A Analysis of Randomized Approximation Algorithms

In this section, we will establish several key properties of column projection and random column sampling that will aid us in deriving DFC recovery guarantees. Hereafter, $\epsilon \in (0,1]$ represents a prescribed error tolerance, and $\delta \in (0,1]$ designates a target failure probability.

Our first theorem asserts that, with high probability, column projection produces an approximation nearly as good as a given rank-r target by sampling a number of columns proportional to the coherence and $r \log r$.

Theorem 6. *Given a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$ and a rank-r approximation $\mathbf{L} \in \mathbb{R}^{m\times n}$, choose $l \geq cr\mu_0(\mathbf{V}_L)\log(n)\log(1/\delta)/\epsilon^2$, where c is a fixed positive constant, and let $\mathbf{C} \in \mathbb{R}^{m\times l}$ be a matrix of l columns of $\mathbf{M}$ sampled uniformly without replacement. Then,*

$$\|\mathbf{M} - \mathbf{L}^{proj}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F$$

with probability at least $1-\delta$.

The proof of Thm. 6 builds upon the randomized ℓ_2 regression work of [6] and will be given in Sec. A.1.

Our next lemma bounds the μ_0 and μ_1-coherence of a uniformly sampled submatrix in terms of the coherence of the full matrix. These properties will allow for accurate submatrix completion or outlier removal using standard MC and RMF algorithms. Its proof is given in Sec. A.2.

Lemma 7. *Let $\mathbf{L} \in \mathbb{R}^{m\times n}$ be a rank-r matrix and $\mathbf{L}_C \in \mathbb{R}^{m\times l}$ be a matrix of l columns of $\mathbf{L}$ sampled uniformly without replacement. If $l \geq cr\mu_0(\mathbf{V}_L)\log(n)\log(1/\delta)/\epsilon^2$, where c is a fixed positive constant defined in Thm. 6, then*

i) $\operatorname{rank}(\mathbf{L}_C) = \operatorname{rank}(\mathbf{L})$

ii) $\mu_0(\mathbf{U}_{L_C}) = \mu_0(\mathbf{U}_L)$

iii) $\mu_0(\mathbf{V}_{L_C}) \leq \dfrac{\mu_0(\mathbf{V}_L)}{1-\epsilon/2}$

iv) $\mu_1^2(\mathbf{L}_C) \leq \dfrac{r\mu_0(\mathbf{U}_L)\mu_0(\mathbf{V}_L)}{1-\epsilon/2}$

all hold jointly with probability at least $1-\delta/n$.

A.1 Proof of Theorem 6

We now give a proof of Thm. 6. While the results of this section are stated in terms of *i.i.d.* with-replacement sampling of columns and rows, a simple argument due to [10, Sec. 6] implies the same conclusions when columns and rows are sampled without replacement.

Our proof of Thm. 6 will require a strengthened version of the randomized ℓ_2 regression work of [6, Thm. 5]. The proof of Thm. 5 of [6] relies heavily on the fact that $\|\mathbf{AB} - \mathbf{GH}\|_F \leq \frac{\epsilon}{2}\|\mathbf{A}\|_F\|\mathbf{B}\|_F$ with probability at least 0.9, when $\mathbf{G}$ and $\mathbf{H}$ contain sufficiently many rescaled columns and rows of $\mathbf{A}$ and $\mathbf{B}$, sampled according to a particular non-uniform probability distribution. A result of [11], modified to allow for slack in the probabilities, shows that a related claim holds with probability $1-\delta$ for arbitrary $\delta \in (0,1]$.

Lemma 8 (Sec. 3.4.3 of [11])**.** *Given matrices $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$ with $r \geq \max(\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B}))$, an error tolerance $\epsilon \in (0,1]$, and a failure probability $\delta \in (0,1]$, define probabilities p_j satisfying*

$$p_j \geq \frac{\beta}{Z}\|\mathbf{A}_{(j)}\|\|\mathbf{B}_{(j)}\|, \quad Z = \sum_j \|\mathbf{A}_{(j)}\|\|\mathbf{B}_{(j)}\|, \quad \textit{and} \quad \textstyle\sum_{j=1}^k p_j = 1 \tag{1}$$

for some $\beta \in (0,1]$. Let $\mathbf{G} \in \mathbb{R}^{m\times l}$ be a column submatrix of $\mathbf{A}$ in which exactly $l \geq 48r\log(4r/(\beta\delta))/(\beta\epsilon^2)$ columns are selected in i.i.d. trials in which the j-th column is chosen *with probability p_j, and let $\mathbf{H} \in \mathbb{R}^{l \times n}$ be a matrix containing the corresponding rows of $\mathbf{B}$. Further, let $\mathbf{D} \in \mathbb{R}^{l \times l}$ be a diagonal rescaling matrix with entry $\mathbf{D}_{tt} = 1/\sqrt{lp_j}$ whenever the j-th column of $\mathbf{A}$ is selected on the t-th sampling trial, for $t = 1, \dots, l$. Then, with probability at least $1-\delta$,*

$$\|\mathbf{AB} - \mathbf{GDDH}\|_2 \le \frac{\epsilon}{2}\|\mathbf{A}\|_2\|\mathbf{B}\|_2.$$

Using Lemma 8, we now establish a stronger version of Lemma 1 of [6]. For a given $\beta \in (0, 1]$ and $\mathbf{L} \in \mathbb{R}^{m \times n}$ with rank r, we first define column sampling probabilities p_j satisfying

$$p_j \ge \frac{\beta}{r}\|(\mathbf{V}_L)_{(j)}\|^2 \quad \text{and} \quad \textstyle\sum_{j=1}^{n} p_j = 1. \tag{2}$$

We further let $\mathbf{S} \in \mathbb{R}^{n \times l}$ be a random binary matrix with independent columns, where a single 1 appears in each column, and $\mathbf{S}_{jt} = 1$ with probability p_j for each $t \in \{1, \dots, l\}$. Moreover, let $\mathbf{D} \in \mathbb{R}^{l \times l}$ be a diagonal rescaling matrix with entry $\mathbf{D}_{tt} = 1/\sqrt{lp_j}$ whenever $\mathbf{S}_{jt} = 1$. Postmultiplication by $\mathbf{S}$ is equivalent to selecting l random columns of a matrix, independently and with replacement. Under this notation, we establish the following lemma:

Lemma 9. *Let $\epsilon \in (0, 1]$, and define $\mathbf{V}_l^\top = \mathbf{V}_L^\top \mathbf{S}$ and $\Gamma = (\mathbf{V}_l^\top \mathbf{D})^+ - (\mathbf{V}_l^\top \mathbf{D})^\top$. If $l \ge 48r\log(4r/(\beta\delta))/(\beta\epsilon^2)$ for $\delta \in (0, 1]$ then with probability at least $1-\delta$:*

$$\begin{aligned}
\operatorname{rank}(\mathbf{V}_l) = \operatorname{rank}(\mathbf{V}_L) &= \operatorname{rank}(\mathbf{L}) \\
\|\Gamma\|_2 &= \left\|\mathbf{\Sigma}_{V_l^\top D}^{-1} - \mathbf{\Sigma}_{V_l^\top D}\right\|_2 \\
(\mathbf{LSD})^+ &= (\mathbf{V}_l^\top \mathbf{D})^+ \mathbf{\Sigma}_L^{-1}\mathbf{U}_L^\top \\
\left\|\mathbf{\Sigma}_{V_l^\top D}^{-1} - \mathbf{\Sigma}_{V_l^\top D}\right\|_2 &\le \epsilon/\sqrt{2}.
\end{aligned}$$

Proof By Lemma 8, for all $1 \le i \le r$,

$$\begin{aligned}
|1 - \sigma_i^2(\mathbf{V}_l^\top \mathbf{D})| &= |\sigma_i(\mathbf{V}_L^\top \mathbf{V}_L) - \sigma_i(\mathbf{V}_l^\top \mathbf{DDV}_l)| \\
&\le \|\mathbf{V}_L^\top \mathbf{V}_L - \mathbf{V}_L^\top \mathbf{SDDS}^\top \mathbf{V}_L\|_2 \\
&\le \epsilon/2\|\mathbf{V}_L^\top\|_2\|\mathbf{V}_L\|_2 = \epsilon/2,
\end{aligned}$$

where $\sigma_i(\cdot)$ is the i-th largest singular value of a given matrix. Since $\epsilon/2 \le 1/2$, each singular value of $\mathbf{V}_l$ is positive, and so $\operatorname{rank}(\mathbf{V}_l) = \operatorname{rank}(\mathbf{V}_L) = \operatorname{rank}(\mathbf{L})$. The remainder of the proof is identical to that of Lemma 1 of [6]. □

Lemma 9 immediately yields improved sampling complexity for the randomized ℓ_2 regression of [6]:

Proposition 10. *Suppose $\mathbf{B} \in \mathbb{R}^{p \times n}$ and $\epsilon \in (0, 1]$. If $l \ge 3200r\log(4r/(\beta\delta))/(\beta\epsilon^2)$ for $\delta \in (0, 1]$, then with probability at least $1 - \delta - 0.2$:*

$$\|\mathbf{B} - \mathbf{BSD}(\mathbf{LSD})^+\mathbf{L}\|_F \le (1+\epsilon)\|\mathbf{B} - \mathbf{BL}^+\mathbf{L}\|_F$$

.

Proof The proof is identical to that of Thm. 5 of [6] once Lemma 9 is substituted for Lemma 1 of [6]. □

Prop. 10 allows us to prove a generalization of Thm. 1 of [6] that will provide recovery guarantees for column projection relative to an arbitrary low-rank approximation.

Theorem 11. *Suppose $\mathbf{M} \in \mathbb{R}^{m \times n}$ and $\epsilon \in (0, 1]$. If $l \ge 3200r\log(4r/(\beta\delta))/(\beta\epsilon^2)$ for $\delta \in (0, 1]$, then with probability at least $1 - \delta - 0.2$,*

$$\|\mathbf{M} - \mathbf{MSD}(\mathbf{MSD})^+\mathbf{M}\|_F \le (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F.$$

Proof Since $(\mathbf{MSD})^+\mathbf{M}$ minimizes $\|\mathbf{M} - \mathbf{MSDX}\|_F$ over all $\mathbf{X} \in \mathbb{R}^{l \times n}$, it follows that

$$\|\mathbf{M} - \mathbf{MSD}(\mathbf{MSD})^+\mathbf{M}\|_F \le \|\mathbf{M} - \mathbf{MSD}(\mathbf{LSD})^+\mathbf{L}\|_F.$$

Further, by Prop. 10,

$$\|\mathbf{M} - \mathbf{MSD}(\mathbf{LSD})^{+}\mathbf{L}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{ML}^{+}\mathbf{L}\|_F.$$

Finally, noting that $\mathbf{ML}^{+}$ minimizes $\|\mathbf{M} - \mathbf{YL}\|_F$ over all $\mathbf{Y} \in \mathbb{R}^{m\times m}$ yields

$$\|\mathbf{M} - \mathbf{ML}^{+}\mathbf{L}\|_F \leq \|\mathbf{M} - \mathbf{L}\|_F,$$

establishing the result. □

By sampling an additional factor of O($\log(1/\delta)$) columns, we may boost the success probability of Thm. 11 to an arbitrarily large value of $1-\delta$.

Corollary 12. *Suppose* $\mathbf{M} \in \mathbb{R}^{m\times n}$ *and* $\epsilon \in (0,1]$. *If*

$$l \geq 3200r\log(16r/\beta)(\log(\delta)/\log(0.45))/(\beta\epsilon^2)$$

for $\delta \in (0,1]$, *then with probability at least* $1-\delta$,

$$\|\mathbf{M} - \mathbf{MSD}(\mathbf{MSD})^{+}\mathbf{M}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F.$$

Proof Partition the columns of $\mathbf{S}$ into $b = \log(\delta)/\log(0.45)$ submatrices, $\mathbf{S} = [\mathbf{S}_1, \cdots, \mathbf{S}_b]$, each with $a = l/b$ columns.[6] Associate with each submatrix $\mathbf{S}_i$ a diagonal matrix $\mathbf{D}_i$ with entry $\mathbf{D}_{itt} = 1/\sqrt{ap_j}$ whenever $\mathbf{S}_{ijt} = 1$. Since

$$a \geq 3200r\log(4r/(0.25\beta))/(\beta\epsilon^2),$$

we may apply Thm. 11 independently for each i to yield

$$\|\mathbf{M} - \mathbf{MS}_i\mathbf{D}_i(\mathbf{MS}_i\mathbf{D}_i)^{+}\mathbf{M}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F \tag{3}$$

with probability at least 0.55. Moreover, for any i,

$$\begin{aligned}\|\mathbf{M} - \mathbf{P}_{MS_iD_i}\mathbf{M}\|_F^2 &= \|\mathbf{P}^{\perp}_{MS_iD_i}\mathbf{M}\|_F^2 = \|\mathbf{P}_{MSD}\mathbf{P}^{\perp}_{MS_iD_i}\mathbf{M}\|_F^2 + \|\mathbf{P}^{\perp}_{MSD}\mathbf{P}^{\perp}_{MS_iD_i}\mathbf{M}\|_F^2 \\ &\geq \|\mathbf{P}^{\perp}_{MSD}\mathbf{P}^{\perp}_{MS_iD_i}\mathbf{M}\|_F^2 = \|\mathbf{P}^{\perp}_{MSD}\mathbf{M}\|_F^2 = \|\mathbf{M} - \mathbf{P}_{MSD}\mathbf{M}\|_F^2,\end{aligned}$$

where the penultimate equality holds since $\mathbf{MS}_i\mathbf{D}_i$ is contained in the column space of $\mathbf{MSD}$ for all i. Hence, if

$$\|\mathbf{M} - \mathbf{P}_{MSD}\mathbf{M}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F,$$

fails to hold, then, for each i, Eq. 3 also fails to hold. The desired conclusion therefore must hold with probability at least $1 - 0.45^b = 1-\delta$. □

A typical application of Cor. 12 would involve performing a truncated SVD of $\mathbf{M}$ to obtain the *statistical leverage scores*, $\|(\mathbf{V}_L)_{(j)}\|^2$, used to compute the column sampling probabilities of Eq. (2). Here, we will take advantage of the slack term, β, allowed in the sampling probabilities of Eq. (2) to show that uniform column sampling gives rise to the same recovery guarantees for column projection approximations when $\mathbf{L}$ is sufficiently incoherent.

Fix $c = 16000/\log(1/0.45)$. To prove Thm. 6, we first notice that for $n > 1$,

$$16000\log(n) = 3200\log(n^5) \geq 3200\log(16n).$$

Hence

$$\begin{aligned}l &\geq 3200r\mu_0(\mathbf{V}_L)\log(16n)(\log(\delta)/\log(0.45))/\epsilon^2 \\ &\geq 3200r\mu_0(\mathbf{V}_L)\log(16r\mu_0(\mathbf{V}_L))(\log(\delta)/\log(0.45))/\epsilon^2 \\ &\geq 3200r\log(16r/\beta)(\log(\delta)/\log(0.45))/(\beta\epsilon^2)\end{aligned}$$

whenever $\beta \geq 1/\mu_0(\mathbf{V}_L)$. Thus, we may apply Cor. 12 with $\beta = 1/\mu_0(\mathbf{V}_L) \in (0,1]$ and $p_j = 1/n$ by noting that

$$\frac{\beta}{r}\|(\mathbf{V}_L)_{(j)}\|^2 \leq \frac{\beta}{r}\frac{r}{n}\mu_0(\mathbf{V}_L) = \frac{1}{n} = p_j$$

for all j, by the definition of $\mu_0(\mathbf{V}_L)$. By our choice of probabilities, $\mathbf{D} = \mathbf{I}\sqrt{n/l}$, and hence

$$\|\mathbf{M} - \mathbf{CC}^{+}\mathbf{M}\|_F = \|\mathbf{M} - \mathbf{CD}(\mathbf{CD})^{+}\mathbf{M}\|_F \leq (1+\epsilon)\|\mathbf{M} - \mathbf{L}\|_F$$

with probability at least $1-\delta$, as desired.

[6] For simplicity, we assume that b divides l evenly.

A.2 Proof of Lemma 7

Since for all $n > 1$,

$$c\log(n)\log(1/\delta) = (c/4)\log(n^4)\log(1/\delta) \geq 48\log(4n^2/\delta) \geq 48\log(4r\mu_0(\mathbf{V}_L)/(\delta/n))$$

as $n \geq r\mu_0(\mathbf{V}_L)$, claim i follows immediately from Lemma 9 with $\beta = 1/\mu_0(\mathbf{V}_L)$, $p_j = 1/n$ for all j, and $\mathbf{D} = \mathbf{I}\sqrt{n/l}$. When $\operatorname{rank}(\mathbf{L}_C) = \operatorname{rank}(\mathbf{L})$, Lemma 1 of [18] implies that $\mathbf{P}_{U_{L_C}} = \mathbf{P}_{U_L}$, which in turn implies claim ii.

To prove claim iii given the conclusions of Lemma 9, assume, without loss of generality, that $\mathbf{V}_l$ consists of the first l rows of $\mathbf{V}_L$. Then if $\mathbf{L}_C = \mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top$ has $\operatorname{rank}(\mathbf{L}_C) = \operatorname{rank}(\mathbf{L}) = r$, the matrix $\mathbf{V}_l$ must have full column rank. Thus we can write

$$\begin{aligned}
\mathbf{L}_C^+\mathbf{L}_C &= (\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top)^+\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top \\
&= (\mathbf{\Sigma}_L\mathbf{V}_l^\top)^+\mathbf{U}_L^+\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top \\
&= (\mathbf{\Sigma}_L\mathbf{V}_l^\top)^+\mathbf{\Sigma}_L\mathbf{V}_l^\top \\
&= (\mathbf{V}_l^\top)^+\mathbf{\Sigma}_L^+\mathbf{\Sigma}_L\mathbf{V}_l^\top \\
&= (\mathbf{V}_l^\top)^+\mathbf{V}_l^\top \\
&= \mathbf{V}_l(\mathbf{V}_l^\top\mathbf{V}_l)^{-1}\mathbf{V}_l^\top,
\end{aligned}$$

where the second and third equalities follow from $\mathbf{U}_L$ having orthonormal columns, the fourth and fifth result from $\mathbf{\Sigma}_L$ having full rank and $\mathbf{V}_l$ having full column rank, and the sixth follows from $\mathbf{V}_l^\top$ having full row rank.

Now, denote the right singular vectors of $\mathbf{L}_C$ by $\mathbf{V}_{L_C} \in \mathbb{R}^{l\times r}$. Observe that $\mathbf{P}_{V_{L_C}} = \mathbf{V}_{L_C}\mathbf{V}_{L_C}^\top = \mathbf{L}_C^+\mathbf{L}_C$, and define $\mathbf{e}_{i,l}$ as the ith column of $\mathbf{I}_l$ and $\mathbf{e}_{i,n}$ as the ith column of $\mathbf{I}_n$. Then we have,

$$\begin{aligned}
\mu_0(\mathbf{V}_{L_C}) &= \frac{l}{r}\max_{1\leq i\leq l}\|\mathbf{P}_{V_{L_C}}\mathbf{e}_{i,l}\|^2 \\
&= \frac{l}{r}\max_{1\leq i\leq l}\mathbf{e}_{i,l}^\top\mathbf{L}_C^+\mathbf{L}_C\mathbf{e}_{i,l} \\
&= \frac{l}{r}\max_{1\leq i\leq l}\mathbf{e}_{i,l}^\top(\mathbf{V}_l^\top)^+\mathbf{V}_l^\top\mathbf{e}_{i,l} \\
&= \frac{l}{r}\max_{1\leq i\leq l}\mathbf{e}_{i,l}^\top\mathbf{V}_l(\mathbf{V}_l^\top\mathbf{V}_l)^{-1}\mathbf{V}_l^\top\mathbf{e}_{i,l} \\
&= \frac{l}{r}\max_{1\leq i\leq l}\mathbf{e}_{i,n}^\top\mathbf{V}_L(\mathbf{V}_l^\top\mathbf{V}_l)^{-1}\mathbf{V}_L^\top\mathbf{e}_{i,n},
\end{aligned}$$

where the final equality follows from $\mathbf{V}_l^\top\mathbf{e}_{i,l} = \mathbf{V}_L^\top\mathbf{e}_{i,n}$ for all $1 \leq i \leq l$.

Now, defining $\mathbf{Q} = \mathbf{V}_l^\top\mathbf{V}_l$ we have

$$\begin{aligned}
\mu_0(\mathbf{V}_{L_C}) &= \frac{l}{r}\max_{1\leq i\leq l}\mathbf{e}_{i,n}^\top\mathbf{V}_L\mathbf{Q}^{-1}\mathbf{V}_L^\top\mathbf{e}_{i,n} \\
&= \frac{l}{r}\max_{1\leq i\leq l}\operatorname{Tr}\left[\mathbf{e}_{i,n}^\top\mathbf{V}_L\mathbf{Q}^{-1}\mathbf{V}_L^\top\mathbf{e}_{i,n}\right] \\
&= \frac{l}{r}\max_{1\leq i\leq l}\operatorname{Tr}\left[\mathbf{Q}^{-1}\mathbf{V}_L^\top\mathbf{e}_{i,n}\mathbf{e}_{i,n}^\top\mathbf{V}_L\right] \\
&\leq \frac{l}{r}\|\mathbf{Q}^{-1}\|_2\max_{1\leq i\leq l}\left\|\mathbf{V}_L^\top\mathbf{e}_{i,n}\mathbf{e}_{i,n}^\top\mathbf{V}_L\right\|_*,
\end{aligned}$$

by Hölder's inequality for Schatten p-norms. Since $\mathbf{V}_L^\top \mathbf{e}_{i,n}\mathbf{e}_{i,n}^\top \mathbf{V}_L$ has rank one, we can explicitly compute its trace norm as $\|\mathbf{V}_L^\top \mathbf{e}_{i,n}\|^2 = \|\mathbf{P}_{V_L}\mathbf{e}_{i,n}\|^2$. Hence,

$$\begin{aligned}
\mu_0(\mathbf{V}_{L_C}) &\leq \frac{l}{r}\|\mathbf{Q}^{-1}\|_2 \max_{1\leq i\leq l} \|\mathbf{P}_{V_L}\mathbf{e}_{i,n}\|^2 \\
&\leq \frac{l}{r}\frac{r}{n}\|\mathbf{Q}^{-1}\|_2 \left(\frac{n}{r}\max_{1\leq i\leq n} \|\mathbf{P}_{V_L}\mathbf{e}_{i,n}\|^2\right) \\
&= \frac{l}{n}\|\mathbf{Q}^{-1}\|_2 \mu_0(\mathbf{V}_L)\,,
\end{aligned}$$

by the definition of μ_0-coherence. The proof of Lemma 9 established that the smallest singular value of $\frac{n}{l}\mathbf{Q} = \mathbf{V}_l^\top \mathbf{D}\mathbf{D}\mathbf{V}_l$ is lower bounded by $1-\frac{\epsilon}{2}$ and hence $\|\mathbf{Q}^{-1}\|_2 \leq \frac{n}{l(1-\epsilon/2)}$. Thus, we conclude that $\mu_0(\mathbf{V}_{L_C}) \leq \mu_0(\mathbf{V}_L)/(1-\epsilon/2)$.

To prove claim iv under Lemma 9, note that $\mathbf{P}_{U_L} = \mathbf{P}_{U_{L_C}}$ implies $\mathbf{U}_L\mathbf{U}_L^\top \mathbf{U}_{L_C} = \mathbf{U}_{L_C}$. We thus observe that,

$$\begin{aligned}
\mathbf{U}_{L_C}\mathbf{V}_{L_C}^\top &= \mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top \mathbf{L}_C \\
&= \mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top \mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top \\
&= \mathbf{U}_L\mathbf{U}_L^\top \mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top \mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top\,.
\end{aligned}$$

Letting $\mathbf{B} = \mathbf{U}_L^\top \mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top \mathbf{U}_L\mathbf{\Sigma}_L$, we have

$$\begin{aligned}
\mu_1(\mathbf{L}_C) &= \sqrt{\frac{ml}{r}} \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} |\mathbf{e}_{i,m}^\top \mathbf{U}_{L_C}\mathbf{V}_{L_C}^\top \mathbf{e}_{j,l}| \\
&= \sqrt{\frac{ml}{r}} \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} |\mathbf{e}_{i,m}^\top \mathbf{U}_L\mathbf{B}\mathbf{V}_l^\top \mathbf{e}_{j,l}| \\
&= \sqrt{\frac{ml}{r}} \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} |\mathbf{e}_{i,m}^\top \mathbf{U}_L\mathbf{B}\mathbf{V}_L^\top \mathbf{e}_{j,n}| \\
&= \sqrt{\frac{ml}{r}} \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} |\mathrm{Tr}\big[\mathbf{e}_{i,m}^\top \mathbf{U}_L\mathbf{B}\mathbf{V}_L^\top \mathbf{e}_{j,n}\big]| \\
&= \sqrt{\frac{ml}{r}} \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} |\mathrm{Tr}\big[\mathbf{B}\mathbf{V}_L^\top \mathbf{e}_{j,n}\mathbf{e}_{i,m}^\top \mathbf{U}_L\big]| \\
&\leq \sqrt{\frac{ml}{r}}\|\mathbf{B}\|_2 \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} \|\mathbf{V}_L^\top \mathbf{e}_{j,n}\mathbf{e}_{i,m}^\top \mathbf{U}_L\|_*\,,
\end{aligned}$$

by Hölder's inequality for Schatten p-norms. Since $\mathbf{V}_L^\top \mathbf{e}_{j,n}\mathbf{e}_{i,m}^\top \mathbf{U}_L$ has rank one, we can explicitly compute its trace norm as $\|\mathbf{U}_L^\top \mathbf{e}_{i,m}\|\|\mathbf{V}_L^\top \mathbf{e}_{j,n}\| = \|\mathbf{P}_{U_L}\mathbf{e}_{i,m}\|\|\mathbf{P}_{V_L}\mathbf{e}_{j,n}\|$. Hence,

$$\begin{aligned}
\mu_1(\mathbf{L}_C) &\leq \sqrt{\frac{ml}{r}}\|\mathbf{B}\|_2 \max_{\substack{1\leq i\leq m\\1\leq j\leq l}} \|\mathbf{P}_{U_L}\mathbf{e}_{i,m}\|\|\mathbf{P}_{V_L}\mathbf{e}_{j,n}\| \\
&= \sqrt{\frac{mlr^2}{mnr}}\|\mathbf{B}\|_2 \left(\sqrt{\frac{m}{r}}\max_{1\leq i\leq m}\|\mathbf{P}_{U_L}\mathbf{e}_{i,m}\|\right)\left(\sqrt{\frac{n}{r}}\max_{1\leq j\leq l}\|\mathbf{P}_{V_L}\mathbf{e}_{j,n}\|\right) \\
&\leq \sqrt{\frac{mlr^2}{mnr}}\|\mathbf{B}\|_2 \left(\sqrt{\frac{m}{r}}\max_{1\leq i\leq m}\|\mathbf{P}_{U_L}\mathbf{e}_{i,m}\|\right)\left(\sqrt{\frac{n}{r}}\max_{1\leq j\leq n}\|\mathbf{P}_{V_L}\mathbf{e}_{j,n}\|\right) \\
&= \sqrt{\frac{lr}{n}}\|\mathbf{B}\|_2\sqrt{\mu_0(\mathbf{U}_L)\mu_0(\mathbf{V}_L)}\,,
\end{aligned}$$

by the definitition of μ_0-coherence.

Next, we notice that

$$\begin{aligned}
\mathbf{B}^\top\mathbf{B} &= \mathbf{\Sigma}_L\mathbf{U}_L^\top\mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top\mathbf{U}_L\mathbf{U}_L^\top\mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top\mathbf{U}_L\mathbf{\Sigma}_L \\
&= \mathbf{\Sigma}_L\mathbf{U}_L^\top\mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top\mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-1}\mathbf{U}_{L_C}^\top\mathbf{U}_L\mathbf{\Sigma}_L \\
&= \mathbf{\Sigma}_L\mathbf{U}_L^\top\mathbf{U}_{L_C}\mathbf{\Sigma}_{L_C}^{-2}\mathbf{U}_{L_C}^\top\mathbf{U}_L\mathbf{\Sigma}_L \\
&= \mathbf{\Sigma}_L\mathbf{U}_L^\top(\mathbf{L}_C\mathbf{L}_C^\top)^+\mathbf{U}_L\mathbf{\Sigma}_L \\
&= \mathbf{\Sigma}_L\mathbf{U}_L^\top(\mathbf{U}_L\mathbf{\Sigma}_L\mathbf{V}_l^\top\mathbf{V}_l\mathbf{\Sigma}_L\mathbf{U}_L^\top)^+\mathbf{U}_L\mathbf{\Sigma}_L \\
&= \mathbf{\Sigma}_L\mathbf{U}_L^\top\mathbf{U}_L\mathbf{\Sigma}_L^{-1}(\mathbf{V}_l^\top\mathbf{V}_l)^{-1}\mathbf{\Sigma}_L^{-1}\mathbf{U}_L^\top\mathbf{U}_L\mathbf{\Sigma}_L \\
&= (\mathbf{V}_l^\top\mathbf{V}_l)^{-1},
\end{aligned}$$

where the penultimate equality follows from $\mathbf{U}_L$ having orthogonal columns and $\mathbf{\Sigma}_L\mathbf{V}_l^\top\mathbf{V}_l\mathbf{\Sigma}_L$ having full rank. The proof of Lemma 9 established that the smallest singular value of $\frac{n}{l}\mathbf{V}_l^\top\mathbf{V}_l = \mathbf{V}_l^\top\mathbf{D}\mathbf{D}\mathbf{V}_l$ is lower bounded by $1-\epsilon/2$ and hence that $\|\mathbf{B}^\top\mathbf{B}\|_2 \le \frac{n}{l(1-\epsilon/2)}$ and $\|\mathbf{B}\|_2 \le \sqrt{\frac{n}{l(1-\epsilon/2)}}$. Thus, we conclude that $\mu_1(\mathbf{L}_C) \le \sqrt{r\mu_0(\mathbf{U}_L)\mu_0(\mathbf{V}_L)}/\sqrt{1-\epsilon/2}$.

B Proof of Theorem 3

Let $\mathbf{L}_0 = [\mathbf{C}_{0,1},\ldots,\mathbf{C}_{0,t}]$ and $\hat{\mathbf{L}} = [\hat{\mathbf{C}}_1,\ldots,\hat{\mathbf{C}}_t]$. Define G as the event $\|\mathbf{L}_0 - \hat{\mathbf{L}}^{proj}\|_F \le (2+\epsilon)c_e\sqrt{mn}\Delta$, H as the event $\|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{proj}\|_F \le (1+\epsilon)\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F$, and B_i as the event $\|\mathbf{C}_{0,i} - \hat{\mathbf{C}}_i\|_F \le c_e\sqrt{ml}\Delta$, for each $i \in \{1,\ldots,t\}$. When H holds, we have that

$$\|\mathbf{L}_0 - \hat{\mathbf{L}}^{proj}\|_F \le \|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F + \|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{proj}\|_F \le (2+\epsilon)\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F,$$

by the triangle inequality, and hence

$$\mathbf{P}(G) \ge \mathbf{P}(\textstyle\bigcap_i B_i \cap H \cap \bigcap_i A(\mathbf{C}_{0,i})) = \mathbf{P}(\bigcap_i B_i \mid H \cap \bigcap_i A(\mathbf{C}_{0,i}))\mathbf{P}(H \cap \bigcap_i A(\mathbf{C}_{0,i})).$$

Our choice of l, with a factor of $\log(2/\delta)$, implies that each $A(\mathbf{C}_{0,i})$ holds with probability at least $1-\delta/(2n)$ by Lemma 7, while H holds with probability at least $1-\delta/2$ by Thm. 6. Hence, by the union bound,

$$\mathbf{P}(H \cap \textstyle\bigcap_i A(\mathbf{C}_{0,i})) \ge 1 - \mathbf{P}(H^c) - \sum_i \mathbf{P}(A(\mathbf{C}_{0,i})^c) \ge 1 - \delta/2 - t\delta/(2n) \ge 1-\delta.$$

Further, by a union bound and our base MF assumption,

$$\mathbf{P}(\textstyle\bigcap_i B_i \mid H \cap \bigcap_i A(\mathbf{C}_{0,i})) \ge 1 - \sum_i \mathbf{P}(B_i^c \mid A(\mathbf{C}_{0,i})) \ge 1 - t\delta_C$$

yielding the desired bound on $\mathbf{P}(G)$.

To prove the second statement, we redefine $\hat{\mathbf{L}}$ and write it in block notation as:

$$\hat{\mathbf{L}} = \begin{bmatrix}\hat{\mathbf{C}}_1 & \hat{\mathbf{R}}_2 \\ \hat{\mathbf{C}}_2 & \mathbf{L}_{0,22}\end{bmatrix}, \quad \text{where} \quad \hat{\mathbf{C}} = \begin{bmatrix}\hat{\mathbf{C}}_1 \\ \hat{\mathbf{C}}_2\end{bmatrix}, \quad \hat{\mathbf{R}} = \begin{bmatrix}\hat{\mathbf{R}}_1 & \hat{\mathbf{R}}_2\end{bmatrix}$$

and $\mathbf{L}_{0,22} \in \mathbb{R}^{(m-d)\times(n-l)}$ is the bottom right submatrix of $\mathbf{L}_0$. We further define J as the event $\|\mathbf{L}_0 - \hat{\mathbf{L}}^{nys}\|_F \le (2+3\epsilon)c_e\sqrt{ml+nd}\Delta$, K_1 as the event $\|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{nys}\|_F \le (1+\epsilon)\|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{proj}\|_F$, K_2 as the event $\|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{proj}\|_F \le (1+\epsilon)\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F$, and B_C and B_R as the events $\|\mathbf{C}_0 - \hat{\mathbf{C}}\|_F \le c_e\sqrt{ml}\Delta$ and $\|\mathbf{R}_0 - \hat{\mathbf{R}}\|_F \le c_e\sqrt{dn}\Delta$. As above,

$$\|\mathbf{L}_0 - \hat{\mathbf{L}}^{nys}\|_F \le \|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F + \|\hat{\mathbf{L}} - \hat{\mathbf{L}}^{nys}\|_F \le (2+2\epsilon+\epsilon^2)\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F \le (2+3\epsilon)\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F,$$

when $K_1 \cap K_2$ holds, by triangle inequality. Reasoning identical to that above now yields

$$\begin{aligned}
\mathbf{P}(J) &\ge \mathbf{P}(B_C \cap B_R \mid K_1 \cap K_2 \cap A(\mathbf{C}) \cap A(\mathbf{R}))\mathbf{P}(K_1 \cap A(\mathbf{C}))\mathbf{P}(K_2 \cap A(\mathbf{R})) \\
&\ge (1-\delta)^2(1-\delta_C-\delta_R).
\end{aligned}$$

C Proof of Corollary 4

Cor. 4 is based on a new noisy MC theorem, which we prove in Sec. E. A similar recovery guarantee is obtained by [3] under stronger assumptions.

Theorem 13. *Suppose that $\mathbf{L}_0 \in \mathbb{R}^{m\times n}$ is (μ, r)-coherent and that, for some target rate parameter $\beta > 1$,*

$$s \geq 32\mu r(m+n)\beta \log^2(m+n)$$

entries of $\mathbf{M}$ are observed with locations Ω sampled uniformly without replacement. Then, if $m \leq n$ and $\|\mathcal{P}_\Omega(\mathbf{M}) - \mathcal{P}_\Omega(\mathbf{L}_0)\|_F \leq \Delta$ a.s., the minimizer $\hat{\mathbf{L}}$ to the problem

$$\text{minimize}_{\mathbf{L}} \quad \|\mathbf{L}\|_* \quad \text{subject to} \quad \|\mathcal{P}_\Omega(\mathbf{M}-\mathbf{L})\|_F \leq \Delta \tag{4}$$

satisfies

$$\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F \leq 8\sqrt{\frac{2m^2n}{s} + m + \frac{1}{16}}\Delta \leq c'_e\sqrt{mn}\Delta$$

with probability at least $1 - 4\log(n)n^{2-2\beta}$ for c'_e a positive constant.

We begin by proving the DFC-PROJ bound. For each $i \in \{1,\dots,t\}$, let B_i be the event that $\|\mathbf{C}_{0,i} - \hat{\mathbf{C}}_i\|_F > c'_e\sqrt{ml}\Delta$ and D_i be the event that $s_i < 32\mu' r(m+l)\beta' \log^2(m+l)$, where s_i is the number of revealed entries in $\mathbf{C}_{0,i}$,

$$\mu' \triangleq \frac{\mu^2 r}{1-\epsilon/2}, \qquad \text{and} \qquad \beta' \triangleq \frac{\beta\log(n)}{\log(\max(m,l))}.$$

Then, by Thm. 3, it suffices to establish that

$$\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) \leq (4\log(\bar{n})+1)\bar{n}^{2-2\beta}$$

for each i. By Thm. 13 and our choice of β',

$$\begin{aligned}\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) &\leq \mathbf{P}(B_i \mid A(\mathbf{C}_{0,i}), D_i^c) + \mathbf{P}(D_i \mid A(\mathbf{C}_{0,i}))\\ &\leq 4\log(\max(m,l))\max(m,l)^{2-2\beta'} + \mathbf{P}(D_i)\\ &\leq 4\log(\bar{n})\bar{n}^{2-2\beta} + \mathbf{P}(D_i).\end{aligned}$$

Further, since the support of $\mathbf{S}_0$ is uniformly distributed and of cardinality s, the variable s_i has a hypergeometric distribution with $\mathrm{E}s_i = \frac{sl}{n}$ and hence satisfies Hoeffding's inequality for the hypergeometric distribution [10, Sec. 6]:

$$\mathbf{P}(s_i \leq \mathrm{E}s_i - st) \leq \exp\big(-2st^2\big).$$

It therefore follows that

$$\begin{aligned}\mathbf{P}(D_i) &= \mathbf{P}\left(s_i < \mathrm{E}s_i - s\left(\frac{l}{n} - \frac{32\mu' r(m+l)\beta'\log^2(m+l)}{s}\right)\right)\\ &= \mathbf{P}\left(s_i < \mathrm{E}s_i - s\left(\frac{l}{n} - \frac{\beta(m+l)\log^2(m+l)}{\beta_s(m+n)\log^2(m+n)}\frac{\log(\bar{n})}{\log(\max(m,l))}\right)\right)\\ &\leq \mathbf{P}\left(s_i < \mathrm{E}s_i - s\left(\frac{l}{n} - \frac{\beta}{\beta_s}\right)\right)\\ &\leq \mathbf{P}\left(s_i < \mathrm{E}s_i - s\sqrt{\frac{\beta-1}{n\beta_s}}\right)\\ &\leq \exp\left(-2s\frac{\beta-1}{n\beta_s}\right) \leq \exp(-2\log(\bar{n})(\beta-1)) = \bar{n}^{2-2\beta}\end{aligned}$$

by our assumptions on s and l. Hence, $\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) \leq (4\log(\bar{n})+1)\bar{n}^{2-2\beta}$ for each i, and the DFC-PROJ result follows from Thm. 3.

For DFC-NYS, let B_C be the event that $\|\mathbf{C}_0 - \hat{\mathbf{C}}\|_F > c'_e\sqrt{ml}\Delta$ and B_R be the event that $\|\mathbf{R}_0 - \hat{\mathbf{R}}\|_F > c'_e\sqrt{dn}\Delta$. Reasoning identical to that above yields $\mathbf{P}(B_C \mid A(\mathbf{C})) \leq (4\log(\bar{n})+1)\bar{n}^{2-2\beta}$ and $\mathbf{P}(B_R \mid A(\mathbf{R})) \leq (4\log(\bar{n})+1)\bar{n}^{2-2\beta}$. Thus, the DFC-NYS bound also follows from Thm. 3.

D Proof of Corollary 5

Cor. 5 is based on the following theorem of Zhou et al. [25], reformulated for a generic rate parameter β, as described in [2, Section 3.1].

Theorem 14 (Thm. 2 of [25])**.** *Suppose that $\mathbf{L}_0$ is (μ, r)-coherent and that the support set of $\mathbf{S}_0$ is uniformly distributed among all sets of cardinality s. Then, if $m \leq n$ and $\|\mathbf{M} - \mathbf{L}_0 - \mathbf{S}_0\|_F \leq \Delta$ a.s., there is a constant c_p such that with probability at least $1 - c_p n^{-\beta}$, the minimizer $(\hat{\mathbf{L}}, \hat{\mathbf{S}})$ to the problem*

$$\textit{minimize}_{\mathbf{L},\mathbf{S}} \quad \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \quad \textit{subject to} \quad \|\mathbf{M} - \mathbf{L} - \mathbf{S}\|_F \leq \Delta \tag{5}$$

with $\lambda = 1/\sqrt{n}$ satisfies $\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F^2 + \|\mathbf{S}_0 - \hat{\mathbf{S}}\|_F^2 \leq c_e''^2 mn\Delta^2$, provided that

$$r \leq \frac{\rho_r m}{\mu \log^2(n)} \quad \textit{and} \quad s \leq (1 - \rho_s\beta)mn$$

for target rate parameter $\beta > 2$, and positive constants ρ_r, ρ_s, and c_e''.

We begin by proving the DFC-Proj bound. For each $i \in \{1, \ldots, t\}$, let B_i be the event that $\|\mathbf{C}_{0,i} - \hat{\mathbf{C}}_i\|_F > c_e''\sqrt{ml}\Delta$, and further define $\bar{m} \triangleq \max(m, l)$ and $\beta'' \triangleq \beta\log(\bar{n})/\log(\bar{m}) \leq \beta'$. Then, by Thm. 3, it suffices to establish that

$$\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) \leq (c_p + 1)\bar{n}^{-\beta}$$

for each i. By Thm. 14 and the definitions of β' and β'',

$$\begin{aligned}
\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) &\leq \mathbf{P}(B_i \mid A(\mathbf{C}_{0,i}), s_i \leq (1 - \rho_s\beta'')ml) + \mathbf{P}(s_i > (1 - \rho_s\beta'')ml \mid A(\mathbf{C}_{0,i})) \\
&\leq c_p\bar{m}^{-\beta''} + \mathbf{P}(s_i > (1 - \rho_s\beta'')ml) \\
&\leq c_p\bar{n}^{-\beta} + \mathbf{P}(s_i > (1 - \rho_s\beta')ml),
\end{aligned}$$

where s_i is the number of corrupted entries in $\mathbf{C}_{0,i}$. Further, since the support of $\mathbf{S}_0$ is uniformly distributed and of cardinality s, the variable s_i has a hypergeometric distribution with $\mathrm{E}s_i = \frac{sl}{n}$ and hence satisfies Bernstein's inequality for the hypergeometric [10, Sec. 6]:

$$\mathbf{P}(s_i \geq \mathrm{E}s_i + st) \leq \exp\big(-st^2/(2\sigma^2 + 2t/3)\big) \leq \exp\big(-st^2 n/4l\big),$$

for all $0 \leq t \leq 3l/n$ and $\sigma^2 \triangleq \frac{l}{n}(1 - \frac{l}{n}) \leq \frac{l}{n}$. It therefore follows that

$$\begin{aligned}
\mathbf{P}(s_i > (1 - \rho_s\beta')ml) &= \mathbf{P}\left(s_i > \mathrm{E}s_i + s\left(\frac{(1 - \rho_s\beta')ml}{s} - \frac{l}{n}\right)\right) \\
&= \mathbf{P}\left(s_i > \mathrm{E}s_i + s\frac{l}{n}\left(\frac{(1 - \rho_s\beta')}{(1 - \rho_s\beta_s)} - 1\right)\right) \\
&\leq \exp\left(-s\frac{l}{4n}\left(\frac{(1 - \rho_s\beta')}{(1 - \rho_s\beta_s)} - 1\right)^2\right) \\
&= \exp\left(-\frac{ml}{4}\frac{(\rho_s\beta_s - \rho_s\beta')^2}{(1 - \rho_s\beta_s)}\right) \leq \bar{n}^{-\beta}
\end{aligned}$$

by our assumptions on s and l and the fact that $\frac{l}{n}\left(\frac{(1-\rho_s\beta')}{(1-\rho_s\beta_s)} - 1\right) \leq 3l/n$ whenever $4\beta_s - 3/\rho_s \leq \beta'$. Hence, $\mathbf{P}(B_i \mid A(\mathbf{C}_{0,i})) \leq (c_p + 1)\bar{n}^{-\beta}$ for each i, and the DFC-Proj result follows from Thm. 3.

For DFC-Nys, let B_C be the event that $\|\mathbf{C}_0 - \hat{\mathbf{C}}\|_F > c_e''\sqrt{ml}\Delta$ and B_R be the event that $\|\mathbf{R}_0 - \hat{\mathbf{R}}\|_F > c_e''\sqrt{dn}\Delta$. Reasoning identical to that above yields $\mathbf{P}(B_C \mid A(\mathbf{C})) \leq (c_p + 1)\bar{n}^{-\beta}$ and $\mathbf{P}(B_R \mid A(\mathbf{R})) \leq (c_p + 1)\bar{n}^{-\beta}$. Thus, the DFC-Nys bound also follows from Thm. 3.

E Proof of Theorem 13

In the spirit of [3], our proof will extend the noiseless analysis of [22] to the noisy matrix completion setting. As suggested in [9], we will obtain strengthened results, even in the noiseless case, by reasoning directly about the without-replacement sampling model, rather than appealing to a with-replacement surrogate, as done in [22].

For $\mathbf{U}_{L_0}\mathbf{\Sigma}_{L_0}\mathbf{V}_{L_0}^\top$ the compact SVD of $\mathbf{L}_0$, we let $T = \{\mathbf{U}_{L_0}\mathbf{X}+\mathbf{Y}\mathbf{V}_{L_0}^\top : \mathbf{X} \in \mathbb{R}^{m\times r}, \mathbf{Y} \in \mathbb{R}^{m\times r}\}$, $\mathcal{P}_T$ denote orthogonal projection onto the space T, and $\mathcal{P}_{T^\perp}$ represent orthogonal projection onto the orthogonal complement of T. We further define $\mathcal{I}$ as the identity operator on $\mathbb{R}^{m\times n}$ and the spectral norm of an operator $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ as $\|\mathcal{A}\|_2 = \sup_{\|\mathbf{X}\|_F\le 1} \|\mathcal{A}(\mathbf{X})\|_F$.

We begin with a theorem providing sufficient conditions for our desired recovery guarantee.

Theorem 15. *Under the assumptions of Thm. 13, suppose that*

$$\frac{mn}{s}\Big\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T - \frac{s}{mn}\mathcal{P}_T\Big\|_2 \le \frac{1}{2} \tag{6}$$

and that there exists a $\mathbf{Y} = \mathcal{P}_\Omega(\mathbf{Y}) \in \mathbb{R}^{m\times n}$ *satisfying*

$$\|\mathcal{P}_T(\mathbf{Y}) - \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top\|_F \le \sqrt{\frac{s}{32mn}} \quad \textit{and} \quad \|\mathcal{P}_{T^\perp}(\mathbf{Y})\|_2 < \frac{1}{2}. \tag{7}$$

Then,

$$\|\mathbf{L}_0 - \hat{\mathbf{L}}\|_F \le 8\sqrt{\frac{2m^2n}{s} + m + \frac{1}{16}}\Delta \le c_e\sqrt{mn}\Delta.$$

Proof We may write $\hat{\mathbf{L}}$ as $\mathbf{L}_0 + \mathbf{G} + \mathbf{H}$, where $\mathcal{P}_\Omega(\mathbf{G}) = \mathbf{G}$ and $\mathcal{P}_\Omega(\mathbf{H}) = \mathbf{0}$. Then, under Eq. (6),

$$\|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{H})\|_F^2 = \big\langle \mathbf{H}, \mathcal{P}_T\mathcal{P}_\Omega^2\mathcal{P}_T(\mathbf{H})\big\rangle \ge \langle \mathbf{H}, \mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{H})\rangle \ge \frac{s}{2mn}\|\mathcal{P}_T(\mathbf{H})\|_F^2.$$

Furthermore, by the triangle inequality, $0 = \|\mathcal{P}_\Omega(\mathbf{H})\|_F \ge \|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{H})\|_F - \|\mathcal{P}_\Omega\mathcal{P}_{T^\perp}(\mathbf{H})\|_F$. Hence, we have

$$\sqrt{\frac{s}{2mn}}\|\mathcal{P}_T(\mathbf{H})\|_F \le \|\mathcal{P}_\Omega\mathcal{P}_T(\mathbf{H})\|_F \le \|\mathcal{P}_\Omega\mathcal{P}_{T^\perp}(\mathbf{H})\|_F \le \|\mathcal{P}_{T^\perp}(\mathbf{H})\|_F \le \|\mathcal{P}_{T^\perp}(\mathbf{H})\|_*, \tag{8}$$

where the penultimate inequality follows as $\mathcal{P}_\Omega$ is an orthogonal projection operator.

Next we select $\mathbf{U}_\perp$ and $\mathbf{V}_\perp$ such that $[\mathbf{U}_{L_0}, \mathbf{U}_\perp]$ and $[\mathbf{V}_{L_0}, \mathbf{V}_\perp]$ are orthonormal and $\big\langle \mathbf{U}_\perp\mathbf{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathbf{H})\big\rangle = \|\mathcal{P}_{T^\perp}(\mathbf{H})\|_*$ and note that

$$\begin{aligned}
\|\mathbf{L}_0 + \mathbf{H}\|_* &\ge \big\langle \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top + \mathbf{U}_\perp\mathbf{V}_\perp^\top, \mathbf{L}_0 + \mathbf{H}\big\rangle \\
&= \|\mathbf{L}_0\|_* + \big\langle \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top + \mathbf{U}_\perp\mathbf{V}_\perp^\top - \mathbf{Y}, \mathbf{H}\big\rangle \\
&= \|\mathbf{L}_0\|_* + \big\langle \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top - \mathcal{P}_T(\mathbf{Y}), \mathcal{P}_T(\mathbf{H})\big\rangle + \big\langle \mathbf{U}_\perp\mathbf{V}_\perp^\top, \mathcal{P}_{T^\perp}(\mathbf{H})\big\rangle - \langle \mathcal{P}_{T^\perp}(\mathbf{Y}), \mathcal{P}_{T^\perp}(\mathbf{H})\rangle \\
&\ge \|\mathbf{L}_0\|_* - \|\mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top - \mathcal{P}_T(\mathbf{Y})\|_F\|\mathcal{P}_T(\mathbf{H})\|_F + \|\mathcal{P}_{T^\perp}(\mathbf{H})\|_* - \|\mathcal{P}_{T^\perp}(\mathbf{Y})\|_2\|\mathcal{P}_{T^\perp}(\mathbf{H})\|_* \\
&> \|\mathbf{L}_0\|_* + \frac{1}{2}\|\mathcal{P}_{T^\perp}(\mathbf{H})\|_* - \sqrt{\frac{s}{32mn}}\|\mathcal{P}_T(\mathbf{H})\|_F \\
&\ge \|\mathbf{L}_0\|_* + \frac{1}{4}\|\mathcal{P}_{T^\perp}(\mathbf{H})\|_F
\end{aligned}$$

where the first inequality follows from the variational representation of the trace norm, $\|\mathbf{A}\|_* = \sup_{\|\mathbf{B}\|_2\le 1}\langle \mathbf{A}, \mathbf{B}\rangle$, the first equality follows from the fact that $\langle \mathbf{Y}, \mathbf{H}\rangle = 0$ for $\mathbf{Y} = \mathcal{P}_\Omega(\mathbf{Y})$, the second inequality follows from Hölder's inequality for Schatten p-norms, the third inequality follows from Eq. (7), and the final inequality follows from Eq. (8).

Since $\mathbf{L}_0$ is feasible for Eq. (4), $\|\mathbf{L}_0\|_* \ge \|\hat{\mathbf{L}}\|_*$, and, by the triangle inequality, $\|\hat{\mathbf{L}}\|_* \ge \|\mathbf{L}_0 + \mathbf{H}\|_* - \|\mathbf{G}\|_*$. Since $\|\mathbf{G}\|_* \le \sqrt{m}\|\mathbf{G}\|_F$ and $\|\mathbf{G}\|_F \le \|\mathcal{P}_\Omega(\hat{\mathbf{L}} - \mathbf{M})\|_F +$

$\|\mathcal{P}_\Omega(\mathbf{M}-\mathbf{L}_0)\|_F \le 2\Delta$, we conclude that

$$\begin{aligned}\|\mathbf{L}_0-\hat{\mathbf{L}}\|_F^2 &= \|\mathcal{P}_T(\mathbf{H})\|_F^2+\|\mathcal{P}_{T^\perp}(\mathbf{H})\|_F^2+\|\mathbf{G}\|_F^2\\ &\le \left(\frac{2mn}{s}+1\right)\|\mathcal{P}_{T^\perp}(\mathbf{H})\|_F^2+\|\mathbf{G}\|_F^2\\ &\le 16\left(\frac{2mn}{s}+1\right)\|\mathbf{G}\|_*^2+\|\mathbf{G}\|_F^2\\ &\le 64\left(\frac{2m^2n}{s}+m+\frac{1}{16}\right)\Delta^2.\end{aligned}$$

Hence

$$\|\mathbf{L}_0-\hat{\mathbf{L}}\|_F \le 8\sqrt{\frac{2m^2n}{s}+m+\frac{1}{16}}\Delta \le c_e\sqrt{mn}\Delta$$

for some constant c_e, by our assumption on s. □

To show that the sufficient conditions of Thm. 15 hold with high probability, we will require four lemmas. The first establishes that the operator $\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T$ is nearly an isometry on T when sufficiently many entries are sampled.

Lemma 16. *For all $\beta > 1$,*

$$\frac{mn}{s}\left\|\mathcal{P}_T\mathcal{P}_\Omega\mathcal{P}_T-\frac{s}{mn}\mathcal{P}_T\right\|_2 \le \sqrt{\frac{16\mu r(m+n)\beta\log(n)}{3s}}$$

with probability at least $1-2n^{2-2\beta}$ provided that $s > \frac{16}{3}\mu r(n+m)\beta\log(n)$.

The second states that a sparsely but uniformly observed matrix is close to a multiple of the original matrix under the spectral norm.

Lemma 17. *Let $\mathbf{Z}$ be a fixed matrix in $\mathbb{R}^{m\times n}$. Then for all $\beta > 1$,*

$$\left\|\left(\frac{mn}{s}\mathcal{P}_\Omega-\mathcal{I}\right)(\mathbf{Z})\right\|_2 \le \sqrt{\frac{8\beta mn^2\log(m+n)}{3s}}\|\mathbf{Z}\|_\infty$$

with probability at least $1-(m+n)^{1-\beta}$ provided that $s > 6\beta m\log(m+n)$.

The third asserts that the matrix infinity norm of a matrix in T does not increase under the operator $\mathcal{P}_T\mathcal{P}_\Omega$.

Lemma 18. *Let $\mathbf{Z}\in T$ be a fixed matrix. Then for all $\beta > 2$*

$$\left\|\frac{mn}{s}\mathcal{P}_T\mathcal{P}_\Omega(\mathbf{Z})-\mathbf{Z}\right\|_\infty \le \sqrt{\frac{8\beta\mu r(m+n)\log(n)}{3s}}\|\mathbf{Z}\|_\infty$$

with probability at least $1-2n^{2-\beta}$ provided that $s > \frac{8}{3}\beta\mu r(m+n)\log(n)$.

These three lemmas were proved in [22, Thm. 3.4, Thm. 3.5, and Lemma 3.6] under the assumption that entry locations in Ω were sampled *with* replacement. They admit identical proofs under the sampling without replacement model by noting that the referenced Noncommutative Bernstein Inequality [22, Thm. 3.2] also holds under sampling without replacement, as shown in [9].

Lemma 16 guarantees that Eq. (6) holds with high probability. To construct a matrix $\mathbf{Y}=\mathcal{P}_\Omega(\mathbf{Y})$ satisfying Eq. (7), we consider a sampling with batch replacement scheme recommended in [9] and developed in [5]. Let $\tilde{\Omega}_1,\ldots,\tilde{\Omega}_p$ be independent sets, each consisting of q random entry locations sampled without replacement, where $pq=s$. Let $\tilde{\Omega}=\cup_{i=1}^p\tilde{\Omega}_i$, and note that there exist p and q satisfying

$$q \ge \frac{128}{3}\mu r(m+n)\beta\log(m+n) \quad \text{and} \quad p \ge \frac{3}{4}\log(n/2).$$

It suffices to establish Eq. (7) under this batch replacement scheme, as shown in the next lemma.

Lemma 19. *For any location set $\Omega_0 \subset \{1,\ldots,m\}\times\{1,\ldots,n\}$, let $A(\Omega_0)$ be the event that there exists $\mathbf{Y} = \mathcal{P}_{\Omega_0}(\mathbf{Y}) \in \mathbb{R}^{m\times n}$ satisfying Eq. (7). If $\Omega(s)$ consists of s locations sampled uniformly without replacement and $\tilde{\Omega}(s)$ is sampled via batch replacement with p batches of size q for $pq = s$, then $\mathbf{P}(A(\tilde{\Omega}(s))) \leq \mathbf{P}(A(\Omega(s)))$.*

Proof As sketched in [9]

$$\begin{aligned}
\mathbf{P}\Big(A(\tilde{\Omega(s)})\Big) &= \sum_{i=1}^{s} \mathbf{P}(|\tilde{\Omega}| = i)\mathbf{P}(A(\tilde{\Omega}(i)) \mid |\tilde{\Omega}| = i) \\
&\leq \sum_{i=1}^{s} \mathbf{P}(|\tilde{\Omega}| = i)\mathbf{P}(A(\Omega(i))) \\
&\leq \sum_{i=1}^{s} \mathbf{P}(|\tilde{\Omega}| = i)\mathbf{P}(A(\Omega(s))) = \mathbf{P}(A(\Omega(s))),
\end{aligned}$$

since the probability of existence never decreases with more entries sampled without replacement and, given the size of $\tilde{\Omega}$, the locations of $\tilde{\Omega}$ are conditionally distributed uniformly (without replacement). □

We now follow the construction of [22] to obtain $\mathbf{Y} = \mathcal{P}_{\tilde{\Omega}}(\mathbf{Y})$ satisfying Eq. (7). Let $\mathbf{W}_0 = \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top$ and define $\mathbf{Y}_k = \frac{mn}{q}\sum_{j=1}^{k}\mathcal{P}_{\tilde{\Omega}_j}(\mathbf{W}_{j-1})$ and $\mathbf{W}_k = \mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top - \mathcal{P}_T(\mathbf{Y}_k)$ for $k = 1,\ldots,p$. Assume that

$$\frac{mn}{q}\left\|\mathcal{P}_T\mathcal{P}_{\tilde{\Omega}_k}\mathcal{P}_T - \frac{q}{mn}\mathcal{P}_T\right\|_2 \leq \frac{1}{2} \tag{9}$$

for all k. Then

$$\|\mathbf{W}_k\|_F = \left\|\mathbf{W}_{k-1} - \frac{mn}{q}\mathcal{P}_T\mathcal{P}_{\tilde{\Omega}_k}(\mathbf{W}_{k-1})\right\|_F = \left\|(\mathcal{P}_T - \frac{mn}{q}\mathcal{P}_T\mathcal{P}_{\tilde{\Omega}_k}\mathcal{P}_T)(\mathbf{W}_{k-1})\right\|_F \leq \frac{1}{2}\|\mathbf{W}_{k-1}\|_F$$

and hence $\|\mathbf{W}_k\|_F \leq 2^{-k}\|\mathbf{W}_0\|_F = 2^{-k}\sqrt{r}$. Since $p \geq \frac{3}{4}\log(n/2) \geq \frac{1}{2}\log_2(n/2) \geq \log_2\sqrt{32rmn/s}$, $\mathbf{Y} \triangleq \mathbf{Y}_p$ satisfies the first condition of Eq. (7).

The second condition of Eq. (7) follows from the assumptions

$$\left\|\mathbf{W}_{k-1} - \frac{mn}{q}\mathcal{P}_T\mathcal{P}_{\tilde{\Omega}_k}(\mathbf{W}_{k-1})\right\|_\infty \leq \frac{1}{2}\|\mathbf{W}_{k-1}\|_\infty \tag{10}$$

$$\left\|\left(\frac{mn}{q}\mathcal{P}_{\tilde{\Omega}_k} - \mathcal{I}\right)(\mathbf{W}_{k-1})\right\|_2 \leq \sqrt{\frac{8mn^2\beta\log(m+n)}{3q}}\|\mathbf{W}_{k-1}\|_\infty \tag{11}$$

for all k, since Eq. (10) implies $\|\mathbf{W}_k\|_\infty \leq 2^{-k}\|\mathbf{U}_{L_0}\mathbf{V}_{L_0}^\top\|_\infty$, and thus

$$\begin{aligned}
\|\mathcal{P}_{T^\perp}(\mathbf{Y}_p)\|_2 &\leq \sum_{j=1}^{p}\left\|\frac{mn}{q}\mathcal{P}_{T^\perp}\mathcal{P}_{\tilde{\Omega}_j}(\mathbf{W}_{j-1})\right\|_2 \\
&= \sum_{j=1}^{p}\left\|\mathcal{P}_{T^\perp}(\frac{mn}{q}\mathcal{P}_{\tilde{\Omega}_j}(\mathbf{W}_{j-1}) - \mathbf{W}_{j-1})\right\|_2 \\
&\leq \sum_{j=1}^{p}\left\|(\frac{mn}{q}\mathcal{P}_{\tilde{\Omega}_j} - \mathcal{I})(\mathbf{W}_{j-1})\right\|_2 \\
&\leq \sum_{j=1}^{p}\sqrt{\frac{8mn^2\beta\log(m+n)}{3q}}\|\mathbf{W}_{j-1}\|_\infty \\
&= 2\sum_{j=1}^{p}2^{-j}\sqrt{\frac{8mn^2\beta\log(m+n)}{3q}}\|\mathbf{U}_W\mathbf{V}_W^\top\|_\infty < \sqrt{\frac{32\mu rn\beta\log(m+n)}{3q}} < 1/2
\end{aligned}$$

by our assumption on q. The first line applies the triangle inequality; the second holds since $\mathbf{W}_{j-1} \in T$ for each j; the third follows because $\mathcal{P}_{T^\perp}$ is an orthogonal projection; and the final line exploits (μ, r)-coherence.

We conclude by bounding the probability of any assumed event failing. Lemma 16 implies that Eq. (6) fails to hold with probability at most $2n^{2-2\beta}$. For each k, Eq. (9) fails to hold with probability at most $2n^{2-2\beta}$ by Lemma 16, Eq. (10) fails to hold with probability at most $2n^{2-2\beta}$ by Lemma 18, and Eq. (11) fails to hold with probability at most $(m+n)^{1-2\beta}$ by Lemma 17. Hence, by the union bound, the conclusion of Thm. 15 holds with probability at least

$$1 - 2n^{2-2\beta} - \frac{3}{4}\log(n/2)(4n^{2-2\beta} + (m+n)^{1-2\beta}) \geq 1 - \frac{15}{4}\log(n)n^{2-2\beta} \geq 1 - 4\log(n)n^{2-2\beta}.$$