MC^2: A Two-Phase Algorithm for Leveraged Matrix Completion

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
MC\textsuperscript{2}: A Two-Phase Algorithm for Leveraged Matrix Completion
Armin Eftekhari, Michael B. Wakin, and Rachel A. Ward\textsuperscript{*}

September 8, 2016

Abstract

Leverage scores, loosely speaking, reflect the importance of the rows and columns of a matrix. Ideally, given the leverage scores of a rank-\(r\) matrix \(M \in \mathbb{R}^{n \times n}\), that matrix can be reliably completed from just \(O(rn \log^2 n)\) samples if the samples are chosen randomly from a nonuniform distribution induced by the leverage scores. In practice, however, the leverage scores are often unknown \textit{a priori}. As such, the sample complexity in standard matrix completion—using uniform random sampling—increases to \(O(\eta(M) \cdot rn \log^2 n)\), where \(\eta(M)\) is the largest leverage score of \(M\). In this paper, we propose a two-phase algorithm called MC\textsuperscript{2} for matrix completion: in the first phase, the leverage scores are estimated based on uniform random samples, and then in the second phase the matrix is resampled nonuniformly based on the estimated leverage scores and then completed. The total sample complexity of MC\textsuperscript{2} is provably smaller than standard matrix completion—substantially so for coherent but well-conditioned matrices. In fact, the dependence on condition number appears to be an artifact of the proof techniques, as numerical simulations suggest that the algorithm outperforms standard matrix completion in a much broader class of matrices.

1 Introduction

Matrix completion is commonly defined as the problem of recovering a low-rank matrix \(M \in \mathbb{R}^{n_1 \times n_2}\) from a fraction of its entries, observed on an often random index set \([1, 2, 3]\). To be concrete,\textsuperscript{1} let \(n_1 = n_2 = n\) and set \(r = \text{rank}(M)\) for short. Also let \(M = U_r \Sigma_r V_r^*\) be the singular value decomposition (SVD) of \(M\), where \(U_r, V_r \in \mathbb{R}^{n \times r}\) have orthonormal columns and the diagonal matrix \(\Sigma_r \in \mathbb{R}^{r \times r}\) contains the singular values of \(M\).

In standard low-rank matrix completion (SMC), each entry of \(M\) is observed with a probability of \(p \in (0, 1]\) so that, in expectation, \(pn^2\) entries of \(M\) are revealed. Remarkably, from these uniform samples, \(M\) can be successfully reconstructed (via convex programming, for example) provided that\textsuperscript{2}

\[
\eta(M) \cdot \frac{r \log^2 n}{n} \lesssim p \leq 1,
\]

where \(\eta(M)\), the \textit{coherence} of \(M\), in a sense measures how “diffuse” \(M\) is. Above, the dependence of \(p\) on \(r\) and \(n\) is optimal, up to a logarithmic factor, and also

\[
\eta(M) := \frac{n}{r} \left( \|U_r\|_{2 \to \infty}^2 \vee \|V_r\|_{2 \to \infty}^2 \right) = \frac{n}{r} \left( \max_{i \in [1:n]} \|U_r[i,:]\|_2^2 \vee \max_{j \in [1:n]} \|V_r[j,:]\|_2^2 \right),
\]

with \(U_r[i,:]\) and \(V_r[j,:]\) standing for the corresponding rows of \(U_r\) and \(V_r\), respectively. We also used the conventions \(a \vee b = \max\{a, b\}\) and \([c : d] = \{c, c+1, \cdots, d\}\) for integers \(c \leq d\). One may verify that

\textsuperscript{*}AE is with the Alan Turing Institute in London. MBW is with the Department of Electrical Engineering and Computer Science at the Colorado School of Mines. RW is with the Department of Mathematics at the University of Texas at Austin and a member of the Institute for Computational Engineering and Sciences (ICES). Email: armin.eftekhari@gmail.com. MBW was partially supported by NSF CAREER Grant CCF-1149225 and NSF Grant CCF-1409258. RW was partially funded by NSF CAREER Grant CCF-1255631.

\textsuperscript{1}To make the introduction as accessible as possible, technical details are kept to a minimum in this section and deferred to Sections 2 and 3.

\textsuperscript{2}Throughout, we often use \(\lesssim\) to simplify the presentation by omitting universal constant factors.
\( \eta(M) \in [1, \frac{2}{3}] \). It is also common to say that \( M \) is coherent (incoherent) when \( \eta(M) \) is very large (small). Loosely speaking, a coherent matrix is “spiky” whereas an incoherent matrix is “diffuse.” For example, if \( M[1, 1] = 1 \) is the only nonzero entry of \( M \), then \( M \) is extremely coherent since \( \eta(M) = \frac{2}{3} = n \).

Roughly speaking then, in SMC, we can expect to successfully recover \( M \) from \( O(\eta(M) \cdot r n \log^2 n) \) uniform samples. In particular, when \( M \) is incoherent, say \( \eta(M) \approx 1 \), then \( M \) can be completed from \( O(r n \log^2 n) \ll n^2 \) uniform samples. In contrast, when \( M \) is coherent, say \( \eta(M) \approx \frac{2}{3} \), then one needs to observe nearly all entries of \( M \).

For instance, if \( M[1, 1] = 1 \) is the only nonzero entry of \( M \), then uniform sampling will collect \( M[1, 1] \) with a probability of \( p \). So, unless \( p \approx 1 \), \( M[1, 1] \) is not observed and successful reconstruction of this coherent matrix from uniform samples is highly unlikely.

The poor performance of SMC in completing coherent matrices can be remedied by means of leveraged (rather than uniform) sampling [4]. In the example above, \( M[1, 1] \) is by far the most important entry of \( M \). Therefore, a better sampling strategy might be to measure \( M[1, 1] \) with more likelihood than the rest of the entries. More generally, the importance of the rows and columns of a rank-\( r \) matrix \( M \in \mathbb{R}^{n \times n} \) are often measured by its leverage scores defined as

\[
\mu_i(M) := \frac{n}{r} \| U_r[i, :]\|_2^2, \quad \nu_j(M) := \frac{n}{r} \| V_r[j, :]\|_2^2, \quad i, j \in [1:n].
\]

It is easily verified that \( \mu_i(M), \nu_j(M) \in [0, \frac{2}{3}] \) for all \( i, j \in [1:n] \), and that the coherence \( \eta(M) \) is simply the largest leverage score of \( M \). Moreover,

\[
\sum_{i=1}^{n} \mu_i(M) = \sum_{j=1}^{n} \nu_j(M) = n,
\]

since \( U_r^* U_r = V_r^* V_r = I_r \), with \( I_r \in \mathbb{R}^{r \times r} \) being the identity matrix. Naturally, we might consider \( \mu_i(M) + \nu_j(M) \) as an indicator of the importance of \( M[i, j] \). In the example where \( M[1, 1] = 1 \) is the only nonzero entry of \( M \), \( \mu_1(M) + \nu_1(M) = 2n \), whereas \( \mu_i(M) + \nu_j(M) = 0 \) for every \( i, j > 1 \), suggesting the importance of the first row and column of \( M \).

If the leverage scores of \( M \) were known in advance, a good sampling strategy would be to measure the entries of \( M \) according to their importance \( \mu_i(M) + \nu_j(M) \). More specifically, in leveraged low-rank matrix completion (LMC), \( M \) can be recovered (via convex programming, for instance) provided that each entry \( M[i, j] \) is observed with a probability of \( P[i, j] \) that satisfies

\[
(\mu_i(M) + \nu_j(M)) \frac{r \log^2 n}{n} \lesssim P[i, j] \leq 1, \quad \forall i, j \in [1:n].
\]

That is, we can expect to recover \( M \) from

\[
O(1) \sum_{i, j=1}^{n} P[i, j] = O(1) \sum_{i, j} (\mu_i(M) + \nu_j(M)) \cdot \frac{r \log^2 n}{n} = O(\eta(M) r n \log^2 n) \quad \text{(see (4))}
\]

entries, as opposed to \( O(\eta(M) \cdot r n \log^2 n) \) uniform samples required in SMC, thereby removing any dependence on coherence.

In practice, however, the leverage scores of \( M \) are often unknown \textit{a priori}, suggesting the need for a matrix completion scheme that would improve over SMC, particularly in completing coherent matrices, and yet would \textit{not} require much prior knowledge about \( M \). In this paper, we propose a two-phase algorithm for matrix completion—dubbed MC²—which first estimates the relatively large leverage scores of \( M \) from uniform samples. These estimated leverage scores are then used to complete \( M \) by applying LMC. MC² is developed in Section 3 and summarized in Figure 2.

Unlike LMC, MC² requires little prior knowledge about \( M \) and yet substantially improves over SMC when, loosely speaking, \( M \) is coherent (\( \eta(M) \gg 1 \)) but well-conditioned. In fact, the dependence on the condition number appears to be an artifact of the proof techniques; in numerical simulations, MC² improves over SMC for a much broader class of matrices. The performance of MC² and SMC are on par for completion of incoherent matrices. A variant of MC² first appeared in [4] where it was (heuristically) shown to outperform
SMC. Our main contribution in this paper is in carefully studying the performance of MC², the summary of which appears in Corollary 7. The other two contributions of this work are outlined later in this section.

For now, we briefly demonstrate these points with an example (and postpone additional numerical simulations to Section 5). Set $n = 50$ and $r = 5$. Let $U_r, V_r \in \mathbb{R}^{n \times r}$ (with orthonormal columns) span generic $r$-dimensional subspaces in $\mathbb{R}^n$ and also draw the nonzero entries of the diagonal matrix $\Sigma_r \in \mathbb{R}^{r \times r}$ independently from the uniform distribution on the interval $(0, 1]$. Set $M = U_r \Sigma_r V_r^\ast \in \mathbb{R}^{n \times n}$. By design, $M$ is often incoherent, $\eta(M) \approx 1$. When each entry of $M$ is revealed with a probability of $p$, we compare the average recovery error of MC² and SMC as $p$ increases (over many trials). The results, presented in Figure 1a, suggest that MC² and SMC perform similarly in completing incoherent matrices.

In contrast, set $M' = DMD$, where $D \in \mathbb{R}^{n \times n}$ is diagonal, with its nonzero entries following the power law, $D[i,i] \propto i^{-1}$, for every $i \in [1 : n]$. By design, $M'$ is often coherent, $\eta(M') \gg 1$. Figure 1b compares the average performance of the two algorithms in reconstructing $M'$. We repeat the same experiment when $D[i,i] \propto i^{-2}$ (see Figure 1c). This experiment suggests that MC² substantially outperforms SMC in completing coherent matrices.

---

Figure 1: Comparison of MC² and SMC for completing rank-5 matrices of size $50 \times 50$, in the absence of noise. See Section 5 for additional numerical comparisons.
other contributions It is worth noting that Proposition 2 in Section 2 extends the recent result on leveraged matrix completion in [4] to account for noise and to handle nearly low-rank matrices.

Additionally, as a byproduct of the analysis here, we also find a way to estimate the leverage scores of a matrix from a fraction of its entries (and up to a multiplicative factor) which is perhaps interesting in its own right, given the additive nature of the most available bounds in the literature of numerical linear algebra. More specifically, Corollary 7 in Section 4 describes how the leverage scores of M can be estimated from the row and column norms of a (zero-filled) partially-observed copy of M.

organization The rest of this paper is organized as follows. After a brief review, in Section 2, of the relevant concepts in matrix completion, MC^2 is developed in Section 3 and detailed in Figure 2. The accompanying theoretical guarantees are given in Corollary 3. Without being exhaustive, Section 4 compares MC^2 and SMC numerically. Related work is discussed in Section 6 and the proofs are deferred to the appendices. Section 4 might be of independent interest, in that it provides a class of estimators for the leverage scores of a matrix from uniform samples.

2 Matrix Completion: A Brief Review

Consider a matrix M ∈ R^{n×n} with the SVD M = UΣV^*. Here, U, V ∈ R^{n×n} are orthonormal bases, and the diagonal matrix Σ ∈ R^{n×n} collects the singular values of M, namely σ_1(M) ≥ σ_2(M) ≥ · · · ≥ σ_n(M), in non-increasing order. Let U_r, V_r ∈ R^{n×r} comprise of the first r columns of U, V, respectively, and let Σ_r ∈ R^{r×r} contain the first r (largest) singular values of M. Ties are broken arbitrarily. Then, M_r = U_rΣ_rV_r^* ∈ R^{n×n} is a rank-r truncation of M. We also set M_{r+} = M − M_r to be the residual.

Central to this work, the row and column leverage scores of M_r are defined as

μ_i(M_r) := \frac{n}{r} \|U_r[i, :]\|^2_2, \quad ν_j(M_r) := \frac{n}{r} \|V_r[:, j]\|^2_2, \quad i, j ∈ [1 : n]. \quad (7)

It is easy to see that μ_i(M_r), ν_j(M_r) ∈ [0, \frac{n}{r}] for every i, j ∈ [1 : n], and that

\sum_{i=1}^{n} μ_i(M_r) = \sum_{j=1}^{n} ν_j(M_r) = n, \quad (8)

so that {μ_i(M_r)/n}_i, {ν_j(M_r)/n}_j might be interpreted as probability distributions on the rows and columns of M_r, respectively. In a sense, leverage scores capture the importance of rows and columns of M_r. The coherence of M_r is set to be the largest leverage score, namely

η(M_r) = \max_{i \in [1 : n]} μ_i(M_r) \vee \max_{j \in [1 : n]} ν_j(M_r). \quad (9)

We may only access M through the linear measurement operator Ṙ_p(·). More specifically, for a probability p ∈ (0, 1], let

\begin{align*}
\mathcal{R}_p(M) := \sum_{i,j=1}^{n} \frac{ε_{i,j}}{p} · M[i,j] · E_{i,j} ∈ \mathbb{R}^{n×n},
\end{align*}

where {ε_{i,j}} are independent Bernoulli random variables. Each ε_{i,j} takes 1 with probability p and 0 otherwise. Also, E_{i,j} ∈ R^{n×n} is the [i,j]th canonical matrix, i.e., E_{i,j}[i,j] = 1 is the only nonzero entry of E_{i,j}. In standard low-rank matrix completion (SMC), the objective is to (approximately) recover M from \mathcal{R}_p(M) (or its noisy copy). The following result is a small modification of Theorem 7 in [5].

Proposition 1. [Standard matrix completion] For a matrix M ∈ R^{n×n} and probability p ∈ (0, 1], set Y_p = \mathcal{R}_p(M + D) to be a possibly noisy and partially observed (scaled) copy of M. Suppose that the noise matrix D ∈ R^{n×n} satisfies ∥\mathcal{R}_p(D)∥_F ≤ δ with δ ≥ 0. Let \tilde{M} ∈ R^{n×n} be a solution of

\begin{align*}
\begin{cases}
\min_{X} ∥X∥_*, \\
∥\mathcal{R}_p(X) − Y_p∥_F ≤ δ,
\end{cases}
\end{align*}

(11)
where \(\| \cdot \|_*\) and \(\| \cdot \|_F\) denote, respectively, nuclear norm (sum of the singular values) and Frobenius norm.

Fix an integer \(r \leq n\), let \(M_r\) be a rank-\(r\) truncation of \(M\), and set \(M_{r+} = M - M_r\) to be the residual. Then, except with a probability of \(o(n^{-19})\), it holds that

\[
\| \hat{M} - M \|_F \lesssim \frac{\|M_{r+}\|_*}{\sqrt{p}} + \delta \sqrt{pn},
\]

(12)

provided that

\[
\eta(M_r) \cdot \frac{r \log^2 n}{n} \lesssim p \leq 1, \quad i, j \in [1:n].
\]

(13)

Qualitatively speaking, then, SMC approximately completes a nearly rank-\(r\) matrix \(M \in \mathbb{R}^{n \times n}\) after observing \(pn^2 = O(\eta(M_r) \cdot rn \log^2 n)\) of its entries, uniformly at random, and possibly corrupted by noise. SMC is particularly powerful when \(M_r\) is incoherent, say \(\eta(M_r) \approx 1\), in which case SMC requires only \(O(rn \log^2 n)\) uniform samples. For more coherent matrices, SMC requires increasingly more uniform samples. At worst, when \(\eta(M_r) = \frac{2}{3}\), we must observe nearly all entries of \(M\).

It is worth noting that, under different coherence assumptions, [6] has improved the requisite number of uniform samples to \(O(rn \log n)\) and it remains to be seen if Proposition 1 could be similarly improved. We also remark that there are alternatives to Program (11) for matrix completion, see for example [7, 8, 9, 10], among many other algorithms.

The poor performance of SMC in completing coherent matrices stems from the uniform sampling strategy. If the leverage scores of \(M\) were known in advance, a better sampling strategy might have been to measure important entries of \(M\) (namely those corresponding to large leverage scores) with more likelihood (rather than sampling \(M\) uniformly at random). Indeed, leveraged sampling leads to substantial improvement over SMC, as we next describe.

For a matrix of probabilities \(P \in (0,1]^{n \times n}\), let

\[
\mathcal{R}_P(M) := \sum_{i,j=1}^{n} \frac{\epsilon_{i,j}}{P[i,j]} \cdot M[i,j] \cdot E_{i,j} \in \mathbb{R}^{n \times n},
\]

(14)

where \(\{\epsilon_{i,j}\}\) are independent Bernoulli random variables: each \(\epsilon_{i,j}\) takes 1 with probability \(P[i,j]\) and 0 otherwise. In leveraged matrix completion (LMC), we aim to recover \(M\) from \(\mathcal{R}_P(M)\) (or its noisy copy). The following result, proved in Appendix A, is a straightforward extension of Theorem 2 in [4] to noisy and nearly low-rank matrices (from noise-free and exactly low-rank).

**Proposition 2. [Leveraged matrix completion]** Fix \(0 < l \leq h \leq 1\). For a matrix \(M \in \mathbb{R}^{n \times n}\) and matrix of probabilities \(P \in [l,h]^{n \times n}\), set \(Y_P = \mathcal{R}_P(M + D)\). Here, the noise matrix \(D \in \mathbb{R}^{n \times n}\) satisfies \(\| \mathcal{R}_P(D) \|_F \leq \delta\) with \(\delta \geq 0\). Let \(\hat{M} \in \mathbb{R}^{n \times n}\) be a solution of

\[
\begin{dcases}
\min_X \|X\|_* \\
\|\mathcal{R}_P(X) - Y_P\|_F \leq \delta.
\end{dcases}
\]

(15)

Fix an integer \(r \leq n\) and let \(M_r\) be a rank-\(r\) truncation of \(M\), and set \(M_{r+} = M - M_r\) to be the residual. Then, if \(l^{-1} = 1/(n)\) is bounded by a polynomial in \(n\) (of finite degree) and except with a probability of \(o(n^{-19})\), it holds that

\[
\| \hat{M} - M \|_F \lesssim \frac{\sqrt{h}}{l} \|M_{r+}\|_* + \frac{\delta \sqrt{nh^2}}{l},
\]

provided that

\[
l \lesssim (\mu_i(M_r) + \nu_j(M_r)) \frac{r \log^2 n}{n} \lesssim P[i,j] \leq 1, \quad i, j \in [1:n].
\]

In words, LMC approximately completes a nearly rank-\(r\) matrix \(M \in \mathbb{R}^{n \times n}\) after observing

\[
O(1) \cdot \left(\sum_{i,j=1}^{n} \mu_i(M_r) + \nu_j(M_r)\right) \cdot \frac{r \log^2 n}{n} = O(1) \cdot rn \log^2 n, \quad \text{(see (8))}
\]

entries of \(M\) regardless of its coherence. In practice, however, the leverage scores of \(M\) are often not known \textit{a priori} and this impedes the implementation of LMC. We set out to address this problem next.
3 MC\(^2\)

So far, we reviewed the standard and leveraged matrix completion in Section 2, and explained how the lack of a priori knowledge about the leverage scores impedes the implementation of leveraged sampling in practice. To resolve this issue, consider a two-phase algorithm which, in Phase I, estimates the relatively large leverage scores from a small number of uniform samples. Then, in Phase II, these estimated leverage scores are used for leveraged matrix completion.

To formally present the algorithm, let us introduce some additional notation first. With \(\{\sigma_i(M)\}_i\) standing for the singular values of \(M\) in a non-increasing order and for integer \(r \leq n\), let

\[
\kappa_r(M) = \frac{\sigma_1(M)}{\sigma_r(M)}, \quad \gamma_r(M) = \frac{\sigma_{r+1}(M)}{\sigma_r(M)}.
\]

Note that \(\kappa_r(M)\) is the condition number of \(M_r\) (rank-\(r\) truncation of \(M\)) and that \(\gamma_r(M)\) might be interpreted as an inverse "spectral gap." In particular, for a rank-\(r\) matrix \(M\), \(\gamma_r(M) = 0\). Let also \(a_+ = \max[a,0]\) be the positive part of \(a \in \mathbb{R}\). Additionally, for \(p, p', q \in [0,1]\), consider three independent Bernoulli random variables

\[
\epsilon_p \sim \text{Benoulli}(p), \quad \epsilon_{p'} \sim \text{Benoulli}(p'), \quad \epsilon_q \sim \text{Benoulli}(q).
\]

(Here, for example, \(\epsilon_{p'}\) takes 1 with probability \(p'\) and 0 otherwise.) Then, note that

\[
q \ll 1, \; p \approx p' + q \Rightarrow \epsilon_p \overset{\text{dist}}{=} \epsilon_{p'} \lor \epsilon_q,
\]

where \(\overset{\text{dist}}{=}\) denotes equality in distribution. More generally, we will use the following map to combine different sampling probabilities in the two phases of MC\(^2\):

\[
p' = \Pi(p, q) := \begin{cases} 1 - \frac{1-p}{1-q} & p \geq q, \\ 0 & p < q. \end{cases}
\]

(16)

In particular, note that \(\Pi(p, q) \approx p - q\) when \(q \ll 1\).

With this setup, MC\(^2\), the proposed two-phase algorithm for leveraged matrix completion is detailed in Figure 2. The following discussion might help clarify each of the steps of the algorithm.

**Discussion:** As Figure 2 suggests, MC\(^2\) requires prior knowledge about the rank and spectral gap of \(M\), as well as the noise level. Crucially, MC\(^2\) is allowed to make queries about the entries of a possibly noisy copy of the true matrix, namely \(\hat{M} = M + D\), where \(D \in \mathbb{R}^{n \times n}\) represents noise. More specifically, in MC\(^2\), the overall sampling budget is \(pn^2\). That is, if \(P[i,j]\) is the probability of observing \(\hat{M}[i,j]\) in MC\(^2\), then we are constrained by \(\sum_{i,j=1}^n P[i,j] \leq pn^2\). Half of this budget is spent in Phase I (comprising of the first two steps of MC\(^2\)) and the other half is invested in Phase II (the rest of steps in MC\(^2\)). There is nothing special about this budget and a more general analysis is given in Appendix B.

- In Step 1 of MC\(^2\), every entry of \(\hat{M}\) is observed with a probability of \(p/2\); the measurements are stored in \(Y \in \mathbb{R}^{n \times n}\) and the corresponding index set is \(\Omega \subset [1:n]^2\). We therefore spent half of our sampling budget in Step 1.

- In Step 2, \(Y\) is used to estimate the leverage scores of \(M_r\). More specifically, we calculate \(\{\mu_i(M_r), \nu_i(M_r)\}_i\) as estimates of \(\{\mu_i(M_r), \nu_i(M_r)\}_i\). As discussed later, the sampling budget of \(p/2\) per entry in Step 1 is often not enough to guarantee that \(\mu_r(M) \approx \mu_i(M_r)\) and \(\nu_r(M) \approx \nu_i(M_r)\) for all \(i\). However, what matters most is having reliable estimates for the large leverage scores, which likely correspond to the most informative entries of \(M\). Indeed, when most leverage scores are small, we will show that a small sampling budget suffices to guarantee that \(\mu_r(M) \approx \mu_i(M_r)\) and \(\nu_r(M) \approx \nu_i(M_r)\) when either of the leverage scores is large.

- In Step 3, loosely speaking, we observe the most important entries of \(\hat{M}\), as determined by the estimated leverage scores \(\{\overline{\mu}_{r,i}, \overline{\nu}_{r,i}\}_i\). More specifically, in Step 3, every entry \(\hat{M}[i,j]\) is collected with a
Input:

- Rank $r$, inverse of spectral gap $\gamma_r \in [0, 1]$, measurement budget of $p \in (0, 1]$ per entry, noise level $\delta \geq 0$, and tuning factor $\alpha > 0$.
- Access to the entries of (a possibly noisy copy of the true matrix) $\tilde{M} \in \mathbb{R}^{n \times n}$.

Output:

- Estimate $\hat{M} \in \mathbb{R}^{n \times n}$.

Body:

1. **(Uniform sampling)** Observe every entry of $\tilde{M}$ with a probability of $p/2$: let $Y \in \mathbb{R}^{n \times n}$ store the measurements, filled with zeros elsewhere. Let also $\Omega \subseteq [1 : n]^2$ be the corresponding index set over which $\tilde{M}$ is observed.

2. **(Estimating the large leverage scores)** With $C_1 = \frac{147}{16}, C_2 = 12, C_3 = \frac{147}{64}$, set
   
   \[
   \mu_{r,i} \leftarrow \left( \frac{C_1 \|Y[i,:]\|_2^2}{\sigma_r^2(Y)} + \frac{C_2 \gamma_r^2}{p^2} + \frac{C_3 p^2 \delta^2}{\sigma_r^2(Y)} \right) \frac{p n r}{r}, \quad i \in [1 : n],
   \]
   
   \[
   \nu_{r,j} \leftarrow \left( \frac{C_1 \|Y[,j]\|_2^2}{\sigma_r^2(Y)} + \frac{C_2 \gamma_r^2}{p^2} + \frac{C_3 p^2 \delta^2}{\sigma_r^2(Y)} \right) \frac{p n r}{r}, \quad j \in [1 : n].
   \]  

3. **(Leveraged sampling)** Set
   
   \[
   P[i,j] \leftarrow \frac{p}{2} + \left( \alpha^2 (\mu_{r,i} + \nu_{r,j}) \frac{r \log^2 n}{n} - \frac{p}{2} \right)_{+}, \quad i,j \in [1 : n].
   \]  

   If $\sum_{i,j=1}^n P[i,j] > p n^2$, then set to $p/2$ the smallest entry of $P$ that is strictly larger than $p/2$ and repeat until $\sum_{i,j=1}^n P[i,j] \leq p n^2$. Then, set
   
   \[
   P'[i,j] \leftarrow \Pi \left( P[i,j], \frac{P}{2} \right), \quad i,j \in [1 : n].
   \]

   Lastly, observe the $[i,j]$th entry of $\tilde{M}$ with a probability of $P'[i,j]$, for every $i,j \in [1 : n]$. Append the corresponding index set to $\Omega$, and append the new measurements to $Y$. Readjust the entries of $Y$ by dividing every $Y[i,j]$ by $P'[i,j]$.

4. **(Matrix completion)** Let $\hat{M}$ be a solution of the program

   \[
   \min_X \|X\|_*, \quad \|R_P(X) - Y\|_F \leq \delta.
   \]

   Above, for every $i,j \in [1 : n]$, the $[i,j]$th entry of $R_P(X)$ equals $X[i,j]/P[i,j]$ when $[i,j] \in \Omega$, and is zero otherwise.

Figure 2: MC$^2$: A two-phase algorithm for leveraged matrix completion.
probability of $P'[i, j]$. Critically, $P'[i, j] > 0$ whenever either of $\pi_{r,i}$ or $\pi_{r,j}$ is large (see (16-19)). The collected samples are then appended to $Y$ and the support $\Omega$ is updated accordingly. In fact, by design, this is equivalent to generating $Y'$ and $\Omega$ through the measurement operator $R_{P'}(\cdot)$ (see (14)). Care is also taken in Step 3 so as not to exceed the overall sampling budget of $pn^2$, so that $\sum_{i,j} P'[i,j] \leq pn^2$.

At this point, as later discussed in Appendix B, $Y$ is the output of leveraged sampling of $M$.

- In Step 4, we deploy LMC to reconstruct $M$ by solving Program (20) and obtaining $\tilde{M}$.

We study the performance of MC$^2$ in Appendix B, where we prove a statement more general than the following.

**Corollary 3. [Performance of MC$^2$]** Consider a matrix $M \in \mathbb{R}^{n \times n}$ and its possibly noisy copy $\tilde{M} = M + D \in \mathbb{R}^{n \times n}$. Apply MC$^2$ with input parameters $\gamma, \sigma, \delta, \alpha$ as detailed in Figure 2. Suppose that $\|R(D)\|_F \leq \delta$ and $\|R(D')\|_F \leq \delta$ (see (10), (14), and (18)). Suppose also that

$$\gamma_r(M) \leq \gamma_r \leq \frac{p}{8}, \quad \delta \leq \frac{\sigma_r(M)}{4},$$

where $\sigma_r(M)$ is the $r$th largest singular value of $M$. Assume also that

$$\# \left\{ i : \mu_i(M) \vee \nu_i(M) \geq \frac{\eta(M_i)}{\log^2 n} \right\} = O(1).$$

Lastly, assume that

$$\alpha^4 \left( \frac{\kappa_r(M)^2}{\lambda r^2} + \frac{\gamma^2 n}{\lambda pr^2} + \frac{\delta^2 \lambda p r}{\sigma_r(M)^2} \right)^2 \left( \frac{\eta(M_r) + \log^3 n}{n} \right) \frac{p \log n}{n} \lesssim p \leq 1.$$

Then, except with a probability of at most $\varepsilon^{-\alpha} + o(n^{-19})$, the output of MC$^2$, namely $\tilde{M}$, satisfies the following:

$$\|\tilde{M} - M\|_F \lesssim \frac{\|M_i\|_F}{l} + \frac{\delta \sqrt{n} h^2}{l}. \tag{22}$$

A few remarks are in order regarding Corollary 3.

**Remark 4. [Improvement over SMC]** For the sake of simplicity, suppose that $M$ is rank-$r$ (so that $M = M_r$, $M^+ = 0$, $\gamma_r(M) = 0$) and that $\delta = 0$ (noise free). Then, Corollary 7 guarantees that $\tilde{M} = M$ with high probability if

$$\alpha^2 \kappa_r(M)^4 \left( \frac{\eta(M_r) + \log^3 n}{n} \right) \frac{p \log n}{n} \lesssim p \leq 1. \tag{23}$$

Since, in expectation, MC$^2$ takes $pn^2$ measurements, (23) implies that, loosely speaking, MC$^2$ succeeds with

$$O(1) \cdot \alpha^2 \kappa_r(M)^4 \left( \frac{\eta(M_r) + \log^3 n}{n} \right) r n \log n$$

samples. In contrast, SMC requires $O(1) \cdot \eta(M) r n \log^2 n$ samples to complete $M$ (see Proposition 1). By comparing the two sampling budgets, we find that MC$^2$ has a smaller sample complexity than SMC if

$$\eta(M_r) \gtrsim \log^3 n, \quad \kappa_r(M)^4 \lesssim \log n.$$  

In words, MC$^2$ improves over SMC for coherent but well-conditioned matrices. In practice, we find that the condition number of $M$ is inconsequential and likely an artifact of the proof techniques: in our simulations in Section 5, MC$^2$ substantially outperforms SMC for coherent matrices regardless of their condition number. On the other hand, up to a factor of $1/2$, MC$^2$ spends the same budget on uniform sampling as SMC does. Since $l \geq \frac{\delta}{\kappa}$, comparing (12) and (22) therefore implies that MC$^2$ is never outperformed by SMC by more than the constant factor.

**Remark 5. [Large leverage scores]** According to (21), for MC$^2$ to succeed, the number of large leverage scores must be small. There is nothing special about the threshold $\eta(M_r)/\log^2 n$ in (21) and, in the more general Theorem 10, an arbitrary threshold is considered. Essentially, this requirement guarantees that the sampling rate in Phase I of MC$^2$ is large enough to correctly estimate all significant leverage scores.

**Remark 6.** In numerical simulations, we set $C_1 = C_2 = 1$ in (17) as we found that it led to better empirical performance.
4 Estimating Leverage Scores

A byproduct of the analysis of MC$^2$ (more specifically, Lemma 12 followed by the union bound) is a simple recipe for estimating the leverages scores of a matrix given a fraction of its entries, which is summarized below (without proof).

**Corollary 7.** For a matrix $M \in \mathbb{R}^{n \times n}$, noise level $\delta \geq 0$, and probability $p \in (0,1]$, set $Y_p = R_p(M + D)$ for noise matrix $D \in \mathbb{R}^{n \times n}$ and with $\|R_p(D)\|_F \leq \delta$. Take

$$
\bar{\nu}_{r,i} := \left( \frac{C_1 \|Y_p[i,\cdot]\|_2^2}{\sigma_r^2(Y_p)} + \frac{C_2 \gamma_r(M)^2}{p^2} + \frac{C_1 \delta^2}{\sigma_r^2(Y_p)} \right) \frac{pm}{r}, \quad i \in [1:n],
$$

$$
\bar{\nu}_{r,j} := \left( \frac{C_1 \|Y_p[\cdot,j]\|_2^2}{\sigma_r^2(Y_p)} + \frac{C_2 \gamma_r(M)^2}{p^2} + \frac{C_1 \delta^2}{\sigma_r^2(Y_p)} \right) \frac{pm}{r}, \quad j \in [1:n],
$$

where $C_1 = \frac{147}{n}$, $C_2 = 6$. Then, $\bar{\nu}_{r,i} \geq \nu_r(M_r)$ and $\bar{\nu}_{r,j} \geq \nu_j(M_r)$, for all $i,j \in [1:n]$, provided that

$$
p \geq \alpha^2 \kappa_r(M)^2 \frac{\eta(M) \log^2 n}{n}, \quad \gamma_r(M) \leq \frac{p}{4}, \quad \delta \leq \frac{\sigma_r(M)}{4},
$$

and except with a probability of at most $n^{-\alpha}$.

5 Numerical Simulations

In this section, we compare MC$^2$ (Figure 2) with SMC (see Proposition 1). For the sake of brevity, we take $n = 50$ and $r = 5$—the results are similar for different $n$ and $r$. First, let us assume that $M \in \mathbb{R}^{n \times n}$ with rank($M$) = $r$ is incoherent. More specifically, we set $M = M_r = U_r \Sigma_r V_r^*$ where $U_r, V_r \in \mathbb{R}^{n \times r}$ are generic matrices with orthonormal columns, and $\Sigma_r \in \mathbb{R}^{r \times r}$ is diagonal with nonzero entries drawn from the uniform distribution on the interval $(0,1]$. Lastly, for easier comparison, $M$ is normalized to have unit Frobenius norm. Typically, $M$ is incoherent, $\eta(M) \approx 1$. Note that $\gamma_r(M) = 0$ and we also set $\delta = 0$ (noise free). Then, we compare the two algorithms when the overall sampling budget $p$ varies in the interval $(0,1]$. (A sampling budget of $p$ translates to observing $pm^2$ entries of $M$, in expectation.) The results, depicted in Figure 1a, report the average (over 100 trials) relative recovery error (in Frobenius norm). This suggests that the two algorithms perform similarly for incoherent matrices.

Next, let us compare the two algorithms for coherent matrices. This time, we set $M' = DMD$, where the nonzero entries of the diagonal matrix $D \in \mathbb{R}^{n \times n}$ follow the power law, $D[i,i] \propto i^{-1}$, $i \in [1:n]$. Typically, $M'$ is highly coherent, $\eta(M') \gg 1$. Again, with $\delta = 0$, Figure 1b compares the two algorithms for various budgets. Note that MC$^2$ meaningfully improves over SMC. We also repeat all experiments when $D[i,i] \propto i^{-2}$ and $\eta(M') \approx \frac{2}{n}$. See Figure 1c. In Figure 3, we repeat these experiments with ill-conditioned matrices, by replacing the $i$th singular value $\sigma_i$ of $M$ above with $e^{10(\sigma_i^{-1})}$. The average condition number was about $8 \cdot 10^4$.

Lastly, we compare the two algorithms when the spectral gap and noise level are nonzero. Figures 4 and 5 summarize the results. In particular, to generate the noisy example, entries of $M$ were perturbed by additive white Gaussian noise with standard deviation of $10^{-3}$.

6 Related Work

Matrix completion is an active research topic with a myriad of practical applications, and the large body of related literature includes [1, 11, 3, 5, 12, 8, 4, 13, 14, 15, 2, 16]. In particular, it is worth noting that [6] recently improved the logarithmic factor in Proposition 1 from $\log^2 n = \log n$, under different coherence assumptions (and the new result is optimal up to a constant factor). It remains to be seen if Proposition 2 and Corollary 3 here may also be improved under similar coherence assumptions. An extension of the ideas in [4] to tensor completion recently appeared in [17].
Leverage scores are of particular interest in numerical linear algebra. For instance, in a large linear regression problem, working with a randomly-selected small row-subset of the design matrix will significantly improve the processing time, without adversely affecting the performance. On this front, a few relevant references are [18, 19, 20, 21, 22, 23], where estimation of leverage scores is discussed. It is worth pointing out that, rather than row and column norms used in Corollary 7, one may alternatively utilize the column and row subspaces of $Y_p$ (the measurement matrix), after truncation, to estimate the leverage scores. This, however, evidently leads to an additive (rather than multiplicative) error bounds (akin to [22]) which are in fact not suited for the analysis of $MC^2$ here.

**Acknowledgments**

The authors thank Dehui Yang at Colorado School of Mines for helpful discussions in early phases of this project. Part of this research was conducted when AE was a visitor at the Institute for Computational and Experimental Research in Mathematics (ICERM). AE is grateful for their hospitality and kindness.
Figure 4: Comparison of MC\(^2\) and SMC when $\gamma_r(M) = 0.1$ and $\delta = 0$.

References


(a) Incoherent, $\eta(M_r) \approx 1$

(b) Coherent, $\eta(M_r) \gg 1$

(c) Coherent, $\eta(M_r) \gg 1$

Figure 5: Comparison of MC$^2$ and SMC when $\gamma_r(M) = 0$ and noise level (per entry) is $10^{-3}$.


We begin with a few observations about the measurement operator. For arbitrary \( A \in \mathbb{R}^{n \times n} \), note that

\[
\| \mathcal{R}_P(A) \|_F = \left\| \sum_{i,j} \frac{\xi_{i,j}}{P_{i,j}} \cdot A[i,j] \cdot E_{i,j} \right\|_F \tag{14}
\]

\[
\leq \sqrt{\sum_{i,j} \frac{A[i,j]^2}{P[i,j]^2}} \leq \frac{1}{\ell} \| A \|_F, \quad \left( \min_{i,j} P[i,j] \geq \ell \right)
\]

which implies that

\[
\| \mathcal{R}_P(\cdot) \|_{F \to F} \leq \ell^{-1}. \tag{24}
\]
Here, $\| \cdot \|_{F \rightarrow F}$ is the operator norm (from Frobenius norm to itself). For arbitrary $A \in \mathbb{R}^{n \times n}$, note also that
\[
\|R_P(A)\|_F^2 = \sum_{i,j} \frac{\epsilon_{i,j}}{P[i,j]^2} \cdot |A[i,j]|^2 \\
\geq h^{-1} \sum_{i,j} \frac{\epsilon_{i,j}}{P[i,j]} \cdot |A[i,j]|^2 \quad \left( \max_{i,j} P[i,j] \leq h \right) \\
= h^{-1} \langle R_P(A), A \rangle ,
\]
or, equivalently, $R_P^2(\cdot) \geq h^{-1} R_P(\cdot)$. Let $\Omega \subset [1 : n]^2$ be the random index set corresponding to the support of the measurement operator $R_P(\cdot)$ and set $P_\Omega(\cdot)$ to be the orthogonal projection onto the span of matrices supported on $\Omega$. For arbitrary $A \in \mathbb{R}^{n \times n}$, it then holds that
\[
\|R_P(A)\|_F = \sqrt{\sum_{i,j} \frac{\epsilon_{i,j} A[i,j]^2}{P[i,j]^2}} \\
\geq h^{-1} \|P_\Omega(A)\|_F . \quad \left( \max_{i,j} P[i,j] \leq h \right) \quad (26)
\]
The proof of Proposition 2 requires only a simple modification of the proof of Theorem 7 in [5], which we include for the sake of completeness. Let $H = \tilde{M} - M$ denote the error and note that
\[
H = P_\Omega(H) + P_{\Omega^C}(H). \quad (27)
\]
By feasibility of both $M$ and $\tilde{M}$ in Program (15), we find the tube constraint:
\[
h^{-1} \|P_\Omega(H)\|_F \leq \|R_P(H)\|_F \quad \text{(see (26))} \\
\leq \|R_P(\tilde{M}) - Y_P\|_F + \|R_P(M) - Y_P\|_F \quad \text{(triangle inequality)} \\
\leq 2h. \quad \text{(see (15))} \quad (28)
\]
Therefore, it remains to control $\|P_{\Omega^C}(H)\|_F$, where $\Omega^C$ is the complement of the index set $\Omega$. Let the subspace $T_r \subset \mathbb{R}^{n \times n}$ be the “support” of $M_r$, defined as
\[
T_r := \{ Z \in \mathbb{R}^{n \times n} : P_{\text{span}(M_r)} \cdot Z + Z \cdot P_{\text{span}(M_r^*)} - P_{\text{span}(M_r)} \cdot Z \cdot P_{\text{span}(M_r^*)} \}
\]
and let $P_{T_r}(\cdot)$ denote the orthogonal projection onto the subspace $T_r$. Since
\[
\|P_{\Omega^C}(H)\|_F^2 = \|P_{T_r} \cdot P_{\Omega^C}(H)\|_F^2 + \|P_{T_r^\perp} \cdot P_{\Omega^C}(H)\|_F^2 ,
\]
it suffices to control each norm on the right-hand side. Here, $T_r^\perp$ is the orthogonal complement of $T_r$. A byproduct of the proof of Theorem 2 in [4] is that
\[
\|M_r\|_* + \frac{1}{2} \|P_{T_r^\perp}(A)\|_* \leq \|M_r + A\|_* , \quad (31)
\]
for any $A \in \mathbb{R}^{n \times n}$ such that $P_\Omega(A) = 0$, under the conditions listed in Proposition 2, and except with a probability of $o(n^{-19})$. In particular, with the choice of $A = P_{\Omega^C}(H)$, we find that
\[
\|M\|_* - \|M_r + H\|_* + \frac{1}{2} \|P_{T_r^\perp} \cdot P_{\Omega^C}(H)\|_* \quad (32)
\]
\[
\leq \|M_r\|_* + \frac{1}{2} \|P_{T_r^\perp} \cdot P_{\Omega^C}(H)\|_* \quad (M = M_r + M_r + \text{and triangle inequality}) \\
\leq \|M_r + P_{\Omega^C}(H)\|_* \quad \text{(see (31))} \\
\leq \|M_r + H\|_* + \|P_{\Omega^C}(H)\|_* \quad \text{(27) and triangle inequality) \\
\leq \|M_r + H\|_* + \|M_r + H\|_* + \|P_{\Omega^C}(H)\|_* \quad (M = M_r + M_r + \text{and triangle inequality}) \\
= \|M_r\|_* + \|M_r\|_* + \|P_{\Omega^C}(H)\|_* \quad \text{(optimality of $\tilde{M}$ in Program (15))} \\
\leq \|M\|_* + \|M_r\|_* + \|P_{\Omega^C}(H)\|_* \quad \text{(see (28))} \quad (33)
\]
which immediately yields that
\[ \|P_{T^*}P_{\Omega^c}(H)\|_F \leq \|P_{T^*}P_{\Omega^c}(H)\|_\infty \leq 4\|M_r\|_\infty + \sqrt{n} \cdot 4h\delta. \]  
(34)

It remains to control the first norm on the right-hand side of (30). Another byproduct of the proof of Theorem 2 in [4] is that
\[ \|(P_T^*R_T^*P_T - P_{T^*})\cdot\|_{F \to F} \leq \frac{1}{2}, \]  
(35)

under the conditions in Proposition 2 and except with a probability of \(o(n^{-19})\). Using this relation, we observe that
\[ \|R_P\cdot F_P (H)\|_F = (R_P\cdot F_P (H), P_T, P_{\Omega^c} (H)) \quad (R_P\cdot F_P (H) \text{ is self-adjoint}) \]
\[ \geq h^{-1} (R_P\cdot F_P (H), P_T, P_{\Omega^c} (H)) \quad (\text{see (25)}) \]
\[ = h^{-1} (P_T, R_P\cdot F_P (H), P_T, P_{\Omega^c} (H)) + h^{-1} (P_T, P_{\Omega^c} (H), P_{\Omega^c} (H)) \]
\[ \geq -h^{-1} (\|P_T, R_P\cdot F_P (H), P_T, P_{\Omega^c} (H)\|_{F \to F} \cdot \|P_{\Omega^c} (H)\|_F^2 + h^{-1} \|P_T, P_{\Omega^c} (H)\|_F^2 \]
\[ = \frac{1}{2h} \|P_T, P_{\Omega^c} (H)\|_F^2. \quad \text{(see (35))} \]
(36)

In light of the above inequality, we focus on controlling \(\|R_P\cdot F_P (H)\|_F\) instead. Note that
\[ \|R_P\cdot F_P (H)\|_F = \|R_P\cdot F_P (H) - R_P\cdot F_P (H)\|_F \]
\[ \leq l^{-1} \|P_T, P_{\Omega^c} (H)\|_F \quad (\text{see (24)}) \]
\[ \leq 4l^{-1} (\|M_r\|_\infty + \sqrt{n} \cdot h\delta). \quad \text{(see (34))} \]
(37)

Overall, we conclude that
\[ \|P_T, P_{\Omega^c} (H)\|_F \leq \sqrt{2h} \|R_P\cdot F_P (H)\|_F \quad \text{(see (36))} \]
\[ \leq \frac{4\sqrt{2h}}{l} (\|M_r\|_\infty + \sqrt{n} \cdot h\delta). \quad \text{(see (37))}. \]
(38)

Putting everything back together, we find that
\[ \frac{1}{2} \|H\|_F^2 = \|P_{\Omega^c} (H)\|_F^2 + \|P_{\Omega^c} (H)\|_F^2 \]
\[ \leq (2h\delta)^2 + \|P_T, P_{\Omega^c} (H)\|_F^2 + \|P_{T^*}, P_{\Omega^c} (H)\|_F^2 \quad \text{(see (34))} \]
\[ \leq (2h\delta)^2 + \|P_{T^*}, P_{\Omega^c} (H)\|_F^2 + 4\|M_r\|_\infty + \sqrt{n} \cdot 4h\delta \]
\[ \leq (2h\delta)^2 + \frac{32h^3}{l^2} \left(\|M_r\|_\infty + \sqrt{n} \cdot h\delta\right)^2 + 4\|M_r\|_\infty + \sqrt{n} \cdot 4h\delta \]
\[ \leq \frac{h^2}{l^2} \|M_r\|_\infty^2 + \frac{n\delta^2}{l^2}, \quad \text{(39)} \]

which completes the proof of Proposition 2.

B Analysis of MC^2

In this section, we study the performance of MC^2 (see Figure 2) and, in particular, prove Corollary 3. Throughout, we assume that \(\gamma_r(\hat{M}) \leq \gamma_r\). Recall that our sampling budget of \(p\) (per entry) is split two- ways: \(\lambda \cdot p\) is spent in Phase 1 of MC^2 (with \(\lambda = \frac{1}{2}\)), where the relatively large leverage scores of \(M_r\) are estimated, and the rest is invested in Phase II of MC^2, where important entries (namely those corresponding to the large leverage scores) are sampled and \(M\) is then completed. After pointing out that the analysis in this appendix is valid for arbitrary \(\lambda \in (0, 1)\), we fix \(\lambda = \frac{1}{2}\) to match MC^2.

First, in Appendix C, we show that MC^2 correctly estimates the spectrum of \(M_r\), if the sampling budget is sufficiently large. The proof involves a simple application of the standard large-deviation bounds.
Lemma 8. For $M \in \mathbb{R}^{n \times n}$, set $Y_{\lambda p} = \mathcal{R}_{\lambda p}(M) = \mathcal{R}_{\lambda p}(M + D)$ with $\|\mathcal{R}_{\lambda p}(D)\|_F \leq \delta$ and $\delta \geq 0$. Then, except with a probability of at most $e^{-\alpha}$, it holds that
\[
\frac{\kappa_r(M)}{7} \leq \kappa_r(Y_{\lambda p}) \leq 7 \kappa_r(M), \quad \frac{\sigma_r(M)}{4} \leq \sigma_r(Y_{\lambda p}) \leq \frac{7 \sigma_r(M)}{4},
\]
provided that
\[
\lambda p \gtrsim \alpha^2 \kappa_r(M)^2 \cdot \eta(M) \frac{r \log n}{n}, \quad \gamma_r \leq \frac{\lambda p}{\delta \sigma_r(M)}, \quad \delta \leq \frac{\sigma_r(M)}{4}.
\]

Lemma 8, in turn, helps us take the second step: In Appendix D, we show that $MC^2$ correctly upper-estimates the leverage scores of $M_r$, provided that the number of large leverage scores, namely those exceeding $\beta^2 \eta(M_r)$, is controlled. Here, $\beta \in (0, 1]$ is a threshold to be set later.

Lemma 9. Consider a matrix $M \in \mathbb{R}^{n \times n}$ and its possibly noisy copy $\tilde{M} = M + D \in \mathbb{R}^{n \times n}$, where $\|\mathcal{R}_{\lambda p}(D)\|_F \leq \delta$ with $\delta \geq 0$. Apply $MC^2$ and compute $(\hat{\mu}_{r,i}, \hat{\nu}_{r,j})$. Also, for $\beta \in (0, 1]$, set
\[
\hat{\mu}_{r,i} := \beta^2 \eta(M_r) + (\hat{\mu}_{r,i} - \beta^2 \eta(M_r))_+ \quad \text{and} \quad \hat{\nu}_{r,j} = \beta^2 \eta(M_r) + (\hat{\nu}_{r,j} - \beta^2 \eta(M_r))_+ \quad \text{for} \quad i \in [1 : n],
\]
and define $\{\hat{\nu}_{r,j}\}$ similarly. Assume that
\[
\# \{i : \mu_i(M_r) \vee \nu_i(M_r) \geq \beta^2 \eta(M_r)\} = O \left( n^3 \right).
\]
Then, except with a probability of at most $e^{-\alpha} + n^{-\alpha}$, it holds that
\[
\max_{i,j \in [1 : n]} (\hat{\mu}_{r,i} + \hat{\nu}_{r,j}) \lesssim \alpha \cdot \kappa_r(M)^2 \eta(M_r) + \frac{\gamma_r n}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r},
\]
provided that
\[
\lambda p \gtrsim \alpha^2 \kappa_r(M)^4 \max \left[ \beta^2 \log n, 1 \right] \cdot \frac{\eta(M_r) r \log n}{n}, \quad \gamma_r \leq \frac{\lambda p}{\delta \frac{\sigma_r(M)}{4}}, \quad \delta \leq \frac{\sigma_r(M)}{4}.
\]

Below, we assume that both Lemmas 8 and 9 are in force. Assume also that $\sum_{i,j} P[i, j] \leq \Pi(p, \frac{\xi}{2}) \cdot n^2$, so as not to exceed the sampling budget of $pn^2$.

By definition of probability matrices $P, P' \in [0, 1]^{n \times n}$ in $MC^2$ and (16), each entry $\tilde{M}[i, j]$ is observed in $MC^2$ with a probability of $P[i, j]$ where
\[
P[i, j] = \lambda p + \left( \alpha^2 (\hat{\mu}_{r,i} + \hat{\nu}_{r,j}) \frac{r \log n}{n} - \lambda p \right)_+ \quad \text{for} \quad i, j \in [1 : n].
\]
That is, the measurement matrix $Y$ and the corresponding index set $\Omega$ in Step 4 of $MC^2$ might have been equivalently generated with an application of $\mathcal{R}_P(\cdot)$:
\[
Y = Y_P = \mathcal{R}_P \left( \tilde{M} \right) = \sum_{[i, j] \in [1 : n]} \frac{\epsilon_{i,j}}{P[i, j]} \cdot \tilde{M}[i, j] \cdot E_{i,j}.
\]

Next, we show that each $P[i, j]$ is large whenever the corresponding entry $M_r[i, j]$ is important, namely $MC^2$ effectively performs leveraged sampling. To that end, first note that
\[
P[i, j] = \lambda p + \left( \alpha^2 (\hat{\mu}_{r,i} + \hat{\nu}_{r,j}) \frac{r \log n}{n} - \lambda p \right)_+ \quad \text{(see (42))}
\]
\[
\gtrsim \frac{4 \alpha^2 \beta^2 \eta(M_r) r \log n}{n} + \left( \alpha^2 (\hat{\mu}_{r,i} + \hat{\nu}_{r,j}) \frac{r \log n}{n} - \frac{4 \alpha^2 \beta^2 \eta(M_r) r \log n}{n} \right)_+ \quad \text{(see (41))}
\]
\[
= \frac{2 \alpha^2 r \log n}{n} \left[ 2 \beta^2 \eta(M_r) + (\hat{\mu}_{r,i} + \hat{\nu}_{r,j} - 2 \beta^2 \eta(M_r))_+ \right],
\]

for every \( i, j \in [1 : n] \). In the last line above, there are four possibilities: \( \overline{\mu}_{r,i}, \overline{\nu}_{r,j} \geq \beta^2 \eta(M_r) \), \( \overline{\mu}_{r,i}, \overline{\nu}_{r,j} \leq \beta^2 \eta(M_r) \), and so forth. In all of these cases, we may verify that

\[
P[i,j] \geq \frac{2\alpha^2 r \log^2 n}{n} \left[ 2\beta^2 \eta(M_r) + \left( \overline{\mu}_{r,i} + \overline{\nu}_{r,j} - 2\beta^2 \eta(M_r) \right)_+ \right] \quad \text{(see (44))}
\]

\[
\geq \frac{2\alpha^2 r \log^2 n}{n} \cdot \frac{\mu_i(M_r) + \nu_j(M_r)}{2} \quad \text{(see (40))}
\]

\[
\geq \frac{2\alpha^2 r \log^2 n}{n} \cdot \frac{\mu_i(M_r) + \nu_j(M_r)}{2}, \quad \text{(see Lemma 9) (45)}
\]

for every \( i, j \in [1 : n] \), namely MC\(^2\) successfully performs leveraged sampling. Additionally, to later control the overall number of measurements in MC\(^2\), let us record that

\[
\sum_{i,j=1}^n P[i,j] = \sum_{P[i,j] \leq \lambda p} \lambda p + \sum_{P[i,j] > \lambda p} P[i,j] \quad \text{(see (42))}
\]

\[
\leq \sum_{P[i,j] \leq \lambda p} \lambda p + \sum_{P[i,j] > \lambda p} \alpha^2 (\overline{\mu}_{r,i} + \overline{\nu}_{r,j}) \frac{r \log^2 n}{n} \quad \text{(see (42))}
\]

\[
\leq \lambda p n^2 + \sum_{\overline{\mu}_{r,i} + \overline{\nu}_{r,j} \geq \beta^2 \eta_r(M)} \alpha^2 (\overline{\mu}_{r,i} + \overline{\nu}_{r,j}) \frac{r \log^2 n}{n} \quad \text{(see the text below)}
\]

\[
\leq \lambda p n^2 + \frac{\alpha^4}{\beta^2} \left( \kappa_r(M)^2 + \frac{\gamma_r^2 n}{\lambda pr} + \frac{\delta^2 \lambda p m}{\sigma_r(M)^2 r} \right)^2 \log^2 n. \quad \text{(see Lemma 9) (46)}
\]

In the fifth line above, we argued that \( P[i,j] > \lambda p \) implies \( \overline{\mu}_{r,i} \lor \overline{\nu}_{r,j} > \beta^2 \eta_r(M) \). Otherwise, by (40), it must be the case that \( \overline{\mu}_{r,i} = \overline{\nu}_{r,j} = \beta^2 \eta_r(M) \) and \( \overline{\mu}_{r,i} \lor \overline{\nu}_{r,j} \leq \beta^2 \eta(M_r) \), and, consequently, it must be the case that

\[
\alpha^2 (\overline{\mu}_{r,i} + \overline{\nu}_{r,j}) \frac{r \log^2 n}{n} \leq 2\beta^2 \frac{\eta(M_r)}{r} \log^2 n \leq \lambda p. \quad \text{(41) with a large enough constant)}
\]

In turn, by (42), it must be the case that \( P[i,j] = \lambda p \), which is a contradiction, thereby verifying the fifth line of (46). Recall (43) and suppose that \( \|r_P(D)\|_F \leq \delta \). Let \( \widetilde{M} \) be a solution of Program (15) with \( P \in (0, 1]^{n \times n} \) specified in (42). Also set \( l = \min_{i,j} P[i,j] \) and \( h = \max_{i,j} P[i,j] \). An application of Proposition 2 now implies that

\[
\|\widetilde{M} - M\|_F \lesssim \frac{1}{l} \|M_r + \| + \frac{\delta \sqrt{\eta h^2}}{l},
\]

except with a probability of \( e^{-\alpha + n^{-\alpha}} + o(n^{-9}) \). Thus, we arrive at the following result, of which Corollary 3 is a special case with \( \beta = 1 / \log n \).

**Theorem 10.** Set \( \lambda = \frac{1}{2} \). Consider a matrix \( M \in \mathbb{R}^{n \times n} \) and its possibly noisy copy \( \widetilde{M} = M + D \in \mathbb{R}^{n \times n} \). Apply MC\(^2\) with input parameters \( r, \gamma_r, p, \alpha \) as prescribed in Figure 2. Suppose that \( \|r_P(D)\|_F \leq \delta \) and \( \|r_P(D)\|_F \leq \delta \) (see (10), (14), and (18)). Suppose that

\[
\gamma_r(M) \leq \gamma_r \leq \frac{\lambda p}{4}, \quad \delta \leq \frac{\sigma_r(M)}{4}.
\]

17
For fixed $\alpha > 0$ and $\beta \in (0, 1]$, assume also that
\[
\# \{ i : \mu_i(M_r) \vee \nu_i(M_r) \geq \beta^2 \eta(M_r) \} = O(n^\beta).
\]
Lastly, with
\[
\lambda_p \gtrsim \alpha^2 \kappa_r(M)^4 \cdot \max \left\{ \beta \log n, 1 \right\} \cdot \frac{\eta(M_r)r \log n}{\eta(M_r)}
\]
avoid that
\[
(1 - \lambda)p \gtrsim \frac{\alpha^2}{\beta^2} \left( \kappa_r(M)^2 + \frac{\gamma^2 n}{\lambda pr} + \frac{\delta^2 \lambda_p n}{\sigma_r(M)^2 r^2} \right)^2 \text{rn} \log^2 n,
\]
with appropriate constants. Then, except with a probability of at most $e^{-\alpha}$, $n^{-\alpha^3} + o(n^{-19})$, the output of $MC^2$, namely $\hat{M}$, satisfies
\[
\| \hat{M} - M \|_F \lesssim \left\| M_r^* \right\|_{l} + \frac{\delta \sqrt{\eta r}}{l},
\]
where $l = \min_i \{ P[i,j] \}$ and $h = \max_{i,j} P[i,j]$.

C Proof of Lemma 8

Let us recall the following standard matrix concentration inequality [24].

**Lemma 11. [Bernstein’s inequality for spectral norm]** Let $(A_i) \subset \mathbb{R}^{n \times n}$ be a finite sequence of zero-mean independent random matrices. Set
\[
b := \max_i \| A_i \|, \quad \sigma^2 := \left\| \sum_i \mathbb{E}[A_i^*A_i] \right\| \vee \left\| \sum_i \mathbb{E}[A_i^*A_i] \right\|.
\]
Then, except with a probability of at most $e^{-\alpha}$, it holds that
\[
\left\| \sum_i A_i \right\| \lesssim \alpha \max \left\{ \log n \cdot b, \sqrt{\log n} \cdot \sigma \right\}.
\]

Next, we observe that
\[
R_{\lambda_p}(M_r) = \sum_{i,j} \frac{\xi_{i,j}}{\lambda_p} \cdot M_r[i,j] \cdot E_{i,j}, \quad \text{(see (10))}
\]
\[
R_{\lambda_p}(M_r) - M_r = R_{\lambda_p}(M_r) - \mathbb{E}[R_{\lambda_p}(M_r)]
\]
\[
= \sum_{i,j} \frac{\xi_{i,j} - p}{p} \cdot M_r[i,j] \cdot E_{i,j}
\]
\[
= : \sum_{i,j} A_{i,j}, \quad \text{(50)}
\]
where $(A_{i,j}) \subset \mathbb{R}^{n \times n}$ are zero-mean independent random matrices. To apply Lemma 11, we make the following calculations:
\[
b \leq \| A_{i,j} \|
\]
\[
\leq \frac{1}{\lambda_p} \cdot \max_{i,j} |M_r[i,j]| \quad \text{(see (50))}
\]
\[
\leq \frac{1}{\lambda_p} \cdot \max_i \| U_r[i, :] \|_2 \cdot \sigma_1(M) \cdot \max_j \| V_r[j, :] \|_2 \quad (M_r = U_r \Sigma_r V_r^*)
\]
\[
= \frac{1}{\lambda_p} \cdot \frac{\eta(M_r)^r}{n} \cdot \sigma_1(M), \quad \text{(see (9))}
\]
\[
\sum_{i,j} \mathbb{E} \left[ A_{i,j} A_{i,j}^* \right] = \sum_{i,j} \mathbb{E} \left[ \frac{(\epsilon_{i,j} - \lambda p)^2}{\lambda^2 p^2} \right] |M_r[i,j]|^2 \cdot E_{i,i} \quad \text{(see (50))}
\]
\[
= 1 - \frac{\lambda p}{\lambda p} \sum_{i} E_{i,i} \left( \sum_{j} |M_r[i,j]|^2 \right) \quad (\epsilon_{i,j} \sim \text{Bernoulli}(\lambda p))
\]
\[
= 1 - \frac{\lambda p}{\lambda p} \sum_{i} \|M_r[i,:]\|^2 \cdot E_{i,i}
\]
\[
= 1 - \frac{\lambda p}{\lambda p} \sum_{i} \|M_r[i,:]\|^2 \cdot E_{i,i}
\]
\[
= 1 - \frac{\lambda p}{\lambda p} \cdot \text{diag}(M_r M_r^*) ,
\]
\[
\left\| \sum_{i,j} \mathbb{E} \left[ A_{i,j} A_{i,j}^* \right] \right\| = \frac{1 - \lambda p}{\lambda p} \cdot \text{max}_i \|M_r[i,:]\|^2
\]
\[
\leq \frac{1}{\lambda p} \cdot \text{max}_i \|U_r[i,:]\|^2 \cdot \sigma_1^2(M) \quad (M_r = U_r \Sigma_r V_r^*)
\]
\[
\leq \frac{1}{\lambda p} \cdot \frac{\eta(M_r)r}{n} \cdot \sigma_1^2(M) , \quad \text{(see (9))}
\]

and, similarly,
\[
\left\| \sum_{i,j} \mathbb{E} \left[ A_{i,j}^* A_{i,j} \right] \right\| \leq \frac{1 - \lambda p}{\lambda p} \cdot \frac{\eta(M_r)r}{n} \cdot \sigma_1^2(M) ,
\]

so that
\[
\sigma^2 = \left\| \sum_{i} \mathbb{E} \left[ A_i A_i^* \right] \right\| \vee \left\| \sum_{i} \mathbb{E} \left[ A_i^* A_i \right] \right\| \leq \frac{\eta(M_r)r}{\lambda pn} \cdot \sigma_1^2(M) .
\]

Therefore, except with a probability of at most \(e^{-\alpha}\), it holds that
\[
\left\| R_{\lambda p} (M_r) - M_r \right\| \leq \left\| \sum_{i,j} A_{i,j} \right\| \leq \alpha \max \left[ \log n \cdot b, \sqrt{\log n} \cdot \sigma \right] \quad \text{(see Lemma 17)}
\]
\[
\leq \alpha \max \left[ \log n \cdot \frac{\eta(M_r)r}{\lambda pn} \cdot \sigma_1(M), \sqrt{\log n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda pn} \cdot \sigma_1(M)} \right]
\]
\[
= \alpha \cdot \sigma_1(M) \sqrt{\log n \cdot \frac{\eta(M_r)r}{\lambda pn}} ,
\]

when \(\lambda p\) is sufficiently large. The above inequality yields that
\[
\left\| R_{\lambda p} (M_r) - M_r \right\| \leq \frac{\sigma_r(M)}{4} , \quad \text{(51)}
\]

provided that
\[
\lambda p \geq \alpha^2 \cdot \kappa_r(M)^2 \cdot \frac{\eta(M_r)r \log n}{n} , \quad \text{(52)}
\]
and except with a probability of at most $e^{-\alpha}$. Moreover, note that
\[
\|Y_{\lambda p} - M_r\| = \|R_{\lambda p}(M + D) - M_r\|
\leq \|R_{\lambda p}(M) - M_r\| + \delta \quad (\|R_{\lambda p}(D)\|_F \leq \delta)
\leq \|R_{\lambda p}(M_r) - M_r\| + \|R_{\lambda p}(M_{r+})\| + \delta \quad (M = M_r + M_{r+})
\leq \frac{\sigma_r(M)}{4} + \frac{\|M_r\|}{\lambda p} + \delta \quad \text{(see (51) and the text below)}
\]
\[
= \frac{\sigma_r(M)}{4} + \frac{\sigma_{r+1}(M)}{\lambda p} + \delta
\leq \frac{\sigma_r(M)}{4} + \frac{\sigma_r(M)}{4} + \delta \quad \left(\text{if } \gamma_r(M) = \frac{\sigma_{r+1}(M)}{\sigma_r(M)} \leq \frac{\lambda p}{4}\right)
\]
(54)
Above, the fourth line above also uses the fact that
\[
\|R_{\lambda p}(Z)\| \leq \frac{\|Z\|}{\lambda p}, \quad \forall Z \in \mathbb{R}^{n \times n},
\]
(55)
which follows from (10) and the fact that the map $M \to \sum_{i,j} \varepsilon_{i,j} M[i,j] E_{i,j}$ is an orthogonal projection (and is therefore non-expansive). Then, an application of the Weyl's inequality to (54) leads to
\[
|\sigma_i(Y_{\lambda p}) - \sigma_i(M_r)| \leq \|Y_{\lambda p} - M_r\| \leq \frac{\sigma_r(M)}{2} + \delta \leq \frac{3\sigma_r(M)}{4}, \quad i \in [1 : r],
\]
provided that $\delta \leq \frac{\sigma_r(M)}{4}$. In particular, it follows that
\[
\frac{\kappa_r(M)}{7} \leq \kappa_r(Y_{\lambda p}) \leq 7\kappa_r(M), \quad \frac{\sigma_r(M)}{4} \leq \sigma_r(Y_{\lambda p}) \leq \frac{7\sigma_r(M)}{4}.
\]
This completes the proof of Lemma 8.

## D Proof of Lemma 9

Throughout this proof, we assume that $\gamma_r(M) \leq \gamma_r$. We begin with finding an upper bound for a leverage score. See Appendix E for the standard proof.

**Lemma 12.** Fix $i \in [1 : n]$ and $\alpha \geq 1$. Form $Y_{\lambda p} = R_{\lambda p}(M + D)$ with $\|R_{\lambda p}(D)\|_F \leq \delta$, and assume that
\[
\lambda p \geq \alpha^2 \kappa_r(M)^4 \cdot \frac{\eta(M_r)r}{n}.
\]
Then, except with a probability of at most $e^{-\alpha}$, it holds that
\[
\mu_i(M_r) \leq \tilde{\mu}_{r,i} \leq \kappa_r(M)^2 \mu_i(M_r) + \frac{\gamma_r^2 n}{\lambda pr} + \frac{\delta^2 \lambda pm}{\sigma_r(M)^2 r}, \quad i \in [1 : n],
\]
where
\[
\tilde{\mu}_{r,i} := 6 \left(\frac{\|Y_{\lambda p}[i,\cdot]\|_2}{\sigma_r(M)^2} + \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r(M)^2}\right) \frac{\lambda pm}{r}.
\]
(56)
A similar result holds for column leverage scores with $\{\tilde{\nu}_{r,j}\}_j$ (defined similarly) which involves norms of columns of $Y_{\lambda p}$.

For $\beta \leq 1$, set
\[
B := \left\{ i : \mu_i(M_r) \vee \nu_i(M_r) \geq \beta^2 \eta(M_r) \right\} \subset [1 : n],
\]
(57)
and, by assumption, note that
\[
\log (\#B) = O (\beta \log n).
\]
(58)
The next result, proved in Appendix F, gives a family of estimators for the leverage scores of $M_r$. 20
Lemma 13. For $\alpha \geq 1$, assume that
\[
\lambda p \gtrsim \alpha^2 \beta^2 \kappa_r(M)^4 \cdot \frac{\eta(M_r) r \log^2 n}{n}, \quad (59)
\]
Then, except with a probability of at most $n^{-\alpha \beta}$, it holds that
\[
\mu_i(M_r) \leq \beta^2 \eta(M_r) + (\tilde{\mu}_{r,i} - \beta^2 \eta(M_r))^+, \quad i \in [1:n]. \quad (60)
\]
A similar result holds for column leverage scores.

Furthermore, if Lemma 8 is in force, we can replace $\{\tilde{\mu}_{r,i}, \tilde{\nu}_{r,i}\}$, with estimates that do not depend on $M$ directly. This is stated below (without proof).

Lemma 14. For $\alpha \geq 1$, suppose that
\[
\lambda p \gtrsim \alpha^2 \kappa_r^4(M) \cdot \max [\beta^2 \log n, 1] \cdot \frac{\eta(M_r) r \log n}{n}, \quad \gamma_r \leq \frac{\lambda p}{4}, \quad \delta \leq \frac{\sigma_r(M)}{4}.
\]
Then, except with a probability of at most $n^{-\alpha \beta}$, it holds that
\[
\mu_i(M_r) \leq \tilde{\mu}_{r,i}, \quad i \in [1:n], \quad (62)
\]
where
\[
\tilde{\mu}_{r,i} := \beta^2 \eta(M_r) + (\bar{\mu}_{r,i} - \beta^2 \eta(M_r))^+, \quad (63)
\]
\[
\bar{\mu}_{r,i} := \left( \frac{C_1 \|Y_p[i,\cdot]\|_2^2}{\sigma_r(Y_p)^2} + \frac{C_2 \gamma_r^2}{\lambda p \delta^2} + \frac{C_2 \delta^2}{\sigma_r(Y_p)^2} \right) \frac{\lambda p n}{r}, \quad (64)
\]
for every $i \in [1:n]$. Also, $\{\bar{\nu}_{r,j}, \tilde{\nu}_{r,j}\}$, defined likewise, satisfy similar properties. Above, $C_1 = \frac{147}{8}$ and $C_2 = 6$.

In addition, let us record that
\[
\mu_{r,i} \lor \tilde{\nu}_{r,j} \geq \beta^2 \eta(M_r) \implies \tilde{\mu}_{r,i} \lor \tilde{\nu}_{r,j} = \tilde{\mu}_{r,i} \lor \bar{\nu}_{r,j}, \quad (65)
\]
\[
\{i : \tilde{\mu}_{r,i} \lor \tilde{\nu}_{r,j} \geq \beta^2 \eta(M_r)\} = \{i : \bar{\nu}_{r,i} \lor \bar{\nu}_{r,j} \geq \beta^2 \eta(M_r)\}, \quad (66)
\]
which we now use in the following argument:
\[
\sum_{\tilde{\mu}_{r,i} \lor \tilde{\nu}_{r,j} \geq \beta^2 \eta(M_r)} (\tilde{\mu}_{r,i} + \tilde{\nu}_{r,j}) \leq 2 \sum_{\tilde{\mu}_{r,i} \lor \tilde{\nu}_{r,j} \geq \beta^2 \eta(M_r)} \tilde{\mu}_{r,i} \lor \tilde{\nu}_{r,j}
\]
\[
= 2 \sum_{\bar{\mu}_{r,i} \lor \bar{\nu}_{r,j} \geq \beta^2 \eta(M_r)} \bar{\mu}_{r,i} \lor \bar{\nu}_{r,j} \quad \text{(see (65))}
\]
\[
= 2 \sum_{\bar{\nu}_{r,i} \lor \bar{\nu}_{r,j} \geq \beta^2 \eta(M_r)} \bar{\nu}_{r,i} \lor \bar{\nu}_{r,j} \quad \text{(see (66))}
\]
\[
\leq 2 \left( \max_i \bar{\nu}_{r,i} \lor \max_{j} \bar{\nu}_{r,j} \right) \cdot \left( \frac{\sum_{i,j=1}^n \bar{\nu}_{r,i} + \bar{\nu}_{r,j}}{\beta^2 \eta(M_r)} \right) \quad \text{(Markov’s inequality)}
\]
\[
= 2n \left( \max_i \bar{\nu}_{r,i} \lor \max_{j} \bar{\nu}_{r,j} \right) \cdot \left( \frac{\sum_{i,j=1}^n \bar{\nu}_{r,i} + \bar{\nu}_{r,j}}{\beta^2 \eta(M_r)} \right). \quad (67)
\]
In Appendix G, we control the maximum in the last line above.
Lemma 15. With \( \{\mathbf{p}_{r,i}, \mathbf{v}_{r,i}\}_i \) as defined in Lemma 14, it holds that
\[
\max_i \mathbf{p}_{r,i} \lor \max_j \mathbf{v}_{r,j} \lesssim \alpha \cdot \kappa_r(M)^2 \eta_r(M_r) + \frac{\gamma_r^2 n^2}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r},
\]
except with a probability of at most \( e^{-\alpha} + n^{-\alpha \beta} \), and when (61) is satisfied.

In Appendix H, we control the sum in the last line of (67).

Lemma 16. With \( \{\mathbf{p}_{r,i}, \mathbf{v}_{r,i}\}_i \) defined in Lemma 14, it holds that
\[
\sum_{i=1}^n \mathbf{p}_{r,i} \lor \mathbf{v}_{r,i} \lesssim \alpha \cdot \kappa_r(M)^2 n + \frac{n^2}{\lambda pr} + \frac{\delta^2 \lambda pn^2}{\sigma_r(M)^2 r},
\]
except with a probability of at most \( e^{-\alpha} \), and when (61) is satisfied.

Using these two lemmas, we now simplify the bound in (67) to read
\[
\sum_{\mathbf{p}_{r,i} \lor \mathbf{v}_{r,i} \geq \beta^2 \eta_r(M_r)} (\mathbf{p}_{r,i} + \mathbf{v}_{r,i}) \lesssim \sum_{i=1}^n \mathbf{p}_{r,i} \lor \mathbf{v}_{r,i} \beta^2 \eta_r(M_r) \quad \text{(see (67))}
\]
\[
\lesssim \frac{n^2}{\beta^2 \eta_r(M_r)} \left( \kappa_r(M)^2 \eta_r(M_r) + \frac{\gamma_r^2 n^2}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r} \right) \left( \alpha \cdot \kappa_r(M)^2 n + \frac{n^2}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r} \right) \quad \text{(see Lemmas 15 and 16)}
\]
\[
\lesssim \frac{\alpha^2 n^2}{\beta^2} \left( \kappa_r(M)^2 + \frac{\gamma_r^2 n^2}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r} \right)^2, \quad (\alpha, \eta(M_r) \geq 1)
\]
except with a probability of \( e^{-\alpha} + n^{-\alpha \beta} \). Also, a byproduct of Lemma 15 is that
\[
\mathbf{p}_{r,i} = \beta^2 \eta_r(M_r) + (\mathbf{p}_{r,i} - \beta^2 \eta(M_r)) \quad \text{(see (63))}
\]
\[
\leq \max \left[ \beta^2 \eta(M_r), \mathbf{p}_{r,i} \right] \lesssim \max \left[ \beta^2 \eta(M_r), \alpha \cdot \kappa_r(M)^2 \eta_r(M_r) + \frac{\gamma_r^2 n}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r} \right] \quad \text{(see Lemma 15)}
\]
\[
= \alpha \cdot \kappa_r(M)^2 \eta_r(M_r) + \frac{\gamma_r^2 n}{\lambda pr} + \frac{\delta^2 \lambda pn}{\sigma_r(M)^2 r}, \quad (\beta \leq 1)
\]
for every \( i \in \{1 : n\} \) and except with a probability of at most \( n^{-\alpha \beta} \). A similar bound holds for \( \{\mathbf{v}_{r,j}\}_j \). This completes the proof of Lemma 9.

E  Proof of Lemma 12

Let us first recall the following Bernstein inequality [25].

Lemma 17. [Bernstein inequality for Frobenius norm] Let \( \{A_i\} \subset \mathbb{R}^{n_1 \times n_2} \) be a finite sequence of zero-mean independent random matrices, and set
\[
b := \max_i \|A_i\|_F,
\]
\[
\sigma^2 := \sum_i \mathbb{E} \left[ \|A_i\|_F^2 \right].
\]

Then, for \( \alpha \geq 1 \) and except with a probability of at most \( e^{-\alpha} \), it holds that
\[
\left\| \sum_i A_i \right\|_F \lesssim \alpha \max[b, \sigma].
\]
We focus on the row leverage scores. With a slight abuse of notation, we set $Y_{\lambda p, r} := R_{\lambda p}(M_r)$ for short. Fix $i \in [1 : n]$ and note that

$$E \left[ \|Y_{\lambda p, r}[i, \cdot]\|_2^2 \right] = E \left[ \sum_j \frac{\epsilon_{i,j}}{\lambda^2 p^2} |M_r[i, j]|^2 \right] \quad ((10) \text{ and } \epsilon_{i,j} \in \{0, 1\})$$

and, consequently,

$$\sigma_r^2(M) \cdot \|U_r[i, \cdot]\|_2^2 \leq \|Y_{\lambda p, r}[i, \cdot]\|_2^2 \leq \sigma_1^2(M) \cdot \|U_r[i, \cdot]\|_2^2 \quad (M_r = U_r \Sigma_r V_r^*)$$

To see how $\|Y_{\lambda r}[i, \cdot]\|_2^2$ concentrates around its expectation, we write that

$$\left| \|Y_{\lambda p, r}[i, \cdot]\|_2^2 - E \left[ \|Y_{\lambda p, r}[i, \cdot]\|_2^2 \right] \right| = \left| \|Y_{\lambda p, r}[i, \cdot]\|_2^2 - \frac{\|M_r[i, \cdot]\|_2^2}{\lambda p} \right| \quad (\text{see (68)})$$

$$= \left| \sum_j \frac{\epsilon_{i,j} - \lambda p}{\lambda^2 p^2} |M_r[i, j]|^2 \right|$$

$$= : \left| \sum_j A_j \right|, \quad (71)$$

where $\{A_j\}$ are independent zero-mean random variables. In order to appeal to Lemma 17 (with $n_1 = n_2 = 1$), we calculate that

$$b = \max_j |A_j|$$

$$\leq \max_j \frac{|M_r[i, j]|^2}{\lambda^2 p^2} \quad (\text{see (71)})$$

$$\leq \frac{\|U_r[i, \cdot]\|_2^2 \cdot \sigma_1(M)^2 \cdot \max_j \|V_j[\cdot, :]\|_2^2}{\lambda^2 p^2} \quad (M_r = U_r \Sigma_r V_r^*)$$

$$= \frac{\mu_i(M_r)^r}{\lambda pm} \cdot \sigma_1(M) \cdot \max_j \frac{\nu_j(M_r)^r}{\lambda pm} \quad (\text{see (7)})$$

$$\leq \frac{\mu_i(M_r)^r}{\lambda pm} \cdot \frac{\sigma_1(M)^2 \cdot \eta(M_r)^r}{\lambda pm}, \quad (\text{see (9)})$$

23
\[ \sigma^2 = \mathbb{E} \left[ \sum_j A_j^2 \right] \]
\[ = \sum_j \frac{\mathbb{E} (\epsilon_{i,j} - \lambda p)^2}{\lambda^4 p^4} |M_r[i,j]|^4 \quad \text{(see (71))} \]
\[ = \sum_j \frac{\lambda p(1 - \lambda p)}{\lambda^4 p^4} |M_r[i,j]|^4 \quad (\epsilon_{i,j} \sim \text{Bernoulli}(\lambda p)) \]
\[ \leq \sum_j \frac{1}{\lambda^3 p^3} |M_r[i,j]|^4 \]
\[ \leq \frac{\max_j |M_r[i,j]|^2}{\lambda^3 p^3} : \|M_r[i,\cdot]\|_2 \]
\[ \leq \frac{\|U_r[i,\cdot]\|_2^2 \cdot \sigma_1(M)^2 \cdot \max_j \|V_r[j,\cdot]\|_2^2}{\lambda^3 p^3} : \|U_r[i,\cdot]\|_2^2 \cdot \sigma_1(M)^2 \quad (M_r = U_r \Sigma_r V_r^*) \]
\[ = \sigma_1^4(M) \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda^2 p^2 n^2} \cdot \frac{\max_j \|\sigma_j(M_r)^2 r^2\|}{\lambda p} \quad \text{(see (7))} \]
\[ \leq \sigma_1^4(M) \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda^2 p^2 n^2} \cdot \frac{\eta(M_r)r}{\lambda m} \quad \text{(see (9))} \]

and, finally,
\[ \max \{b, \sigma\} \leq \sigma_1^2(M) \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \cdot \frac{\max \left[ \frac{\eta(M_r)r}{\lambda m}, \sqrt{\frac{\eta(M_r)r}{\lambda m}} \right]}{\lambda p} \]
\[ \leq \sigma_1^2(M) \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda m}}, \]

where the second line holds assuming that \( \lambda p \geq \eta(M_r) \cdot \frac{2}{n} \). Consequently, the Bernstein inequality (Lemma 17) dictates that
\[ \left\| Y_{\lambda p, r}[i,\cdot] \right\|_2^2 = \left\| \frac{M_r[i,\cdot]}{\lambda p} \right\|_2^2 = \left\| \sum_j A_j \right\| \]
\[ \leq \alpha \max \{b, \sigma\} \]
\[ \leq \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda m}}, \]
\[ \leq \frac{\sigma_1^2(M)^2}{2} \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \cdot \frac{\eta(M_r)r}{\lambda m}, \quad \text{(72)} \]

except with a probability of \( e^{-\alpha} \), and provided that
\[ \lambda p \geq 4\alpha^2 \kappa_r(M)^4 \cdot \frac{\eta(M_r)r}{n}, \]

with \( \kappa_r(M) \) being the condition number of \( M_r \). From (72), it immediately follows that
\[ \left\| Y_{\lambda p, r}[i,\cdot] \right\|_2^2 \leq \frac{\left\| M_r[i,\cdot]\right\|_2^2}{\lambda p} + \frac{\sigma_r(M)^2}{2} \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \]
\[ \leq \frac{\left\| U_r[i,\cdot]\right\|_2^2 \cdot \sigma_1(M)^2}{\lambda p} + \frac{\sigma_r(M)^2}{2} \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \quad (M_r = U_r \Sigma_r V_r^*) \]
\[ = \frac{3\sigma_1^2(M)^2}{2} \cdot \frac{\mu_2(M_r)^2 r^2}{\lambda p} \quad \text{(see (7))} \]
\[
\|Y_{\lambda_p,r}[i,:]\|_2^2 \geq \frac{\|M_r[i,:]\|_2^2}{\lambda_p} - \frac{\sigma_r(M)^2}{2} \cdot \mu_i(M_r) r \\
\geq \frac{\|U_r[i,:]\|_2^2}{\lambda_p} \cdot \sigma_r(M)^2 - \frac{\sigma_r(M)^2}{2} \cdot \mu_i(M_r) r = \frac{\sigma_r^2(M)}{2} \cdot \mu_i(M_r) r - \frac{\mu_i(M_r) r}{\lambda p} \quad (M_r = U_r \Sigma_r Y_r^*)
\]

(74)

Overall, we conclude from (73) and (74) that

\[
\mu_i(M_r) \leq \frac{2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \quad \|Y_{\lambda_p,r}[i,:]\|_2^2 \leq 3\sigma_r(M)^2 \cdot \mu_i(M_r),
\]

(75)

provided that \( \lambda p \geq 4\alpha^2 \kappa_r(M)^4 \cdot \eta(M_r) \cdot \frac{\epsilon_r}{r} \). On the other hand, recall that \( Y_{\lambda_p} = R_{\lambda_p}(M + D) \) where the energy of noise is bounded as \( \|R_{\lambda p}(D)\|_F \leq \delta \). This allows us to write that

\[
\|Y_{\lambda_p}[i,:] - Y_{\lambda_p,r}[i,:]\|_2 \leq \|Y_{\lambda_p} - Y_{\lambda_p,r}\| \\
= \|R_{\lambda_p}(M + D) - R_{\lambda_p}(M_r)\| \\
\leq \|R_{\lambda_p}(M - M_r)\| + \|R_{\lambda_p}(D)\| \quad \text{(linearity and triangle inequality)} \\
\leq \|R_{\lambda_p}(M_{r+})\| + \delta \quad (M = M_r + M_{r+} \text{ and } \|R_{\lambda_p}(D)\|_F \leq \delta) \\
\leq \frac{M_{r+}}{\lambda p} + \delta = \frac{\sigma_{r+1}(M)}{\lambda p} + \delta, \quad \text{(see (55))}
\]

(76)

for a fixed \( i \in [1 : n] \). Therefore, we find that

\[
\mu_i(M_r) \leq \frac{2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \|Y_{\lambda_p,r}[i,:]\|_2^2 \quad \text{(see (75))}
\]

\[
\leq \frac{2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \left( \|Y_{\lambda_p}[i,:]\|_2 + \|Y_{\lambda_p,r}[i,:]\|_2 \right)^2 \quad \text{(triangle inequality)}
\]

\[
\leq \frac{2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \left( \|Y_{\lambda_p}[i,:]\|_2 + \frac{\mu_i(M_r) r}{\lambda p} \right)^2 \quad \text{(see (76))}
\]

\[
\leq \frac{6}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \|Y_{\lambda_p}[i,:]\|_2^2 + \frac{\mu_i(M_r) r}{\lambda p} \cdot \frac{6\sigma_r^2(M)}{\lambda p^2 r^2} + \frac{\lambda p m}{r} \cdot \frac{6\delta^2}{\sigma_r(M)^2} \quad ((a + b + c)^2 \leq 3(a^2 + b^2 + c^2))
\]

\[
\leq \frac{6}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \|Y_{\lambda_p}[i,:]\|_2^2 + \frac{\mu_i(M_r) r}{\lambda p} + \frac{6\gamma_r^2}{\lambda p^2} + \frac{6\delta^2}{\sigma_r(M)^2} \quad (\gamma_r(M) \leq \gamma_r)
\]

\[
= \tilde{\mu}_{r,i}.
\]

To find a matching upper bound for \( \tilde{\mu}_{r,i} \), we record that

\[
\tilde{\mu}_{r,i} = \frac{6}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \|Y_{\lambda_p,r}[i,:]\|_2^2 + 6\gamma_r^2 \cdot \frac{n}{\lambda p} + \frac{6\delta^2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \quad \text{(triangle inequality)}
\]

\[
\leq \frac{2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \left( \|Y_{\lambda_p,r}[i,:]\|_2 + \|Y_{\lambda_p}[i,:] - Y_{\lambda_p,r}[i,:]\|_2 \right)^2 + 6\gamma_r^2 \cdot \frac{n}{\lambda p} + \frac{6\delta^2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \quad \text{(see (76))}
\]

\[
\leq \frac{1}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \|Y_{\lambda_p,r}[i,:]\|_2^2 + \gamma_r^2 \cdot \frac{n}{\lambda p} + \frac{\delta^2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \quad ((a + b + c)^2 \leq 3(a^2 + b^2 + c^2) \text{ and } \gamma_r(M) \leq \gamma_r)
\]

\[
\leq \frac{1}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \left( 3\sigma_1(M)^2 \cdot \frac{\mu_i(M_r) r}{\lambda p} \right) + \gamma_r^2 \cdot \frac{n}{\lambda p} + \frac{\delta^2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r} \quad (\text{see (73))}
\]

\[
\leq \kappa_r(M)^2 \mu_i(M_r) + \gamma_r^2 \cdot \frac{n}{\lambda p} + \frac{\delta^2}{\sigma_r(M)^2} \cdot \frac{\lambda p m}{r}. \tag{77}
\]

This completes the proof of Lemma 12.
F Proof of Lemma 13

We focus on the row leverage scores. For \( i \in [1 : n] \), if

\[
\mu_i(M_r) \geq \beta^2 \eta(M_r),
\]

then \( i \in B \). By Lemma 12, (58), and an application of the union bound, we find that the measurement probability \( \lambda p \) in (59) is large enough to guarantee that

\[
\bar{\mu}_{r,i} \geq \mu_i(M_r) \geq \beta^2 \eta(M_r),
\]

except with a probability of at most \( n^{-\alpha \beta} \). Consequently, in this case, (60) holds because

\[
\beta^2 \eta(M_r) + (\bar{\mu}_{r,i} - \beta^2 \eta(M_r))_+ = \bar{\mu}_{r,i} \geq \mu_i(M_r).
\]

Otherwise, for \( i \in [1 : n] \), if

\[
\mu_i(M_r), \bar{\mu}_{r,i} \leq \beta^2 \eta(M_r),
\]

then (60) still holds because

\[
\beta^2 \eta(M_r) + (\bar{\mu}_{r,i} - \beta^2 \eta(M_r))_+ = \bar{\mu}_{r,i} \geq \beta^2 \eta(M_r) \geq \mu_i(M_r).
\]

Lastly, for \( i \in [1 : n] \), if

\[
\mu_i(M_r) \leq \beta^2 \eta(M_r) \leq \bar{\mu}_{r,i},
\]

then (60) holds again because

\[
\beta^2 \eta(M_r) + (\bar{\mu}_{r,i} - \beta^2 \eta(M_r))_+ = \bar{\mu}_{r,i} \geq \beta^2 \eta(M_r) \geq \mu_i(M_r).
\]

In summary, (60) holds for all \( i \), except with a probability of at most \( n^{-\alpha \beta} \). This completes the proof of Lemma 13, since a similar argument holds for the column leverage scores.

G Proof of Lemma 15

Let us focus on the row leverage scores. Recalling the definition of \( B \subseteq [1 : n] \) from (57), we write that

\[
\max_{i \in [1 : n]} \bar{\mu}_{r,i} = \max_{i \in B} \bar{\mu}_{r,i} \vee \max_{i \in [1 : n] \setminus B} \bar{\mu}_{r,i},
\]

and bound the maximum on \( B \) and its complement \([1 : n] \setminus B\) separately. For the maximum on \( B \), we argue as follows. For fixed \( i \in [1 : n] \), the matching upper bound in Lemma 12 states that

\[
\bar{\mu}_{r,i} \leq \kappa_r(M)^2 \mu_i(M_r) + \frac{\gamma^2 n}{\lambda pr} + \frac{\delta^2 \lambda p n}{\sigma_r(M)^2 r},
\]

except with a probability of at most \( e^{-\alpha} \) and provided that

\[
\lambda p \geq \alpha^2 \kappa_r(M)^4 \cdot \frac{\eta(M_r) r}{n}.
\]

Recall also the cardinality of \( B \) from (58). Then, taking \( \alpha = \alpha' \log n \) and applying the union bound, it follows that

\[
\max_{i \in B} \bar{\mu}_{r,i} \leq \kappa_r(M)^2 \max_{i \in B} \mu_i(M_r) + \frac{\gamma^2 n}{\lambda pr} + \frac{\delta^2 \lambda p n}{\sigma_r(M)^2 r}, \quad \text{if \( \lambda p \) satisfies (59)}
\]

\[
= \kappa_r(M)^2 \eta(M_r) + \frac{\gamma^2 n}{\lambda pr} + \frac{\delta^2 \lambda p n}{\sigma_r(M)^2 r}, \quad \text{(see (9))}
\]

(80)
except with a probability of at most $n^{-\alpha'\beta}$ and provided that

$$\lambda p \gtrsim \alpha' \beta^2 \kappa_r(M)^4 \cdot \frac{\eta(M_r) r \log^2 n}{n}, \quad (81)$$

Let us set $\alpha' = \alpha$ above to be consistent with earlier notation. Then, invoking Lemma 8, it follows that

$$\max_{i \in B} \mu_{r,i} = \max_{i \in B} \left( \frac{C_1 \|Y_{\lambda p}[i, :]\|_2^2}{\sigma_r(Y_{\lambda p})^2} + \frac{C_2 \gamma_r^2}{\lambda p^2} + \frac{C_3 \delta^2}{\sigma_r(Y_{\lambda p})^2} \right) \frac{\lambda p n}{r} \quad \text{(see (64))}$$

$$\lesssim \max_{i \in B} \left( \frac{\|Y_{\lambda p}[i, :]\|_2^2}{\sigma_r(M)^2} + \frac{\gamma_r^2}{\lambda p^2} + \frac{\delta^2}{\sigma_r(M)^2} \right) \frac{\lambda p n}{r} \quad \text{(see Lemma 8)}$$

$$\lesssim \max_{i \in B} \mu_{r,i} \quad \text{(see (56))}$$

$$\lesssim \kappa_r(M)^2 \eta(M_r) + \frac{\gamma_r^2}{\lambda p^2} + \frac{\delta^2}{\sigma_r(M)^2} \rho_r. \quad \text{(see (80))} \quad (82)$$

except with a probability of at most $n^{-\alpha'\beta} + e^{-\alpha}$ and if $\lambda p$ satisfies (61).

For the maximum on the complement of $B$ in (78), we make the following argument. Recall that we set $Y_{\lambda p,r} = R_{\lambda p}(M_r)$ in Appendix E, and then fix $i \in [1 : n]$. Our starting point is (72), which implies that

$$\|Y_{\lambda p,r}[i, :]\|_2^2 \leq \frac{\|Y_{\lambda p}[i, :]\|_2^2}{\lambda p} + \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}}, \quad \text{(third line of (72))}$$

$$\leq \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \frac{\eta(M_r)r}{\lambda p} + \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}}, \quad (M_r = U_r \Sigma_r V_r^* \text{ and (7), (9)})$$

$$\leq (1 + \alpha) \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}}, \quad \text{(if $\lambda p \geq \frac{\eta(M_r)r}{n}$)}$$

$$\lesssim \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}}, \quad (\alpha \geq 1) \quad (83)$$

except with a probability of at most $e^{-\alpha}$. Setting $Y_{\lambda p,r} = R_{\lambda p}(M_{r+})$ with a slight abuse of notation, the above bound implies that

$$\|Y_{\lambda p}[i, :]\|_2^2 = \|Y_{\lambda p}[i, :] + Y_{\lambda p,r}[i, :]\|_2^2 \quad \text{(for $Y_{\lambda p} = R_{\lambda p}(M_r + M_{r+} + D) = Y_{\lambda p,r} + Y_{\lambda p,r+} + R_{\lambda p}(D)$)}$$

$$\leq 3 \left( \|Y_{\lambda p}[i, :]\|_2^2 + \|Y_{\lambda p,r}[i, :]\|_2^2 + \delta^2 \right) \quad \text{($a + b + c^2 \leq 3(a^2 + b^2 + c^2)$ and $\|R_{\lambda p}(D)\|_F \leq \delta$)}$$

$$\lesssim \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}} + \|Y_{\lambda p,r}[i, :]\|_2^2 + \delta^2 \quad \text{(see (83))}$$

$$\leq \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}} + \frac{\|M_{r+}[i, :]\|_2^2}{\lambda p^2} + \delta^2 \quad (Y_{\lambda p,r+} = R_{\lambda p}(M_{r+}) \text{ and (10)})$$

$$\leq \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}} + \frac{\|M_{r+}\|_2^2}{\lambda p^2} + \delta^2$$

$$= \alpha \cdot \sigma_1(M)^2 \cdot \frac{\mu_i(M_r)r}{\lambda p n} \cdot \sqrt{\frac{\eta(M_r)r}{\lambda p}} + \frac{\sigma_{r+}(M)^2}{\lambda p^2} + \delta^2. \quad (84)$$
Appealing to Lemma 8, we therefore find that

\[
\bar{p}_{r,i} = \left( \frac{C_1 \| Y_{\lambda p}[i, :]\|^2}{\delta r (Y_{\lambda p})^2} + \frac{C_2 \gamma_r^2}{\lambda^2 p^2} + \frac{C_3 \delta^2}{\sigma_r (Y_{\lambda p})^2} \right) \frac{\lambda_{pm}}{r} \quad \text{(see (64))}
\]

\[
\lesssim \left( \frac{\| Y_{\lambda p}[i, :]\|^2}{\delta r (M)^2} + \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r} \quad \text{(see Lemma 8)}
\]

\[
\lesssim \alpha \cdot \kappa_r (M)^2 \mu_i (M_r) \sqrt{\frac{\eta (M_r) r}{\lambda_{pm}}} + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r}, \quad (84) \text{ and } \gamma_r (M) \leq \gamma_r
\]

except with a probability of at most \( e^{-\alpha} \). Taking \( \alpha = \alpha' \log n \) and applying the union bound to the set \([1 : n] \setminus B\), we arrive at

\[
\max_{i \in [1 : n] \setminus B} \bar{p}_{r,i} \lesssim \alpha' \log n \cdot \kappa_r (M)^2 \mu_i (M_r) \sqrt{\frac{\eta (M_r) r}{\lambda_{pm}}} + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r},
\]

except with a probability of \( n^{-\alpha'} \). We set \( \alpha' = \alpha \) above to be consistent in our notation and then conclude that

\[
\max_{i \in [1 : n] \setminus B} \bar{p}_{r,i} \lesssim \alpha \log n \cdot \kappa_r (M)^2 \mu_i (M_r) \sqrt{\frac{\eta (M_r) r}{\lambda_{pm}}} + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r}
\]

\[
\lesssim \alpha \log n \cdot \kappa_r (M)^2 \cdot \beta^2 \eta (M_r) \sqrt{\frac{\eta (M_r) r}{\lambda_{pm}}} + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r} \quad \text{(\( \mu_i (M_r) \leq \beta^2 \eta (M_r), i \in [1 : n] \setminus B \))}
\]

\[
\lesssim \alpha \beta \cdot \eta (M_r) + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r} \quad \text{(see (61))}
\]

\[
\leq \alpha \cdot \eta (M_r) + \left( \frac{\gamma_r^2}{\lambda^2 p^2} + \frac{\delta^2}{\sigma_r (M)^2} \right) \frac{\lambda_{pm}}{r}, \quad (\beta \leq 1) \quad \text{(85)}
\]

except with a probability of at most \( n^{-\alpha} \). Combining (78) with (82) and (85), we find that

\[
\max_{i \in [1 : n]} \bar{p}_{r,i} = \max_{i \in B} \bar{p}_{r,i} \lor \max_{i \in [1 : n] \setminus B} \bar{p}_{r,i} \quad \text{(see (78))}
\]

\[
\lesssim \alpha \cdot \kappa_r (M)^2 \eta (M_r) + \frac{\gamma_r^2 n}{\lambda_{pr}} + \frac{\delta^2 \lambda_{pm}}{\sigma_r (M)^2 r} \quad \text{(see (82) and (85))}
\]

except probability of at most \( n^{-\alpha^2} + e^{-\alpha} \). This completes the proof of Lemma 15 since a similar bound holds for the column leverage scores.

H Proof of Lemma 16

Let us focus on the row leverage scores. Note that

\[
\| R_{\lambda p} (M) \|^2_F^2 = \sum_{i,j=1}^n \frac{\varepsilon_{i,j}}{\lambda_{pr}^2} \cdot M[i, j]^2, \quad (\varepsilon_{i,j} \in \{0, 1\} \text{ in (10)})
\]

\[
\mathbb{E} \left[ \| R_{\lambda p} (M) \|^2_F^2 \right] = \mathbb{E} \left[ \sum_{i,j=1}^n \frac{\varepsilon_{i,j}}{\lambda_{pr}^2} \cdot M[i, j]^2 \right] = \frac{\| M \|^2_F}{\lambda_p}, \quad (\varepsilon_{i,j} \sim \text{Bernoulli}(\lambda p))
\]
\[ \| R_{\lambda_2}(M) \|_F^2 - \frac{\| M \|_F^2}{\lambda_2} = \| R_{\lambda_2}(M) \|_F^2 - \mathbb{E} \left[ \| R_{\lambda_2}(M) \|_F^2 \right] \]
\[ = \sum_{i,j=1}^{n} \left( \frac{\epsilon_{i,j} - \lambda_2}{\lambda_2^2 p^2} \right) M[i,j]^2 \]
\[ =: \sum_{i,j} A_{i,j}, \quad (86) \]

where \( \{A_{i,j}\} \) are zero-mean independent random variables. In order to apply Lemma 17 (with \( n_1 = n_2 = 1 \)), note that

\[ | A_{i,j} | \leq \frac{M[i,j]^2}{\lambda_2^2 p^2} \quad \text{(see (86))} \]
\[ \leq \left( \frac{\| U[i,:]\|_2 \cdot \sigma_1(M) \cdot \| V[j,:]\|_2}{\lambda_2} \right)^2 \quad (M_r = U_r \Sigma_r V_r^*) \]
\[ \leq \left( \frac{\sigma_1(M) \cdot \eta(M_r)r}{\lambda_2} \right)^2 =: b, \quad \text{(see (7) and (9))} \]

\[ \sum_{i,j=1}^{n} \mathbb{E} \left[ A_{i,j}^2 \right] = \sum_{i,j=1}^{n} \mathbb{E} \left[ \left( \frac{\epsilon_{i,j} - \lambda_2}{\lambda_2^2 p^2} \right)^2 \right] M[i,j]^4 \quad \text{(see (86))} \]
\[ = \sum_{i,j=1}^{n} \frac{\lambda_2 (1 - \lambda_2)}{\lambda_2^4 p^4} M[i,j]^4 \quad (\epsilon_{i,j} \sim \text{Bernoulli}(\lambda_2)) \]
\[ \leq \sum_{i,j} M[i,j]^4 \frac{\lambda_2}{\lambda_2^4 p^4} \]
\[ \leq \max_{i,j} M[i,j]^2 \cdot \sum_{i,j} M[i,j]^2 \]
\[ \leq \frac{\sigma_1(M)^2}{\lambda_2^4 p^4} \cdot \left( \frac{\eta(M_r)r}{n} \right)^2 \cdot \| M \|_F^2 \]
\[ = \left( \frac{\sigma_1(M) \cdot \eta(M_r)r}{\lambda_2 p n} \right)^2 \cdot \| M \|_F^2 =: \sigma_2. \]

With the choice of

\[ \lambda_2 \geq \frac{\kappa_r(M)^2}{r} \cdot \left( \frac{\eta(M_r)r}{n} \right)^2, \quad (87) \]

it follows that

\[ \max [b, \sigma] \lesssim \frac{\| M \|_F^2}{\lambda_2}. \quad (88) \]

Lemma 17 therefore dictates that

\[ \left| \| R_{\lambda_2}(M) \|_F^2 - \frac{\| M \|_F^2}{\lambda_2} \right| = \sum_{i,j} A_{i,j} \]
\[ \lesssim \alpha \max [b, \sigma] \quad \text{(see Lemma 17)} \]
\[ \lesssim \alpha \| M \|_F^2, \quad \text{(see (88))} \]

except with a probability of at most \( e^{-\alpha} \) and, consequently,

\[ \| R_{\lambda_2}(M) \|_F^2 \lesssim \frac{\alpha \| M \|_F^2}{\lambda_2}, \quad (\alpha \geq 1) \quad (89) \]
and, in turn,
\[
\|Y_{\lambda^p}\|_F^2 = \|R_{\lambda^p}(M + D)\|_F^2 \\
\leq (\|R_{\lambda^p}(M)\|_F + \|R_{\lambda^p}(D)\|_F)^2 \quad \text{(triangle inequality)} \\
\leq 2 \|R_{\lambda^p}(M)\|_F^2 + 2 \|R_{\lambda^p}(D)\|_F^2 \quad \text{\((a + b)^2 \leq 2a^2 + 2b^2\)} \\
\lesssim \alpha \|M\|_{\lambda^p}^2 + \delta^2. \quad \text{(see (89), \(\|R_{\lambda^p}(D)\|_F \leq \delta\))} \quad (90)
\]

It now follows that
\[
\sum_{i=1}^{n} \mu_{r,i} = \frac{C_1}{\sigma_r(Y_{\lambda^p})^2} \cdot \frac{\lambda_{pr}}{r} \sum_{i=1}^{n} ||Y_{\lambda^p}[i,.]||_F^2 + \frac{C_2 \gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(Y_{\lambda^p})^2} \quad \text{(see (64))} \\
= \frac{C_1}{\sigma_r(Y_{\lambda^p})^2} \cdot \frac{\lambda_{pr}}{r} \|Y_{\lambda^p}\|_F^2 + \frac{C_2 \gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(Y_{\lambda^p})^2} \\
\lesssim \frac{1}{\sigma_r(Y_{\lambda^p})^2} \cdot \frac{\lambda_{pr}}{r} \left( \frac{\alpha \|M\|_{\lambda^p}^2}{\lambda_{pr}} + \delta^2 \right) + \frac{C_2 \gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(Y_{\lambda^p})^2} \quad \text{(see (90))} \\
= \frac{\alpha}{\sigma_r(Y_{\lambda^p})^2} \cdot \frac{n \lambda_{pr}}{r} \|M\|_F^2 + \frac{1}{\sigma_r(Y_{\lambda^p})^2} \frac{\lambda_{pr}}{r} \cdot \delta^2 + \frac{\gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(Y_{\lambda^p})^2} \quad \text{(see Lemma 8)} \\
\lesssim \frac{\alpha n \|M\|_F^2}{\sigma_r(M)^2 r} + \frac{\gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(M)^2} \frac{\lambda_{pr}}{r} \quad \langle \gamma(r) \rangle \leq \gamma_r \rangle \\
\leq \frac{\alpha n \cdot \kappa_r(M)^2}{r} + \frac{n^2 \gamma_r(M)^2}{\lambda_{pr}} + \frac{\gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2}{\sigma_r(M)^2 \frac{\lambda_{pr}}{r}} \quad \text{(\(\|M\|_F^2 \leq r \cdot \sigma_1(M)^2 + n \cdot \sigma_{r+1}(M)^2\))} \\
\leq \frac{\alpha n \cdot \kappa_r(M)^2}{r} + \frac{n^2 \gamma_r(M)^2}{\lambda_{pr}} + \frac{\gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2 \lambda_{pr}^2}{\sigma_r(M)^2 r} \quad \langle \gamma(r) \rangle \leq \gamma_r \rangle \\
\lesssim \frac{\alpha n \cdot \kappa_r(M)^2}{r} + \frac{\gamma^2 \sigma^2}{\lambda_{pr}} + \frac{C_1 \delta^2 \lambda_{pr}^2}{\sigma_r(M)^2 \frac{\lambda_{pr}}{r}},
\]
except with a probability of at most \(e^{-\alpha}\) and provided that (87) holds. This completes the proof of Lemma 16, as an identical bound holds for column leverage scores.