M - \hat{M})\|_F.$$

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 $C$ columns in $\mathbf{R}$. By definition, $\mathcal{U}(C) \subseteq \mathcal{U} \cup \mathcal{R}(C)$, where $\mathcal{U} = \text{span}(A)$ denotes the subspace spanned by columns in the deterministic matrix $A$. Consequently, we have the following bound on $\|\mathcal{P}_{\mathcal{U}(C)}e_i\|$ (assuming each entry in $\mathbf{R}$ follows a zero-mean Gaussian distribution with $\sigma^2$ variance):

$$\|\mathcal{P}_{\mathcal{U}(C)}e_i\|_2^2 \leq \|\mathcal{P}_\mathcal{U}e_i\|_2^2 + \|\mathcal{P}_{\mathcal{U} \cap \mathcal{R}(C)}e_i\|_2^2,$$

3The original theorem concerns random samples of rows; it is essentially the same for random samples of columns.
\[ \begin{align*}
&\leq \|P_U e_i\|^2_2 + \|P_{R(C)} e_i\|^2_2 \\
&\leq \frac{k\mu_0}{n_1} + \|R_C\|^2_2 \|(R_C^\top R_C)^{-1}\|_2^2 \|R_C^\top e_i\|^2_2 \\
&\leq \frac{k\mu_0}{n_1} + \left(\frac{\sqrt{n_1} + \sqrt{s + \epsilon}}{\sqrt{n_1} - \sqrt{s - \epsilon}}\right)^2 \sigma^2(s + 2\sqrt{s \log(2/\delta) + 2 \log(2/\delta)}).
\end{align*} \]

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

\[
\mu(U(C)) = \frac{n_1}{s} \max_{1 \leq i \leq n_1} \|P_{U(C)} e_i\|^2_2 \\
\leq \frac{n_1}{s} \left( \frac{k\mu_0}{n_1} + O\left( \frac{s + \sqrt{s \log(1/\delta) + \log(1/\delta)}}{n_1} \right) \right) = O\left( \frac{k\mu_0 + s + \sqrt{s \log(1/\delta) + \log(1/\delta)}}{s} \right). \tag{34}
\]

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

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

The projected vector \(x' = P_{\hat{U}^\perp} x\) can be written as \(\hat{x} = \hat{a} + \hat{r}\), where \(\hat{a} = P_{\hat{U}^\perp} a\) and \(\hat{r} = P_{\hat{U}^\perp} r\). By definition, \(\hat{a}\) lives in the subspace \(U \cap \hat{U}^\perp\). So it satisfies the incoherence assumption

\[
\mu(\hat{a}) = \frac{n_1 \|\hat{a}\|^2_\infty}{\|\hat{a}\|^2_2} \leq k\mu(\hat{U}) \leq k\mu_0. \tag{35}
\]

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

\[
\mu(\hat{x}) = n_1 \left\| \frac{\hat{x}}{\|\hat{x}\|^2_2} \right\|^2_\infty \leq 3n_1 \left\| \frac{\hat{a}}{\|\hat{a}\|^2_2} + \frac{\hat{r}}{\|\hat{r}\|^2_2} \right\|^2_\infty \\
\leq 3n_1 \|\hat{a}\|^2_\infty + 3n_1 \|\hat{r}\|^2_\infty \\
\leq 3k\mu_0 + \frac{6\sigma^2 n_1 \log(2n_1n_2/\delta)}{\sigma^2 (n_1 - k - s) - 2\sigma^2 \sqrt{(n_1 - k - s) \log(n_2/\delta)}}.
\]

For the second inequality we use the fact that \(\sum_i a_i b_i \leq \sum_i a_i \frac{\alpha_i}{\beta_i}\) 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 \(\sigma^2 n_1/4\); subsequently, we can bound \(\mu(\hat{x})\) as

\[
\mu(\hat{x}) \leq 3k\mu_0 + \frac{24\sigma^2 n_1 \log(2n_1n_2/\delta)}{\sigma^2 n_1} \leq 3k\mu_0 + 24\log(2n_1n_2/\delta). \tag{36}
\]

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 \( y = x + v \), where \( x \in \mathcal{U} \) and \( v \in \mathcal{U}^\perp \). Fix \( \delta' > 0 \), \( m \geq \max\{ \frac{k}{3\delta'} \mu(\mathcal{U}) \log \left( \frac{3\delta'}{\delta} \right) \}, 4\mu(\mathcal{U}) \log(1/\delta') \} \) and let \( \Omega \) be an index set with entries sampled uniformly with replacement with probability \( m/n \). Then with probability at least \( 1 - 4\delta' \):

\[
\frac{m(1-\alpha) - k\mu(\mathcal{U})}{n} \frac{\beta}{3\delta} \|v\|_2^2 \leq \|y_\Omega - P_{\mathcal{U}_\Omega}y_\Omega\|_2^2 \leq \frac{(1+\alpha)m}{n} \|v\|_2^2,
\]

where \( \alpha = \sqrt{2\mu(\mathcal{U}) \log(1/\delta') + \frac{2\mu(\mathcal{U})}{3m} \log(1/\delta')} \), \( \beta = \frac{1 + 2\mu(\mathcal{U})}{\delta} \) 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 \( \dim(S_i) = t \) with probability 1. Let \( y = M^{(i)} \) denote the \( i \)-th column of \( M \) and let \( v = P_{S_i} y \) be the projected vector. We can apply Theorem 4 to bound the estimation error between \( \|v\| \) and \( \|y_\Omega - P_{S_\Omega}y_\Omega\| \).

First, when \( m \) is set as in Eq. (19) it is clear that \( m \geq \frac{k}{3}\mu(\mathcal{U}) \log \left( \frac{3\delta'}{\delta} \right) = \Omega(k\mu_0 \log(n/\delta) \log(k/\delta')) \) and \( m \geq 4\mu(\mathcal{U}) \log(1/\delta') = \Omega(k\mu_0 \log(n/\delta) \log(1/\delta')) \) are satisfied. We next turn to the analysis of \( \alpha, \beta \) and \( \gamma \). More specifically, we want \( \alpha = O(1), \gamma = O(1) \) and \( \frac{4\mu(\mathcal{U})}{3m} \beta = O(1) \).

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

For \( \gamma = O(1) \) 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 \( \beta = O(1) \) implies \( m = O(t\mu(\mathcal{U}) \beta) = O(k\mu_0 \log(n/\delta) \log(1/\delta')) \). By carefully selecting constants we have \( \beta \leq 0.2 \). Finlay, combining bounds on \( \alpha, \beta \) and \( \gamma \) 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 \( A \in \mathbb{R}^{m \times n} \) with \( m \leq n \). Suppose \( \sigma_1, \ldots, \sigma_m \) are singular values of \( A \). Then

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

Now we are ready to prove Lemma 6.

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

\[
\hat{p}_C \leq \sum_{\tau \in \Pi_k} \left( \frac{5}{2} \right)^k \frac{\|M^{(\tau(i_1))}\|_F^2 d(M^{(\tau(i_1))}, \mathcal{H}_{\tau,1})^2 \cdots d(M^{(\tau(i_k))}, \mathcal{H}_{\tau,k-1})^2}{\|M\|_F^2 \|M - M_k\|_F^2 \cdots \|M - M_{k-1}\|_F^2} \leq 2.5^k \cdot \frac{\sum_{\tau \in \Pi_k} \|M^{(\tau(i_1))}\|_F^2 d(M^{(\tau(i_1))}, \mathcal{H}_{\tau,1})^2 \cdots d(M^{(\tau(i_k))}, \mathcal{H}_{\tau,k-1})^2}{\|M\|_F^2 \|M - M_k\|_F^2 \cdots \|M - M_{k-1}\|_F^2} = 2.5^k \cdot \frac{(k!)^2 \text{vol}(\Delta(C))^2}{\sum_{1 \leq i_1 < \cdots < i_k \leq n_1} \sigma_{i_1}^2 \sigma_{i_2}^2 \cdots \sigma_{i_k}^2} \leq 2.5^k \cdot \frac{k! \text{vol}(\Delta(C))^2}{\text{vol}(\Delta(T))^2} = 2.5^k k! \hat{p}_C.
\]
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 \( U \in \mathbb{R}^{n \times k} \) 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 \( y \in \mathbb{R}^n \), with probability at least \( 1 - \delta \):

\[
\|U_\Omega^T y_\Omega\|_2^2 \leq \beta \frac{m n k \mu(U)}{n_1} \|y\|_2^2,
\]

where \( \beta \) 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

\[
\|(U_\Omega^T U_\Omega)^{-1}\| \leq \frac{n_1}{(1 - \gamma)m},
\]

provided that \( \gamma < 1 \).

Now we can prove Lemma 7.

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

Write \( a = U\tilde{a} \). By Lemma 10, if \( m \) satisfies the condition given in the Lemma then with probability over \( 1 - \delta - 3\delta \) we know \( (U_\Omega^T U_\Omega) \) is invertible and furthermore, \( \|(U_\Omega^T U_\Omega)^{-1}\|_2 \leq 2n_1/m \). Consequently,

\[
U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T a_\Omega = U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T U_\Omega \tilde{a} = U\tilde{a} = a.
\]

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

Now let’s consider the noise term \( r \). By Corollary 1 with probability \( \geq 1 - \delta \) we can bound the incoherence level of \( y \) as \( \mu(y) = O(k \mu_0 \log(n/\delta)) \). 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 \( \geq 1 - \delta - 2\delta'' \))

\[
\|x - U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T (a + r)\|_2^2 \\
= \|a + r - U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T (a + r)\|_2^2 \\
= \|r - U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T r\|_2^2 \\
\leq \|r\|_2^2 + \|(U_\Omega^T U_\Omega)^{-1}\|_2^2 \|U_\Omega^T r\|_2^2 \\
\leq (1 + O(1))\|r\|_2^2.
\]

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

\[
\|x - U(U_\Omega^T U_\Omega)^{-1} U_\Omega^T x\|_2^2 \leq 2.5\|P_{\mathcal{U}^\perp} x\|_2^2.
\]

Summing over all \( n_2 \) columns yields the desired result.

**Appendix C. Some concentration inequalities**

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

\[
-2\sqrt{d \log(1/\delta)} \leq X - d \leq 2\sqrt{d \log(1/\delta)} + 2 \log(1/\delta).
\]

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

\[
\max_i |X_i| \leq \sigma \sqrt{2 \log(2n/\delta)}.
\]
Lemma 13 ([23]). Let $X$ 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:

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