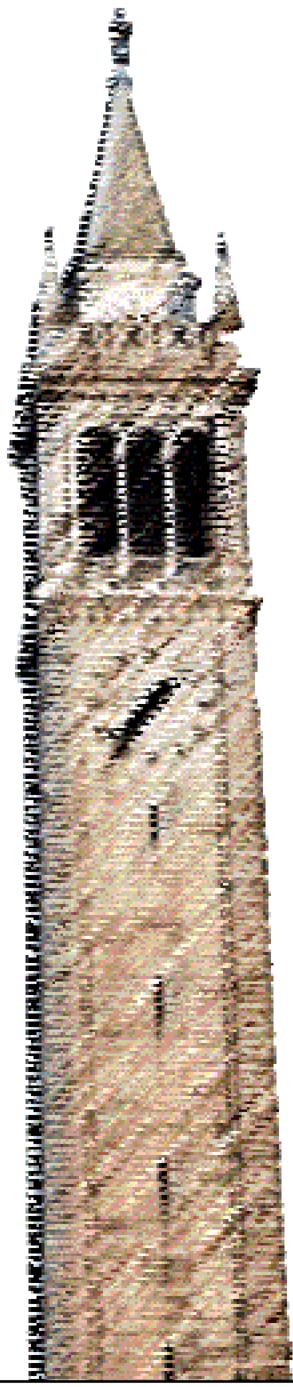
Improved Bounds for Incoherent Matrix Completion



Emaan Hariri

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-33
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-33.html

May 1, 2022

Improved Bounds for Incoherent Matrix Completion

by Emaan Hariri

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Satish Rao
Research Advisor

12/17/2021

(Date)

* * * * * * *

Professor Jelani Nelson
Second Reader

12/17/2021

(Date)

ACKNOWLEDGEMENTS

I am profoundly indebted to my advisor Prof. Satish Rao, whose constant guidance and advice was integral not only to all aspects of this project, but to the latter half of my academic career at Berkeley. Thank you to his former student, Jaya Narasimhan, whose thesis served as the inspiration for this one. Thanks are due to Prof. Tselil Schramm for our insightful conversations relating to ε-nets and to Max Ovsiankin for our illuminating discussion on bounding the expected maximum of simple symmetric random walks. Thank you to Arun Ganesh for his help editing and reviewing as well as to my second reader, Prof. Jelani Nelson, for his comments and for exposing me to the power of sketching algorithms. Finally, I am grateful for my family and friends, for without their endless support and encouragement none of this work would be possible.

To my parents.

IMPROVED BOUNDS FOR INCOHERENT MATRIX COMPLETION

Emaan Hariri
Computer Science Division
University of California, Berkeley
`ehariri@berkeley.edu`

ABSTRACT

In this report, we progress towards removing an extraneous condition imposed in previous sample complexity results for matrix completion via nuclear norm minimization. We build upon the sample complexity results of Candès and Recht (1), Candès and Tao (2), Gross (3), and Recht (4) which employ a dual certificate construction that, with high probability, guarantees the optimality of nuclear norm minimization. The aforementioned authors all make two assumptions in their analysis about the rank-r target matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ with singular value decomposition $\mathbf{U\Sigma V}^*$ which we wish to recover. The first assumption, that $\mathbf{M}$ has *coherence* bounded by μ_0, was shown to be necessary by Candès and Tao (2) and is preserved here. We aim, then, to remove the second assumption, that entries of $\mathbf{UV}^*$ are bounded in magnitude by $\mu_1\sqrt{r}/n$, by improving on the analysis of Recht (4). In particular, Recht (4) constructs the dual via an iterative process wherein the approximation error is tracked using a residual matrix. Recht (4) ensures the maximum entry of the residuals drops geometrically. Here, we track individual matrix entries and their respective columns and rows much more closely, rather than relying on a global bound. As a result, we demonstrate exact recovery of matrices is possible with $m \geq O(\mu_0 n r^{3/2} \beta \log^2 n)$ entries, with probability $1 - n^{-\Omega(\text{poly}(\beta))}$ for $\beta > 0$ of our choosing. This represents an improvement by a factor of $O(\mu_0\sqrt{r})$ from $O(\mu_0^2 n r^2 \beta \log^2 n)$, the best possible result from Recht (4) when not relying on the bounded entry assumption.

1 INTRODUCTION

In the problem of matrix completion, we are given some incomplete sample Ω of $m = |\Omega|$ entries taken uniformly at random from some rank-r matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ and wish to reconstruct $\mathbf{M}$ exactly (and thus "complete" the matrix). This problem is found in numerous contexts including global positioning (5), system detection (6), and collaborative filtering and recommender systems (7). The *Netflix problem*, where we wish to complete some movie-ratings matrix where the $(i, j)^{\text{th}}$ entry is user i's rating of movie j, is perhaps the most famous application of matrix completion for collaborative filtering.

The matrix $\mathbf{M}$ has $O(nr)$ degrees of freedom (or intrinsic dimension), giving us some hope of completing the matrix without the need of all n^2 entries, as would be needed for general matrices. The problem of low-rank matrix completion finds application in numerous contexts and has parallels to the problem of compressed sensing. Under our low-rank assumption on matrices, the problem of matrix completion can be addressed by applying rank minimization via the program

$$\begin{array}{ll} \text{minimize} & \text{rank}(\mathbf{X}) \\ \text{subject to} & X_{ij} = M_{ij} \quad \forall (i, j) \in \Omega. \end{array}$$

As with the problem of determining the sparsest solution to a system of linear equations, the problem of low-rank matrix completion is NP-Hard, but has a variety of heuristics and relaxations which make it reasonable to approximate in polynomial-time. We let $\sigma_i(\mathbf{X})$ be the i^{th} largest singular value of $\mathbf{X}$ and $\sigma(\mathbf{X}) \in \mathbb{R}^n$ be a vector of these values. Consider the *nuclear norm* of $\mathbf{X}$, $\|\mathbf{X}\|_* = \sum_{i=1}^n \sigma_i(\mathbf{X}) = \|\sigma(\mathbf{M})\|_1$, and note $\text{rank}(X) = \|\sigma(\mathbf{X})\|_0$. Fazel (8) showed that nuclear norm is the convex envelope of rank, motivating the use of nuclear norm optimization as a the relaxation of rank optimization to a convex setting as suggested by Fazel, et. al (9) and built upon by Recht, Fazel, and Parillo (10). This relaxation can now be addressed in polynomial-time using semidefinite programming (SDP). In particular, Candès and Recht (1) showed when $m \geq O(\mu_0 n^{1.25} r \log n)$, where μ_0 is the *coherence* (defined below) of $\mathbf{M}$, that, with high probability, the solution to

$$\begin{array}{ll} \text{minimize} & \|\mathbf{X}\|_* \\ \text{subject to} & X_{ij} = M_{ij} \quad \forall (i, j) \in \Omega \end{array} \tag{1.1}$$

is **M**! This is surprising because there is *zero* error in the recovery of **M**. This demands a formal definition of coherence. We must consider matrix such as

$$\mathbf{M} = \vec{e}_1\vec{e}_3^* = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which clearly requires sampling all entries despite being rank-1; this motivates the following definition.

Definition 1.1. (Coherence) The *coherence* of the dimension-r subspace $U \subseteq \mathbb{R}^n$ with respect to the standard basis $\{\vec{e}_i\}_{i\in[n]}$ of $\mathbb{R}^n$ is

$$\mu(U) := \max_{1\le i\le n} \frac{n}{r}\|P_U\vec{e}_i\|_2^2$$

where $P_U : \mathbb{R}^n \to U$ is the projection onto the subspace U.

Remark 1.2. When referring to the "coherence of matrix **M**" or using the notation $\mu(\mathbf{M})$, this refers to the *maximum of the coherences of its row and columns spaces*. Furthermore, the coherences of the orthogonal matrices **U** and **V** refer to the coherences of the column spaces spanned by each matrix, respectively. Similarly, we let $P_\mathbf{U}$ be the projection onto the column space of **U** which we often use in place of P_U. As $P_\mathbf{U} = \mathbf{U}\mathbf{U}^*$, we may also express $\mu(\mathbf{U}) = \frac{n}{r}\max_{1\le i\le n}\|\vec{u}_i\|_2^2$ where $\{\vec{u}_i\}_{i=1}^n$ are the rows of **U**.

Notably, $1 \le \mu(\mathbf{M}) \le n/r$, where **M** with $\mu(\mathbf{M}) = n/r$ is a maximally *coherent* matrix; in the case of the aforementioned $\mathbf{M} = \vec{e}_1\vec{e}_3^*$, $\mu(\mathbf{M}) = n/r$ and **M** is thus maximally coherent. Naturally, it follows that our sample complexity bound for the recovery of **M** must depend on $\mu(\mathbf{M})$, motivating the first of the following assumptions made by Candès and Recht ($\mathbf{U\Sigma V}^*$ is the compact SVD of **M**).

Assumption 1.3. (**A0**) **U** and **V** have coherence bounded by μ_0, or $\max\{\mu(\mathbf{U}), \mu(\mathbf{V})\} \le \mu_0$.

Assumption 1.4. (**A1**) $\mathbf{UV}^*$ has maximum entry bounded by $\mu_1\sqrt{r}/n$, or $\|\mathbf{UV}^*\|_\infty \le \mu_1\sqrt{r}/n$.

In Corollary 2.6 we show that **A0** immediately implies **A1** with $\mu_1 = \mu_0\sqrt{r}$. Furthermore, by **A0** we may take w.l.o.g. $\mu(\mathbf{V}) \le \mu(\mathbf{U})$, which we do in this report for simplicity.

Candès and Recht rely on the above assumptions in their construction of the dual **Y** which verifies that the minimizer of 1.1 is indeed our desired **M**. Candès and Tao (2) improve Candès and Recht's bound to $m \ge O(\mu_2^2 nr\log^6 n)$, relying on a "strong incoherence condition" stronger than **A0** and **A1** where $\max(\mu_0, \mu_1) \le \mu_2$. Recht (4) further improves this bound to $m \ge O(\max\{\mu_1^2, \mu_0\}nr\log^2 n)$, greatly simplifying Candès and Tao's analysis by employing sampling with replacement in place of Bernoulli sampling, but still relying on **A0** and **A1**.

Candès and Tao (2) proved that $\Omega(\mu_0 nr\log n)$ entries are necessary for matrix completion via nuclear norm minimization when sampling uniformly at random. This aligns with our expectation that μ_0 is integral to the sample complexity and, hence, **A0** being *necessary*. As Recht (4) notes, however, it has not been shown that **A1** is necessary.

Our optimality results employ the same setup as Recht (4) but modifies their proof that the dual variable satisfies the needed conditions (discussed below) that certify **M** is the unique minimizer of 1.1.

2 Setup and Framework

We provide notation and outline the framework used in the derivation of our results. We begin with a formal definition of the problem of matrix completion.

Problem 2.1 (Matrix Completion)**.** We are given a rank-r matrix $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$ and some sample $\Omega \subseteq [n_1]\times[n_2]$ of m entries from **M** taken uniformly at random. What is the minimum number of entries $m = |\Omega|$ needed to exactly recover **M** with high probability?

As in previous literature, we concern ourselves with the setting where entries are sampled uniformly at random and also make the simplifying assumption that $n = n_1 = n_2$. Furthermore, throughout this paper we make use of positive constant C (and C_1, C_2, etc.) to absorb other constants and simplify our analysis. Events are said to occur "with high probability" if they happen with probability at least $1 - n^{-\Omega(\text{poly}(\beta))}$ for $\beta > 0$ of our choosing.

Let the compact singular value decomposition of our matrix in question $\mathbf{M} \in \mathbb{R}^{n\times n}$ be $\mathbf{U\Sigma V}^*$; $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times r}$. For general matrix **M**, denote its i^{th} row by $\vec{m}_i$, j^{th} column by $\vec{m}^{(j)}$, and $(i,j)^{\text{th}}$ entry by M_{ij}. We let $\sigma_k(\mathbf{M})$ be the k^{th} largest singular value of **M** and let $\sigma(\mathbf{M}) \in \mathbb{R}^n$ be the vector containing these values. We denote the spectral norm of a matrix $\|\mathbf{M}\| = \sigma_1(\mathbf{M})$, the infinity norm $\|\mathbf{M}\|_\infty = \max_{1\le i,j,\le n}|M_{ij}|$, the nuclear norm $\|\mathbf{M}\|_* = \|\sigma(\mathbf{M})\|_1$ (observe $\text{rank}(\mathbf{M}) = \|\sigma(\mathbf{M})\|_0$), and the Frobenius norm $\|\mathbf{M}\|_F = \|\sigma(\mathbf{M})\|_2 = \sqrt{\sum_{i=1}^n\sum_{j=1}^n M_{ij}^2}$.

Furthermore, denote the maximum of row and column norms of $\mathbf{M}$ with $\|\mathbf{M}\|_r$ and $\|\mathbf{M}\|_c$, respectively, and let $\|\mathbf{M}\|_b := \max\{\|\mathbf{M}\|_r, \|\mathbf{M}\|_c\}$, or the maximum of *both* of these[1]. Matrix operators $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ have operator norm defined $\|\mathcal{A}\| := \sup_{\mathbf{M}:\|\mathbf{M}\|_F \le 1} \|\mathcal{A}(\mathbf{M})\|_F$.

We let the tangent space of $\mathbf{M}$ be $T = \{\mathbf{UX}^* + \mathbf{YV}^* : \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times r}\} \subseteq \mathbb{R}^{n\times n}$. This is useful as $\text{col}(\mathbf{M}) = \text{col}(\mathbf{U})$ and $\text{row}(\mathbf{M}) = \text{row}(\mathbf{V}^*)$, and we may decompose $\mathbb{R}^{n\times n} = T \otimes T^{\perp}$.

Recall the projection operator onto $\text{col}(\mathbf{U})$ is denoted $P_{\mathbf{U}}$, which, for orthonormal $\mathbf{U}$ is $P_{\mathbf{U}} = \mathbf{UU}^*$.The projection operator onto T, $\mathcal{P}_{\mathcal{T}} : \mathbb{R}^{n\times n} \to T$, is defined

$$\mathcal{P}_{\mathcal{T}}(\mathbf{X}) := P_{\mathbf{U}}\mathbf{X} + \mathbf{X}P_{\mathbf{V}} - P_{\mathbf{U}}\mathbf{X}P_{\mathbf{V}}$$

and so $\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{X}) = (\mathcal{I} - \mathcal{P}_{\mathcal{T}})(\mathbf{X}) = (\mathcal{I} - P_{\mathbf{U}})\mathbf{X}(\mathcal{I} - P_{\mathbf{V}})$. For convenience, we define $\mathcal{P}_{\mathbf{UV}^*}(\mathbf{Z}) = P_{\mathbf{U}}\mathbf{Z}P_{\mathbf{V}}$ and we will also use $P_{\mathbf{V}}(\mathbf{Z})$ and $\mathbf{Z}P_{\mathbf{V}}$ interchangeably.

To simplify, we define the sampling operator for $\Omega \subseteq [n] \times [n]$ to be $\mathcal{R}_\Omega$, where $\mathcal{R}_\Omega$ acts as a masking matrix $\mathbf{W}$ where $W_{ij} = 1$ iff $(i, j) \in \Omega$ and $W_{ij} = 0$ otherwise, so $\mathcal{R}_\Omega(\mathbf{X}) = \mathbf{W} \circ \mathbf{X}$. Thus, we may state 1.1 as

$$\begin{aligned} &\text{minimize} && \|\mathbf{X}\|_* \\ &\text{subject to} && \mathcal{R}_\Omega(\mathbf{X}) = \mathcal{R}_\Omega(\mathbf{M}). \end{aligned} \tag{2.1}$$

Recht (4), building off Candès and Recht (1), shows via their Theorem 2 that if there is a dual variable $\mathbf{Y} \in \text{range}(\mathcal{R}_\Omega)$ such that

$$\|\mathcal{P}_{\mathcal{T}}(\mathbf{Y}) - \mathbf{UV}^*\|_F \le \sqrt{\frac{r}{2n}} \tag{2.2}$$

$$\|\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{Y})\| \le \frac{1}{2} \tag{2.3}$$

then $\mathbf{M}$ is the unique minimizer to the nuclear norm minimization program (2.1) with sample complexity $m \ge O(\max\{\mu_1^2, \mu_0\} nr \log^2 n)$, with high probability.

For our result, we use the sampling with replacement model and dual certificate construction from Recht (4). Indeed, for our main result we employ the the majority of Recht (4) Theorem 2 while using modified theorems which avoid reliance on **A1**.

We recall the dual construction from Gross (3) employed by Recht (4) (itself building upon Candès and Recht (1)). Partition Ω into p subsamples of size q, or $\{\Omega_j\}_{j=1}^p$ where $|\Omega_j| = q$ (thus $m = pq$). Initialize $\mathbf{W}_0 = \mathbf{UV}^*$ and iteratively define

$$\mathbf{Y}_k = \frac{n^2}{q} \sum_{j=1}^{k} \mathcal{R}_{\Omega_j}(\mathbf{W}_{j-1}) \quad \text{and} \quad \mathbf{W}_k = \mathcal{P}_{\mathcal{T}}(\mathbf{UV}^* - \mathbf{Y}_k) = \mathbf{UV}^* - \mathcal{P}_{\mathcal{T}}(\mathbf{Y}_k)$$

for $k = 1, \ldots, p$, where in Recht (4) takes $q \ge O(\max\{\mu_0, \mu_1^2\} nr\beta \log n)$ and $p \ge O(\log n)$. The dual certificate in question is thus $\mathbf{Y} = \mathbf{Y}_p$.

In this report, we refine their method for bounding $\|\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{Y}_p)\|$. In particular, we modify the applied Theorem 7 of Recht (4) to use the *maximum of the column and row norms* rather than just the *infinity norm* of our dual variables $\mathbf{W}_k$. We then demonstrate that these norms drop geometrically, allowing us to use $\|\mathbf{UV}^*\|_b = \sqrt{\mu_0 r / n}$ and the trivial bound on $\|\mathbf{UV}^*\|_\infty = \mu_0 r / n$ that follows from **A0** (see Corollary 2.6).

2.1 Useful Theorems

We provide the following theorems which will be useful in our main argument, beginning with some standard inequalities.

Theorem 2.2 (Standard Bernstein Inequality). *Let $X_1 \ldots X_n$ be independent, zero-mean random variables such that $\forall i : |X_i| \le M$, then for $\tau > 0$*

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} X_i\right| \ge \tau\right] \le 2\exp\left(-\frac{\tau^2/2}{\sum_{i=1}^{n} \mathbb{E}[X_i^2] + M\tau/3}\right)$$

[1]Originally from Ding (11) and generalized by Nie (12), the $L_{p,q}$ "entry-wise" norm of matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is $\|\mathbf{A}\|_{p,q} := \left(\sum_{j=1}^{n} (\sum_{i=1}^{n} |A_{ij}|^p)^{(q/p)}\right)^{1/q}$. Observe that $\|\mathbf{A}\|_c = \|\mathbf{A}\|_{2,\infty}$ and similarly $\|\mathbf{A}\|_r = \|\mathbf{A}^*\|_{2,\infty}$, hence, $\|\cdot\|_b$ is clearly a norm.

Corollary 2.3. *Let $X_1, \dots, X_n$ be independent zero-mean random variables such that $\forall i : |X_i| \le M$, $S := \sum_{i=1}^n X_i$ (hence $\mathrm{var}(S) = \sum_{i=1}^n \mathbb{E}\left[X_i^2\right]$), we have*

$$|S| \le 2\max\left\{\sqrt{2\mathrm{var}(S)\beta\log n},\ {}^2\!/\!_3 M\beta\log n\right\}$$

with probability at least $1 - 2n^{-\beta}$ for $\beta > 0$.

Theorem 2.4 (Noncommutative Matrix Bernstein)**.** *Let $\{\mathbf{Z}_k\}$ be a sequence of independent zero-mean random matrices of dimension $d_1 \times d_2$ such that $\forall k : \|\mathbf{Z}_k\| \le M$. Let*

$$\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\left[\mathbf{Z}_k\mathbf{Z}_k^*\right]\right\|, \left\|\sum_k \mathbb{E}\left[\mathbf{Z}_k^*\mathbf{Z}_k\right]\right\|\right\}.$$

Then, for all $\tau \ge 0$,

$$\mathbb{P}\left[\left\|\sum_k \mathbf{Z}_k\right\| \ge \tau\right] \le (d_1 + d_2)\cdot\exp\left(-\frac{\tau^2/2}{\sigma^2 + M\tau/3}\right).$$

Corollary 2.5. *Let $\{\mathbf{Z}_k\}$ be independent zero-mean random matrices of dimension $n \times n$ where $\forall k : \|\mathbf{Z}_k\| \le M$ and $\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\left[\mathbf{Z}_k\mathbf{Z}_k^*\right]\right\|, \left\|\sum_k \mathbb{E}\left[\mathbf{Z}_k^*\mathbf{Z}_k\right]\right\|\right\}$ (as in Theorem 2.4). We have*

$$\left\|\sum_k \mathbf{Z}_k\right\| \le 2\max\left\{\sqrt{2\beta\sigma^2\log n},\ {}^2\!/\!_3\beta M\log n\right\}$$

with probability at least $1 - 2n^{1-\beta}$ for $\beta > 0$.

We briefly note the following property that follows when only using **A0**.

Corollary 2.6. ***A0*** *immediately implies* ***A1*** *with $\mu_1 = \mu_0\sqrt{r}$.*

Proof. **A0** tells us that $\|\mathbf{U}\mathbf{U}^*\|_b^2 = \max_{1\le i\le n}\|P_{\mathbf{U}}\vec{e}_i\|_2^2 \le \mu_0 r/n$. By orthogonality of $\mathbf{U}$, $\max_{1\le i\le n}\|\mathbf{U}\mathbf{U}^*\vec{e}_i\|_2 = \max_{1\le i\le n}\|\mathbf{U}^*\vec{e}_i\|_2 = \|\mathbf{U}\|_r$ and thus $\|\mathbf{U}\|_r \le \sqrt{\mu_0 r/n}$ as well as $\|\mathbf{V}\|_r \le \sqrt{\mu_0 r/n}$ by a similar argument. Observe that $\mathbf{U}\mathbf{V}^* = \left[\vec{u}_i \cdot \vec{v}_j\right]_{ij}$, where $\vec{u}_i$ and $\vec{v}_i$ are the i^{th} rows of $\mathbf{U}$ and $\mathbf{V}$, respectively. Hence, by Cauchy-Schwarz $\|\mathbf{U}\mathbf{V}^*\|_\infty \le \sqrt{\|\mathbf{U}\|_r\|\mathbf{V}\|_r} = \mu_0 r/n$, as desired. □

We present the following analog of Theorem 7 of Recht (4). It notes that $\mathbf{Z} \in \mathbb{R}^{n\times n}$ remains close in spectral norm to the maximum of its row and columns norm after the scaled sampling operator is applied.

Theorem 2.7 (Analog of Recht (4) Theorem 7)**.** *Let $\mathbf{Z}$ be an $n \times n$ matrix and $\mathcal{R}_\Omega$ the sampling operator that samples m entries of a matrix independently and uniformly with replacement, then*

$$\left\|\left(\frac{n^2}{m}\mathcal{R}_\Omega - \mathcal{I}\right)(\mathbf{Z})\right\| \le \max\left\{\frac{\mu_0 nr\beta\log n}{m}\frac{n}{\mu_0 r}\|\mathbf{Z}\|_\infty, \sqrt{\frac{2\mu_0 nr\beta\log n}{m}}\sqrt{\frac{n}{\mu_0 r}}\|\mathbf{Z}\|_b\right\} \tag{2.4}$$

with probability $1 - 2n^{1-\beta}$ for $\beta > 0$.

Proof. As in Recht (4), we decompose $(\frac{n^2}{m}\mathcal{R}_\Omega - \mathcal{I})(\mathbf{Z}) = \frac{1}{m}\sum_{j=1}^m n^2 Z_{a_j b_j}\vec{e}_{a_j}\vec{e}_{b_j}^* - \mathbf{Z}$ using $\Omega = \{(a_j, b_j)\}_{j=1}^m$ and proceed by applying Bernstein. We bound the operator norm

$$\|\mathbf{Z}\| = \sup_{\|x\|=1, \|y\|=1}\sum_{a,b} Z_{ab}y_a x_b \le \left(\sum_{a,b} Z_{ab}^2 y_a^2\right)^{1/2}\left(\sum_{a,b} x_b^2\right)^{1/2} \le \sqrt{n}\max_a\left(\sum_a Z_{ab}^2\right)^{1/2} \le \sqrt{n}\|\mathbf{Z}\|_b$$

We now bound each summand $\|n^2 Z_{a_j b_j}\vec{e}_{a_j}\vec{e}_{b_j}^* - \mathbf{Z}\|$ identically to Recht (4),

$$\left\|n^2 Z_{a_k b_k}\vec{e}_{a_k}\vec{e}_{b_k}^* - \mathbf{Z}\right\| \le \left\|n^2 Z_{a_k b_k}\vec{e}_{a_k}\vec{e}_{b_k}^*\right\| + \|\mathbf{Z}\| < \frac{3}{2}n^2\|\mathbf{Z}\|_\infty.$$

Furthermore, we bound

$$\left\|\mathbb{E}\left[\left(n^2 Z_{a_j b_j}\vec{e}_{a_j}\vec{e}_{b_j}^* - \mathbf{Z}\right)^*\left(n^2 Z_{a_j b_j}\vec{e}_{a_j}\vec{e}_{b_j}^* - \mathbf{Z}\right)\right]\right\| \le \left\|n^2\sum_{c,d} Z_{cd}^2\vec{e}_d\vec{e}_d^* - \mathbf{Z}^*\mathbf{Z}\right\|$$

$$\leq \max\left\{\left\|n^2\sum_{c,d} Z_{cd}^2 \vec{e}_d\vec{e}_d^*\right\|, \|\mathbf{Z}^*\mathbf{Z}\|\right\}$$
$$\leq n^2\|\mathbf{Z}\|_c^2 \leq n^2\|\mathbf{Z}\|_b^2$$

By a similar calculation we get $\left\|\mathbb{E}\left[(n^2 Z_{a_k b_k}\vec{e}_{a_k}\vec{e}_{b_k}^* - \mathbf{Z})(n^2 Z_{a_k b_k}\vec{e}_{a_k}\vec{e}_{b_k}^* - \mathbf{Z})^*)\right]\right\| \leq n^2\|\mathbf{Z}\|_r^2 \leq n^2\|\mathbf{Z}\|_b^2$. Now, we let $\sigma^2 = \frac{n^2}{m}\|\mathbf{Z}\|_b$ and $M = \frac{3}{2}\frac{n^2}{m}\|\mathbf{Z}\|_\infty$ and consider τ such that

$$\begin{aligned}
\tau &\geq \max\left\{\frac{2}{3}M\beta\log n, \sqrt{2\sigma^2\beta\log n}\right\} \\
&\geq \max\left\{\frac{n^2}{m}\|\mathbf{Z}\|_\infty\beta\log n, \sqrt{\frac{2n^2}{m}\|\mathbf{Z}\|_b^2\beta\log n}\right\} \\
&\geq \max\left\{\frac{\mu_0 nr\beta\log n}{m}\frac{n}{\mu_0 r}\|\mathbf{Z}\|_\infty, \sqrt{\frac{2\mu_0 nr\beta\log n}{m}}\sqrt{\frac{n}{\mu_0 r}}\|\mathbf{Z}\|_b\right\}.
\end{aligned}$$

By Theorem 2.4 we have that $\|({}^{n^2}\!/_m\mathcal{R}_\Omega - \mathcal{I})(\mathbf{Z})\| \leq \tau$, with probability at least $1 - 2n^{1-\beta}$, as desired. □

Finally, we will define the constants $0 \leq f_1, \ldots, f_n \leq 1$ for each of the n rows of $\mathbf{U}$ motivated by the maximum possible squared norm of any row ${}^{\mu_0 r}\!/_n$, induced by incoherence. Specifically, we let f_d denote the proportion of ${}^{\mu_0 r}\!/_n$ that the squared norm of column d of $\mathbf{U}$ is. More precisely, if $\vec{u}_d$ is the d^{th} row of $\mathbf{U}$ then

$$f_d := \|\vec{u}_d\|_2^2\left(\frac{\mu_0 r}{n}\right)^{-1} \implies \|\vec{u}_d\|_2^2 = f_d\left(\frac{\mu_0 r}{n}\right).$$

We also define $f_1', \ldots, f_n'$ similarly for $\mathbf{V}$.

Claim 2.8. For $f_1, \ldots, f_n$ and $f_1', \ldots, f_n'$ as defined above,

$$\sum_{d=1}^n f_d = \sum_{d=1}^n f_d' = \frac{n}{\mu_0}. \tag{2.5}$$

Proof. Recalling $\|\mathbf{U}\|_F^2 = \|\mathbf{V}\|_F^2 = r$, we have $\sum_{d=1}^n \|\vec{u}_d\|_2^2 = \sum_{d=1}^n f_d\,({}^{\mu_0 r}\!/_n) = r$ and the claim immediately follows. □

3 Main Result

We provide the following lemma to be used in the refinement of Recht (4) Theorem 2. Briefly, it states that the maximum of row and column norms of the dual variables drop geometrically through the iterations of their construction.

Lemma 3.1. *Let $\{\mathbf{W}_k\}_{k=1}^n$ be the variables defined in the dual construction above with each iteration having sample size q, then*

$$\|\mathbf{W}_k\|_b \leq 2^{-k}\|\mathbf{W}_0\|_b = 2^{-k}\sqrt{\frac{\mu_0 r}{n}} \tag{3.1}$$

with probability $1 - 12n^{2-\beta}$ provided $q \geq C_1\mu_0 nr^{3/2}\beta\log n$ for some $C_1 > 0$.

Proof. Recall the construction of the dual variables, originally from Gross (3): $\mathbf{W}_0 = \mathbf{U}\mathbf{V}^*$, $\mathbf{Y}_k = \frac{n^2}{q}\sum_{j=1}^k \mathcal{R}_{\Omega_j}(\mathbf{W}_{j-1})$, $\mathbf{W}_k = \mathbf{U}\mathbf{V}^* - \mathcal{P}_\mathcal{T}(\mathbf{Y}_k)$ for $k = 1, \ldots, p$. The triangle inequality and our definition of $\mathcal{P}_\mathcal{T}$ together imply $\forall\mathbf{X} \in \mathbb{R}^{n\times n}$

$$\|\mathcal{P}_\mathcal{T}(\mathbf{X})\|_b \leq \|P_\mathbf{U}(\mathbf{X})\|_b + \|P_\mathbf{V}(\mathbf{X})\|_b + \|\mathcal{P}_{\mathbf{U}\mathbf{V}^*}(\mathbf{X})\|_b. \tag{3.2}$$

We recognize that $\mathbf{W}_k$ may be expressed recursively as

$$\mathbf{W}_k = \mathbf{W}_{k-1} - \frac{n^2}{q}\mathcal{P}_\mathcal{T}\mathcal{R}_{\Omega_k}(\mathbf{W}_{k-1}) = \mathcal{P}_\mathcal{T}\left(\mathbf{W}_{k-1} - \frac{n^2}{q}\mathcal{R}_{\Omega_k}(\mathbf{W}_{k-1})\right) = \mathcal{P}_\mathcal{T}\left(\left(\mathcal{I} - \frac{n^2}{q}\mathcal{R}_{\Omega_k}\right)(\mathbf{W}_{k-1})\right)$$

where $\mathbf{X}_k := (\mathcal{I} - \frac{n^2}{q}\mathcal{R}_{\Omega_k})(\mathbf{W}_{k-1})$. Note that $\mathbb{E}[\mathbf{X}_k] = \mathbf{0}$, which implies $\mathbb{E}[\mathbf{W}_k] = \mathbf{0}$. We first inspect $\mathbf{W}_k^{(\mathbf{U})} := P_\mathbf{U}(\mathbf{X}_k)$, $\mathbf{W}_k^{(\mathbf{V})} := P_\mathbf{V}(\mathbf{X}_k)$, and $\mathbf{W}_k^{(\mathbf{U}\mathbf{V}^*)} := \mathcal{P}_{\mathbf{U}\mathbf{V}^*}(\mathbf{X}_k)$ separately, and then apply equation 3.2.

We take $\mathbf{W}_{k-1} \le 2^{-(k-1)}\sqrt{\mu_0 r/n}$ by our inductive hypothesis. Furthermore, Recht (4) Lemma 8 gives us $\|\mathbf{W}_k\|_\infty \le 2^{-k}\|\mathbf{W}_0\|_\infty \le 2^{-k}\,(\mu_0 r/n)$ for q as in our lemma statement. For the remainder of this proof we let $\mathbf{Z}$ and $\mathbf{Z}'$ represent $\mathbf{W}_{k-1}$ and $P_{\mathbf{U}}(\mathbf{W}_{k-1})$, respectively.

We now bound the *row* and *column* norms of $\mathbf{W}_k^{(\mathbf{U})}$ by inspecting one iteration of the dual construction process acting upon $\mathbf{Z}$ and then applying our inductive hypothesis. Here, we employ the technique used in Lemma 8 of Recht (4) wherein we examine the individual entries resulting from one iteration of dual construction, but now inspect only the effects of $P_{\mathbf{U}}$. In particular, we examine entry (c,d) of the resultant matrix $P_{\mathbf{U}}((\frac{n^2}{q}\mathcal{R}_{\Omega_k} - \mathcal{I}))(\mathbf{Z})) = (\frac{n^2}{q}P_{\mathbf{U}}\mathcal{R}_{\Omega_k} - P_{\mathbf{U}})(\mathbf{Z})$. Specifically, for index entry (c,d) of resultant matrix we consider sampling some (a,b) uniformly at random to define

$$\xi_{cd} = \left\langle \vec{e}_c\vec{e}_d^*, n^2\left\langle \vec{e}_a\vec{e}_b^*, \mathbf{Z}\right\rangle P_{\mathbf{U}}(\vec{e}_a\vec{e}_b^*) - \mathbf{Z}'\right\rangle$$

We note that $\mathbb{E}\left[\xi_{cd}\right] = 0$, and now proceed to find an upper bound M_{cd} for $|\xi_{cd}|$. We proceed

$$\begin{aligned}\xi_{cd} &= \left\langle \vec{e}_c\vec{e}_d^*, n^2\left\langle \vec{e}_a\vec{e}_b^*, \mathbf{Z}\right\rangle P_{\mathbf{U}}(\vec{e}_a\vec{e}_b^*) - \mathbf{Z}'\right\rangle \\ &= n^2 Z_{ab}\left\langle \vec{e}_c\vec{e}_d, P_{\mathbf{U}}(\vec{e}_a\vec{e}_b)\right\rangle - Z'_{cd}. && (\langle A, \vec{e}_i\vec{e}_j^*\rangle = A_{ij})\end{aligned}$$

We begin by bounding the magnitude of the first term. Observe that $\forall b \ne d$ we have $\left\langle \vec{e}_c\vec{e}_d^*, P_{\mathbf{U}}(\vec{e}_a\vec{e}_b^*)\right\rangle = 0$ and so we concern ourselves only when $b = d$. Letting $\mathbf{P} = P_{\mathbf{U}} = \mathbf{U}\mathbf{U}^*$, we have

$$\begin{aligned}|n^2 Z_{ad}\left\langle \vec{e}_c\vec{e}_d, P_{\mathbf{U}}(\vec{e}_a\vec{e}_d)\right\rangle| &= n^2|Z_{ad}||\left\langle P_{\mathbf{U}}(\vec{e}_c\vec{e}_d), \vec{e}_a\vec{e}_d\right\rangle| \\ &= n^2|Z_{ad}||P_{ca}| \\ &\le n^2\|\vec{z}^{(d)}\|_\infty \max_a\{|P_{ca}|\} \\ &\le n^2\sqrt{f_c}\frac{\mu_0 r}{n}\|\vec{z}^{(d)}\|_\infty && \text{(Definition of } f_c) \\ &= \mu_0 r n\sqrt{f_c}\|\vec{z}^{(d)}\|_\infty.\end{aligned}$$

Where we use $\max_a\{|P_{ca}|\} = \|\vec{u}_c\|_2 \cdot \max_a\|\vec{u}_a\|_2 = \sqrt{f_c \cdot \mu_0 r/n}\sqrt{\mu_0 r/n} = \sqrt{f_c}\cdot \mu_0 r/n$. Using $\mathbf{Z}' = P_{\mathbf{U}}\mathbf{Z} = \mathbf{U}\mathbf{U}^*\mathbf{Z}$ we observe that second term $Z'_{cd} = \vec{p}_c \cdot \vec{z}^{(d)}$ may be bounded

$$|Z'_{cd}| \le \sqrt{\|\vec{p}_c\|_2^2 \cdot \|\vec{z}^{(d)}\|_2^2} \le \sqrt{f_c\frac{\mu_0 r}{n}\cdot n\|\vec{z}^{(d)}\|_\infty^2} \le \sqrt{\mu_0 r f_c}\|\vec{z}^{(d)}\|_\infty \le \mu_0 r n\sqrt{f_c}\|\vec{z}^{(d)}\|_\infty$$

Using the triangle inequality we now find M_{cd}

$$|\xi_{cd}| \le |n^2 Z_{ad}\left\langle \vec{e}_c\vec{e}_d, P_{\mathbf{U}}(\vec{e}_a\vec{e}_d)\right\rangle| + |Z'_{cd}| \le 2\mu_0 r n\sqrt{f_c}\|\vec{z}^{(d)}\|_\infty =: M_{cd}. \tag{3.3}$$

We proceed with finding our variance term $\mathbb{E}\left[\xi_{cd}^2\right]$ by

$$\begin{aligned}\mathbb{E}\left[\xi_{cd}^2\right] &= \frac{1}{n^2}\sum_{a=1}^n\sum_{b=1}^n\left\langle \vec{e}_c\vec{e}_d^*, n^2\left\langle \vec{e}_a\vec{e}_b^*, \mathbf{Z}\right\rangle P_{\mathbf{U}}(\vec{e}_a\vec{e}_b^*) - \mathbf{Z}'\right\rangle^2 \\ &= n^2\sum_{a=1}^n\sum_{b=1}^n\left\langle P_{\mathbf{U}}(\vec{e}_c\vec{e}_d^*), \vec{e}_a\vec{e}_b^*\right\rangle^2\left\langle \vec{e}_a\vec{e}_b^2, \mathbf{Z}\right\rangle^2 - Z_{cd}'^2 && (\langle A, \vec{e}_i\vec{e}_j^*\rangle = A_{ij}) \\ &= n^2\sum_{a=1}^n\left\langle P_{\mathbf{U}}(\vec{e}_c\vec{e}_d^*), \vec{e}_a\vec{e}_d^*\right\rangle^2\left\langle \vec{e}_a\vec{e}_d^2, \mathbf{Z}\right\rangle^2 - Z_{cd}'^2 && (\forall b \ne d : \left\langle P_{\mathbf{U}}(\vec{e}_c\vec{e}_d^*), \vec{e}_a\vec{e}_b^*\right\rangle = 0) \\ &= n^2\sum_{a=1}^n P_{ca}^2 Z_{ad}^2 - Z_{cd}'^2 \le n^2\sum_{a=1}^n P_{ca}^2 Z_{ad}^2.\end{aligned}$$

We make the following observation

$$\mathbb{E}\left[\xi_{cd}^2\right] \le n^2\sum_{a=1}^n P_{ca}^2 Z_{ad}^2 \le n^2\|\vec{z}^{(d)}\|_\infty^2\sum_{a=1}^n P_{ca}^2 = n^2\|\vec{z}^{(d)}\|_\infty^2 f_c\frac{\mu_0 r}{n} = \mu_0 r n f_c\|\vec{z}^{(d)}\|_\infty^2. \tag{3.4}$$

Notice that each entry of our resultant matrix is distributed like $\frac{1}{q}\sum_{k=1}^q \xi_{cd}^{(k)} =: \zeta_{cd}$ where $\xi_{cd}^{(k)}$ are i.i.d copies of ξ_{cd}. By Bernstein's inequality, specifically Corollary 2.5, we get

$$|\zeta_{cd}| \le \max\left\{\sqrt{\frac{\mathbb{E}\left[\xi_{cd}^2\right]\beta\log n}{q}}, \frac{M_{cd}\beta\log n}{q}\right\}.$$

with probability $1 - 2n^{1-\beta}$. This allows us to make the following claim which is an extension of Recht (4) Lemma 8.

Claim 3.2 (Extension of Recht (4) Lemma 8). With probability $1 - 2n^{1-\beta}$ for q as above we have

$$\|\vec{w}_k^{(d)}\|_\infty \le 2^{-k}\|\vec{w}_0^{(d)}\|_\infty \le 2^{-k}\sqrt{f_d}\left(\frac{\mu_0 r}{n}\right). \tag{3.5}$$

Proof. We observe this by first applying 3.3 and 3.4 in our bound for $|\zeta_{cd}|$ (recalling $\mathbf{Z} = \mathbf{W}_{k-1}$), giving us

$$|\zeta_{cd}| \le \max\left\{\sqrt{\frac{\mu_0 rn\beta\log n}{q}}, \frac{2\mu_0 rn\beta\log n}{q}\right\}\sqrt{f_c}\|\vec{w}_{k-1}^{(d)}\|_\infty.$$

Now, $\|\vec{z}^{(d)}\|_\infty = \max_c|\zeta_{cd}|$ (recall $|f_c| \le 1$) and so for a value of $q \ge C_1\mu_0 nr\beta\log n$ we get $\|\vec{w}_k^{(d)}\|_\infty \le \frac{1}{2}\|\vec{w}_{k-1}^{(d)}\|_\infty$. We consider $\mathbf{W} = \mathbf{W}_0 = \mathbf{U}\mathbf{U}^*$ and in particular some entry $W_{cd} = \vec{u}_c \cdot \vec{u}_d$ and see

$$|W_{cd}| \le \sqrt{\|\vec{u}_c\|_2^2\|\vec{u}_d\|_2^2} = \sqrt{f_c\left(\frac{\mu_0 r}{n}\right)f_d\left(\frac{\mu_0 r}{n}\right)} \le \sqrt{\min(f_c, f_d)}\left(\frac{\mu_0 r}{n}\right)$$

hence $\|\vec{w}^{(d)}\|_\infty \le \sqrt{f_d}\,(\mu_0 r/n)$.Our claim now immediately follows by induction. □

Now, we bound our *squared* row norm for some row c by

$$\begin{aligned}\sum_{d=1}^n|\zeta_{cd}|^2 &\le \sum_{d=1}^n\max\left\{\sqrt{\frac{\mathbb{E}\left[\xi_{cd}^2\right]\beta\log n}{q}}, \frac{M_{cd}\beta\log n}{q}\right\}^2\\ &\le 2\max\left\{\frac{\beta\log n}{q}\sum_{d=1}^n\mathbb{E}\left[\xi_{cd}^2\right], \frac{\beta^2\log^2 n}{q^2}\sum_{d=1}^n M_{cd}^2\right\}\end{aligned}$$

with probability $1 - 2n^{2-\beta}$, using union bound. We bound the first summation in the maximum

$$\begin{aligned}\sum_{d=1}^n\mathbb{E}\left[\xi_{cd}^2\right] &\le \sum_{d=1}^n n^2\sum_{a=1}^n P_{ca}^2 Z_{ad}^2\\ &= n^2\sum_{a=1}^n P_{ca}^2\sum_{d=1}^n Z_{ad}^2\\ &\le n^2\sum_{a=1}^n P_{ca}^2\|\mathbf{Z}\|_b^2\\ &\le n^2\left(\frac{\mu_0 r}{n}\right)\|\mathbf{Z}\|_b^2 && \left(\textstyle\sum_{a=1}^n P_{ca}^2 \le \mu_0 r/n \text{ by } \mathbf{A0}\right)\\ &\le 4^{-(k-1)}\mu_0 nr\left(\frac{\mu_0 r}{n}\right) && \text{(Inductive Hypothesis)}\end{aligned}$$

Similarly, we bound the second summation

$$\begin{aligned}\sum_{d=1}^n M_{cd}^2 &\le \sum_{d=1}^n\left(\mu_0 rn\|\vec{z}^{(d)}\|_\infty\right)^2\\ &\le \sum_{d=1}^n\mu_0^2 r^2 n^2\|\vec{z}^{(d)}\|_\infty^2\\ &\le \mu_0^2 r^2 n^2\sum_{d=1}^n 4^{-(k-1)}f_d\left(\frac{\mu_0 r}{n}\right)^2 && \text{(by Claim 3.2)}\\ &\le 4^{-(k-1)}\mu_0^2 r^2 n^2\left(\frac{\mu_0 r}{n}\right)^2\sum_{d=1}^n f_d\\ &\le 4^{-(k-1)}\mu_0^2 r^2 n^2\left(\frac{\mu_0 r}{n}\right)^2\frac{n}{\mu_0} && \text{(by Claim 2.8)}\\ &\le 4^{-(k-1)}\mu_0^2 r^3 n^2\left(\frac{\mu_0 r}{n}\right)\end{aligned}$$

Hence, with probability $1-2n^{2-\beta}$ our squared row norm can be bounded

$$\sum_{d=1}^{n}|\zeta_{cd}|^2 \leq 2\max\left\{\frac{\mu_0 nr\beta\log n}{q}, \frac{\mu_0^2 n^2 r^3\beta^2\log^2 n}{q^2}\right\}4^{-(k-1)}\left(\frac{\mu_0 r}{n}\right).$$

By the symmetry of $\mathbf{P}$, a similar calculation for some fixed column d produces an identical bound for $\sum_{c=1}^{n}|\zeta_{cd}|^2$. Each row and column is identically distributed to every other row and column, respectively, whence

$$\|\mathbf{W}_k^{(\mathbf{U})}\|_b^2 \leq \max\left\{\sum_{c=1}^{n}|\zeta_{cd}|^2, \sum_{d=1}^{n}|\zeta_{cd}|^2\right\} \leq 2\max\left\{\frac{\mu_0 nr\beta\log n}{q}, \frac{\mu_0^2 n^2 r^3\beta^2\log^2 n}{q^2}\right\}4^{-(k-1)}\left(\frac{\mu_0 r}{n}\right).$$

with probability $1-4n^{3-\beta}$, by union bound. Thus, for $q \geq C_1\mu_0 nr^{3/2}\beta\log n$ for some $C_1 > 0$ we obtain that $\|\mathbf{W}_k\|_b^2 \leq \frac{4^{-k}}{9}\left(\mu_0 r/n\right)$, with aforesaid probability.

As might be expected, the analysis of $\mathbf{W}_k^{(\mathbf{V})}$ produces the same result and is performed nearly-identical to that above, but with $P_{\mathbf{V}}$ in place of $P_{\mathbf{U}}$. We leave our analysis of $\mathcal{P}_{\mathbf{UV}^*}$ to Appendix A, where we follow an entry-wise analysis similar to the above and, again, results identically $\mathbf{W}_k^{(\mathbf{UV}^*)}$. Consequently, we have that $\|\mathbf{W}_k^{(\mathbf{U})}\|, \|\mathbf{W}_k^{(\mathbf{V})}\|, \|\mathbf{W}_k^{(\mathbf{UV}^*)}\| \leq \frac{2^{-k}}{3}\sqrt{\mu_0 r/n}$ each with probability $1-4n^{3-\beta}$. We thus conclude by equation 3.2 that

$$\|\mathbf{W}_k\|_b \leq \|\mathbf{W}_k^{(\mathbf{U})}\| + \|\mathbf{W}_k^{(\mathbf{V})}\| + \|\mathbf{W}_k^{(\mathbf{UV}^*)}\| \leq 2^{-k}\sqrt{\frac{\mu_0 r}{n}}$$

with probability $1-12n^{3-\beta}$, by union bound, given $q \geq C_1\mu_0 nr^{3/2}\beta\log n$ for $\beta > 0$ and some $C_1 > 0$, as desired. □

Using the above derivations, we present the following modification of Theorem 2 of Recht (4) which is the main result of this report.

Theorem 3.3 (Refined Theorem 2 of Recht (4))**.** *Let* $\mathbf{M} \in \mathbb{R}^{n\times n}$ *be a matrix of rank* r *such that* ***A0*** *holds, or* $\mu(\mathbf{M}) \leq \mu_0$*. Given set* Ω *of* m *entries of* $\mathbf{M}$ *sampled uniformly at random with*

$$m \geq C\mu_0 nr^{3/2}\log^2 n = O(\mu_0 nr^{3/2}\log^2 n),$$

the solution to the nuclear norm minimization program (1.1) is exactly $\mathbf{M}$ *with probability at least* $1-36\log n\cdot n^{2-2\beta} - n^{2-2\beta^{1/2}}$ *for* $\beta > 0$ *and some* $C > 0$*.*

Proof. We refer the to the proof of Theorem 2 in Recht (4). We retain his setup and analysis until where it is shown that $\left\|\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{Y}_p)\right\| \leq 1/2$, where improvement can be made by applying our Theorem 2.7 in place of his Theorem 7. Following Recht, we now have

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{Y}_p)\| &\leq \sum_{j=1}^{p}\left\|\frac{n^2}{q}\mathcal{P}_{\mathcal{T}_\perp}\mathcal{R}_{\Omega_j}\mathbf{W}_{j-1}\right\| \\
&= \sum_{j=1}^{p}\left\|\mathcal{P}_{\mathcal{T}_\perp}\left(\frac{n^2}{q}\mathcal{R}_{\Omega_j}\mathbf{W}_{j-1} - \mathbf{W}_{j-1}\right)\right\| && (\mathbf{W}_{j-1} \in T) \\
&\leq \sum_{j=1}^{p}\left\|\left(\frac{n^2}{q}\mathcal{R}_{\Omega_j} - \mathcal{I}\right)(\mathbf{W}_{j-1})\right\| && (\|\mathcal{P}_{\mathcal{T}_\perp}(\mathbf{Z})\| \leq \|\mathbf{Z}\|) \\
&\leq \sum_{j=1}^{p}\max\left\{\frac{\mu_0 nr\log n}{q}\frac{n}{\mu_0 r}\|\mathbf{W}_{j-1}\|_\infty, \sqrt{\frac{2\mu_0 nr\log n}{q}}\sqrt{\frac{n}{\mu_0 r}}\|\mathbf{W}_{j-1}\|_b\right\} && \text{(Theorem 2.7)} \\
&\leq \sum_{j=1}^{p}\max\left\{2^{-(j-1)}\frac{\mu_0 nr\beta\log n}{q}\frac{n}{\mu_0 r}\|\mathbf{W}_0\|_\infty, \sqrt{\frac{2\mu_0 nr\beta\log n}{q}}\sqrt{\frac{n}{\mu_0 r}}\|\mathbf{W}_{j-1}\|_b\right\} && \text{(Recht (4) Lemma 8)} \\
&\leq 2\sum_{j=1}^{p}2^{-j}\max\left\{\frac{\mu_0 nr\beta\log n}{q}\frac{n}{\mu_0 r}\|\mathbf{UV}^*\|_\infty, \sqrt{\frac{2\mu_0 nr\beta\log n}{q}}\sqrt{\frac{n}{\mu_0 r}}\|\mathbf{UV}^*\|_b\right\} && \text{(Lemma 3.1)} \\
&\leq 2\sum_{j=1}^{p}2^{-j}\max\left\{\frac{\mu_0 nr\beta\log n}{q}, \sqrt{\frac{2\mu_0 nr\beta\log n}{q}}\right\}
\end{aligned}$$

$$\leq \max\left\{\frac{2\mu_0 nr\beta\log n}{q}, \sqrt{\frac{4\mu_0 nr\beta\log n}{q}}\right\} \leq \frac{1}{2}$$

where the Theorem 2.7 holds for $q \geq C_1\mu_0 nr^{3/2}\beta\log n$, and Recht (4) Lemma 8 and the final inequality hold for $q \geq C_1\mu_0 nr\beta\log n$. Thus, the above holds under $q \geq C_1\mu_0 nr^{3/2}\beta\log n$ and therefore $m \geq O(\mu_0 nr^{3/2}\beta\log^2 n)$ (recall $p \geq O(\log n)$ and $m = pq$). Observe that we obviate the need for **A1** by using $\|\mathbf{UV}^*\|_\infty = {}^{\mu_0 r}/_n$ rather than $\|\mathbf{UV}^*\|_\infty = {}^{\mu_1\sqrt{r}}/_n$ as used by Recht (4) which requires **A1**.

Now we may preserve and apply the remainder of the proof of Recht (4) Theorem 2, only needing to modify the probability our cited events occur. In addition to the events invoked by Recht (4), each of which maintain their probability of failure for m as in the theorem statement, we incur $12n^{3-2\beta}$ probability of failure of Equation 3.1 in Lemma 3.1 for all $k = 1, \ldots, p$ given said m. Applying union bound, our desired events now hold with probability

$$1 - 18\log n \cdot n^{3-2\beta} - n^{2-2\beta^{1/2}}$$

for our value of m and $\beta > 1$, precisely as desired. □

Thus, we have demonstrated sample complexity results of $m \geq O(\mu_0 nr^{3/2}\beta\log^2 n)$ for *general* (that is, requiring only that the necessary condition **A0** holds) low-rank incoherent matrix recovery, with high probability.

4 Discussion

Observe that in our application of Bernstein's inequality in Theorem 2.7, we apply *both* the variance and the maximum term (as opposed to just the variance term). Our advantage is gained in observing that we may use the maximum of row and column norms in bounding the scaled sampling operator norm, seen in the variance term; our probability of correctness decreases very slightly in this process. Prior to this, past works relied on the infinity norms of the dual variables, whereas we may only apply the trivial assumption on **A1** that follows from **A0**, $\mu_1 = \mu_0\sqrt{r}$.

We make two observations regarding the value of μ_1 in certain matrices which we may wish to complete, demonstrating that μ_1 may not necessarily be as small as desired (that is, $\mu_1 = O(1)$). The first concerns a simplified model of random matrices which were studied by Candès and Recht (1) and Candès and Tao (2).

Observation 4.1 ($\mathbf{U}, \mathbf{V} \in_R \{\pm{}^1/\sqrt{n}\}^{n\times r}$). [2] Suppose $\mathbf{M}$ is a matrix such that the entries of $\mathbf{U}$ and $\mathbf{V}$ are $\pm{}^1/\sqrt{n}$ with equal probability. Then $\mu_0 = 1$ and when $r = \Omega(\log n)$ it follows that $\mathbb{E}[\mu_1] = \Theta(\sqrt{\log n})$.

Proof. We consider $\mathbf{U}', \mathbf{V}' \in_R \{\pm 1\}^{n\times r}$ and let $\mathbf{U} = {}^1/\sqrt{n}\cdot\mathbf{U}'$ and $\mathbf{V} = {}^1/\sqrt{n}\cdot\mathbf{V}'$; clearly, $\mu_0 = 1$. We seek the expectation of $\mu_1 = {}^n/\sqrt{r}\cdot\|\mathbf{UV}^*\|_\infty = {}^1/\sqrt{r}\cdot\|\mathbf{U}'\mathbf{V}'^*\|_\infty$. Each entry of $\mathbf{U}'$ and $\mathbf{V}'$ is an i.i.d. Rademacher random variable and so each entry of $\mathbf{U}'\mathbf{V}'^*$ is equivalent to the sum of r i.i.d. Rademachers. Our problem of finding $\mathbb{E}[\|\mathbf{U}'\mathbf{V}'^*\|_\infty]$ is thus precisely the problem of finding the expected maximum of n simple symmetric random walks of length r on $\mathbb{Z}$. Rademacher random variables are 1-sub-Gaussian and so each entry of $\mathbf{U}'\mathbf{V}'^*$ is r-sub-Gaussian. Applying the maximal inequality (see Rigollet (13) Theorem 1.14) across all n^2 entries, we have $\mathbb{E}[\|\mathbf{U}'\mathbf{V}'^*\|_\infty] \leq \sqrt{4r\log(2n)}$. Furthermore, Orabana and Pal (14) show that for $r \geq 3\log n$, that expected maximum of n simple symmetric length-r random walks on $\mathbb{Z}$ will be $\Omega(\sqrt{r\log n})$. Therefore, we have $\mathbb{E}[\|\mathbf{UV}^*\|_\infty] = \Theta(\sqrt{r\log n})$ and so $\mathbb{E}[\mu_1] = \Theta(\sqrt{\log n})$. □

Remark 4.2. Assuming sufficiently large r, we use $\mathcal{N}(0, r)$ to approximate each entry by the central limit theorem. Kamath (15) shows that the expected maximum of n i.i.d. random variables with distribution $\mathcal{N}(0, \sigma^2)$ is $\Theta(\sigma\sqrt{\log n})$. If our approximation of $\mathcal{N}(0, r)$ suffices across n^2 entries we find $\mathbb{E}[\|\mathbf{U}'\mathbf{V}'^*\|_\infty] = \Theta(\sqrt{r\log n})$ and so $\mathbb{E}[\mu_1] = \Theta(\sqrt{\log n})$. We might expect in cases of large r, even when $r = o(\log n)$ [3], that $\mathbb{E}[\mu_1] = \Theta(\sqrt{\log n})$.

The second observation is regarding positive semi-definite (PSD) matrices.

Observation 4.3. If $\mathbf{M}$ is a PSD matrix, then $\mu_1 = \mu_0\sqrt{r}$.

Proof. We consider some PSD matrix $\mathbf{M}$ with coherence μ_0. Observe that the SVD of $\mathbf{M}$ is $\mathbf{U\Sigma V}^* = \mathbf{U\Sigma U}^*$, or $\mathbf{V} = \mathbf{U}$. Recall our alternative definition of coherence, $\mu(\mathbf{U}) = \max_{1\leq i\leq n}\|\vec{u}_i\|_2^2$. The $(i,i)^{\text{th}}$ entry of $\mathbf{UU}^*$ is precisely $\|\vec{u}_i\|_2^2$, thus $\|\mathbf{UV}^*\|_\infty = \|\mathbf{UU}^*\|_\infty \geq \max_{1\leq i\leq n}\|\vec{u}_i\|_2^2 = {}^{\mu_0 r}/_n$ and so $\mu_1 = \mu_0\sqrt{r}$ (recall $\mu_1 \leq \mu_0\sqrt{r}$ trivially by **A0**). □

---

[2] The notation $x \in_R S$ says x is an element selected uniformly at random from S.

[3] We note that simulating n for sufficiently large r is often computationally intractable — e.g. taking $r \sim \log\log n$ and so $n \sim 2^{2^r}$.

In the above cases, we see that we can produce $\mu_1 \neq O(1)$ even for maximally coherent matrices. It is here where we expect to see improvement in sample complexity over Recht (4). The bottleneck in our above analysis is incurred in the sum of squared maximums in Lemma 3.1 where an extra factor of r is incurred. Inspecting various example matrices, bounding this term may serve as the basis of potential further work.

References

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

[2] E. J. Candès and T. Tao, "The power of convex relaxation: near-optimal matrix completion," *IEEE Trans. Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

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

[4] B. Recht, "A simpler approach to matrix completion," *Journal of Machine Learning Research*, vol. 12, pp. 3413–3430, 2011.

[5] L. T. Nguyen, J. Kim, S. Kim, and B. Shim, "Localization of iot networks via low-rank matrix completion," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5833–5847, 2019.

[6] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *Systems Control Letters*, vol. 62, no. 8, pp. 605–612, 2013.

[7] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[8] M. Fazel, *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[9] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6, pp. 4734–4739, IEEE, 2001.

[10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[11] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd international conference on Machine learning*, pp. 281–288, 2006.

[12] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint 2, 1-norms minimization," *Advances in neural information processing systems*, vol. 23, 2010.

[13] P. Rigollet, "18.s997: High dimensional statistics lecture notes." MIT OpenCourseWare, July 2015.

[14] F. Orabona and D. Pal, "Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice," 2015.

[15] G. Kamath, "Bounds on the expectation of the maximum of samples from a gaussian," 2015.

A Analysis of $\mathcal{P}_{\mathbf{UV}^*}$

We now inspect the isolated effect of $\mathcal{P}_{\mathbf{UV}^*}$ (recall $\mathcal{P}_{\mathbf{UV}^*}(\mathbf{Z}) := P_{\mathbf{U}} X P_{\mathbf{V}}$) in our dual construction process, similarly to how we did $P_{\mathbf{U}}$. We wish to show $\|\mathbf{W}_k^{(\mathbf{UV}^*)}\| \leq 2^{-k}\,(\mu_0 r/n)$, with high probability. We proceed immediately to modeling individual entries of our resultant matrix of interest, $(\mathcal{P}_{\mathbf{UV}^*}(\frac{n^2}{q}\mathcal{R}_{\Omega_k} - \mathcal{I}))(\mathbf{W}_{k-1}) = (\frac{n^2}{q}\mathcal{P}_{\mathbf{UV}^*}\mathcal{R}_{\Omega} - \mathcal{P}_{\mathbf{UV}^*})(\mathbf{W}_{k-1})$.

Now, let $\mathbf{Z}$ and $\mathbf{Z}'$ represent $\mathbf{W}_{k-1}$ and $\mathcal{P}_{\mathbf{UV}^*}(\mathbf{W}_{k-1})$, respectively. For matrix index (c,d), consider sampling (a,b) uniformly at random to define random variable ξ'_{cd}, similar to the variable ξ_{cd} above. We see

$$\xi'_{cd} = \left\langle \vec{e}_c\vec{e}_d^*, n^2 \left\langle \vec{e}_a\vec{e}_b^*, \mathbf{Z}\right\rangle \mathcal{P}_{\mathbf{UV}^*}(\vec{e}_a\vec{e}_b^*) - \mathbf{Z}'\right\rangle$$

Observe that $\mathbb{E}\left[\xi'_{cd}\right] = 0$ as before. We find the maximum magnitude of ξ'_{cd}, $M_c d'$, by decomposing the above form as is done for ξ_{cd}. Again letting $\mathbf{P} = P_{\mathbf{U}} = \mathbf{UU}^*$ and now $\mathbf{P}' = P_{\mathbf{V}} = \mathbf{VV}^*$, we bound the first term

$$|n^2 Z_{ab}\left\langle \vec{e}_c\vec{e}_d^*, \mathcal{P}_{\mathbf{UV}^*}(\vec{e}_a\vec{e}_b)\right\rangle| = n^2|Z_{ab}||\langle \mathcal{P}_{\mathbf{UV}^*}(\vec{e}_c\vec{e}_d), \vec{e}_a\vec{e}_b\rangle|$$

$$
\begin{aligned}
&= n^2 |Z_{ab}||P_{ca}||P'_{db}| \\
&\le n^2 \|\mathbf{Z}\|_\infty \max_a\{|P_{ca}|\} \max_b\{|P_{db}|\} \\
&\le n^2 \sqrt{f_c}\left(\frac{\mu_0 r}{n}\right)\sqrt{f'_d}\left(\frac{\mu_0 r}{n}\right)\|\mathbf{Z}\|_\infty \\
&\le \mu_0^2 r^2 \sqrt{f_c f'_d}\|\mathbf{Z}\|_\infty \le \mu_0 r n \sqrt{f_c f'_d}\|\mathbf{Z}\|_\infty
\end{aligned}
$$

Using $\mathbf{Z}' = \mathcal{P}_{\mathbf{UV}^*}(\mathbf{Z}) = P_{\mathbf{U}}\mathbf{Z}P_{\mathbf{V}}$ we bound the second term

$$
\begin{aligned}
Z'_{cd} &= \sum_{k=1}^n (\mathbf{PZ})_{ck} P'_{kd} \\
&\le \sqrt{\sum_{k=1}^n (\mathbf{PZ})_{ck}^2 \sum_{k=1}^n {P'_{kd}}^2} \\
&\le \sqrt{\sum_{k=1}^n \|\vec{p}_c\|_2^2 \|\vec{z}^{(k)}\|_2^2 \cdot f'_d \left(\frac{\mu_0 r}{n}\right)^2} \\
&\le \sqrt{\sum_{k=1}^n f_c \left(\frac{\mu_0 r}{n}\right) \|\vec{z}^{(k)}\|_2^2 \cdot f'_d \left(\frac{\mu_0 r}{n}\right)} \\
&\le \left(\frac{\mu_0 r}{n}\right) \sqrt{f_c f'_d \sum_{k=1}^n \|\vec{z}^{(k)}\|_2^2} \\
&\le \left(\frac{\mu_0 r}{n}\right) \sqrt{f_c f'_d}\sqrt{n^2 \|\mathbf{Z}\|_\infty^2} \le \mu_0 r n \sqrt{f_c f'_d}\|\mathbf{Z}\|_\infty.
\end{aligned}
$$

Hence, by the triangle inequality we have $|\xi'_{cd}| \le 2\mu_0 r n \sqrt{f_c f'_d}\|\mathbf{Z}\|_\infty =: M'_{cd}$. We now find $\mathbb{E}\left[\xi'^2_{cd}\right]$ as follows.

$$
\begin{aligned}
\mathbb{E}\left[{\xi'_{cd}}^2\right] &= \frac{1}{n^2}\sum_{a=1}^n \sum_{b=1}^n \left\langle \vec{e}_c \vec{e}_d^*, n^2 \left\langle \vec{e}_a \vec{e}_b^*, \mathbf{Z}\right\rangle \mathcal{P}_{\mathbf{UV}^*}(\vec{e}_a \vec{e}_b^*) - \mathbf{Z}' \right\rangle^2 \\
&= n^2 \sum_{a=1}^n \sum_{b=1}^n \left\langle \mathcal{P}_{\mathbf{UV}^*}(\vec{e}_c \vec{e}_d^*), \vec{e}_a \vec{e}_b^* \right\rangle^2 \left\langle \vec{e}_a \vec{e}_b^2, \mathbf{Z}\right\rangle^2 - Z'^2_{cd} && (\langle A, \vec{e}_i \vec{e}_j^* \rangle = A_{ij}) \\
&= n^2 \sum_{a=1}^n \sum_{b=1}^n (P_{ac} P'_{db})^2 Z_{ab}^2 - Z'^2_{cd} \le n^2 \sum_{a=1}^n \sum_{b=1}^n P_{ac}^2 P'^2_{db} Z_{ab}^2 && (\mathcal{P}_{\mathbf{UV}^*}(\vec{e}_c \vec{e}_d^*) = \left[P_{ic} P'_{dj}\right]_{ij}).
\end{aligned}
$$

Entry (c,d) of our resultant matrix is distributed as $\frac{1}{m}\sum_{k=1}^m {\xi'_{cd}}^{(k)} =: \zeta'_{cd}$ where ${\xi'_{cd}}^{(k)}$ are i.i.d. copies of ξ'_{cd}. By Corollary 2.5 we have

$$
|\zeta'_{cd}| \le \max\left\{\sqrt{\frac{\mathbb{E}\left[\xi'^2_{cd}\right]\beta \log n}{q}}, \frac{M'_{cd}\beta \log n}{q}\right\}.
$$

with probability $1 - 2n^{1-\beta}$. As before, we may bound the squared row norm for any row c

$$
\begin{aligned}
\sum_{d=1}^n |\zeta'_{cd}|^2 &\le \sum_{d=1}^n \max\left\{\sqrt{\frac{\mathbb{E}\left[{\xi'_{cd}}^2\right]\log n}{q}}, \frac{M'_{cd}\log n}{q}\right\}^2 \\
&\le 2\max\left\{\frac{\log n}{q}\sum_{d=1}^n \mathbb{E}\left[{\xi'_{cd}}^2\right], \frac{\log^2 n}{q^2}\sum_{d=1}^n {M'_{cd}}^2\right\}
\end{aligned}
$$

with probability $1 - 2n^{2-\beta}$, using union bound. (Notice that in this case a column calculation will give the same result, hence precluding the need for an application of Recht (4) Theorem 6). We bound the first summation in the maximum

$$
\sum_{d=1}^n \mathbb{E}\left[{\xi'_{cd}}^2\right] \le \sum_{d=1}^n n^2 \sum_{a=1}^n \sum_{b=1}^n P_{ac}^2 P'^2_{db} Z_{ab}^2
$$

$$
\begin{aligned}
&= n^2 \sum_{a=1}^{n} P_{ac}^2 \sum_{b=1}^{n} Z_{ab}^2 \sum_{d=1}^{n} P_{db}'^2 \\
&\leq n^2 \left(\frac{\mu_0 r}{n}\right) \sum_{a=1}^{n} P_{ac}^2 \|\mathbf{Z}\|_b^2 \\
&\leq n^2 \left(\frac{\mu_0 r}{n}\right)^2 \|\mathbf{Z}\|_b^2 \\
&\leq 4^{-(k-1)} (\mu_0 r)^2 \left(\frac{\mu_0 r}{n}\right).
\end{aligned}
$$

Observe that $\sum_{d=1}^{n} \mathbb{E}\left[{\xi_{cd}'}^2\right] \leq 4^{-(k-1)} (\mu_0 r)^2 \left({}^{\mu_0 r}/_{n}\right) \leq 4^{-(k-1)} \mu_0 n r \left({}^{\mu_0 r}/_{n}\right) = \sum_{d=1}^{n} \mathbb{E}\left[\xi_{cd}^2\right]$. Now we may bound our second summation similarly to$\sum_{d=1}^{n} M_{cd}^2$:

$$
\begin{aligned}
\sum_{d=1}^{n} M_{cd}'^2 &\leq \sum_{d=1}^{n} \mu_0^2 r^2 n^2 f_c f_d' \|\mathbf{Z}\|_\infty^2 \\
&\leq 4^{-(k-1)} \mu_0^2 r^2 n^2 \left(\frac{\mu_0 r}{n}\right)^2 \sum_{d=1}^{n} f_d \\
&\leq 4^{-(k-1)} \mu_0^2 r^3 n^2 \left(\frac{\mu_0 r}{n}\right)
\end{aligned}
$$

(Note that we obtain the same result for $\sum_{c=1}^{n} M_{cd}'^2$.) Thus, we may use the same bound for $\sum_{d=1}^{n} |\zeta_{cd}'|^2$ as used for $\sum_{d=1}^{n} |\zeta_{cd}|^2$ (which may not be tight). It follows that, with probability $1 - 2n^{2-\beta}$,our squared row norm can be bounded

$$
\sum_{d=1}^{n} |\zeta_{cd}'|^2 \leq 2 \max\left\{\frac{\mu_0 n r \log n}{q}, \frac{\mu_0^2 n^2 r^3 \log^2 n}{q^2}\right\} 4^{-(k-1)} \left(\frac{\mu_0 r}{n}\right),
$$

thus with probability $1 - 4n^{3-\beta}$, by union bound and applying corresponding across rows c, we have

$$
\|\mathbf{W}_k^{(\mathbf{UV}^*)}\|_b^2 \leq 2 \max\left\{\frac{\mu_0 n r \log n}{q}, \frac{\mu_0^2 n^2 r^3 \log^2 n}{q^2}\right\} 4^{-(k-1)} \left(\frac{\mu_0 r}{n}\right).
$$

So, as was the case for $\mathbf{W}_k^{(\mathbf{U})}$, for $q \geq C_1 \mu_0 n r^{3/2} \log n$ we obtain $\|\mathbf{W}_k^{(\mathbf{UV}^*)}\|_b \leq 2^{-k} \left({}^{\mu_0 r}/_{n}\right)$, as desired.