Electronic Colloquium on Computational Complexity, Report No. 173 (2024)

Lower Bounds in the Query-with-Sketch Model and a Barrier in Derandomizing BPL

Songhua He
Rutgers University
sh1511@scarletmail.rutgers.edu

Yuanzhi Li
Carnegie Mellon University
yuanzhil@andrew.cmu.edu

Periklis A. Papakonstantinou
Rutgers University
periklis.research@gmail.com

Xin Yang
Snap Inc.
yx1992@cs.washington.edu

Abstract

This work makes two distinct yet related contributions. The first contribution is a new information-theoretic model, the query-with-sketch model, and tools to show lower bounds within it. The second contribution is conceptual, technically builds on the first contribution, and is a barrier in the derandomization of randomized logarithmic space (BPL).

(1) The query-with-sketch model generalizes the query complexity model for computing multi-bit functions $f : \{0,1\}^N \to \{0,1\}^M$. In this model, computation unfolds in two phases. Initially, the algorithm sends an agent to evaluate an arbitrary but length-restricted sketch of the input. Subsequently, the algorithm proceeds with queries. The main technical contribution is a lower bound in this model for the Approximate Matrix Powering (AMP) problem. To that end, we introduce a constrained form of conditional min-entropy that characterizes the number of queries in the model. We bound this entropy by developing tools that blend geometry, a generalization of tools from Lipschitz analysis for polynomials and low-distortion spaces, and probability theory. The main result is that AMP requires polynomial query complexity or super-polylogarithmic sketch size. We note that AMP and the query-with-sketch model are natural and interesting in their own right, in addition to the following conceptual contribution.

(2) Derandomizing BPL is an open question in computational complexity. The most successful derandomization algorithms of BPL make recursive use of pseudorandom generators or similar pseudorandom objects. The best-known derandomization places BPL inside $\mathsf{DSPACE}(\frac{\log^{3/2} n}{\sqrt{\log\log n}})$. We ask whether these algorithms can be substantially improved if we keep fixed the pseudorandom object and the main idea, which is to approximate $\mathbf{M}^n$ of the computation matrix $\mathbf{M}$ by relying on intermediate powers such as $\mathbf{M}^{n/2}$. We answer this question in the negative in a recursive generalization of the query-with-sketch model. We show that in this recursive and space-bounded model AMP needs super-polynomially many queries (time) or super-polylogarithmic sketch size (space). Specifically, in our model an algorithm that uses approximations of elements of the matrix $\mathbf{M}^{n/2}$ to approximate an element of $\mathbf{M}^n$, it must first determine the values of these elements in $\mathbf{M}^{n/2}$ and it can do this recursively. Other than this restriction, an algorithm is free to determine arbitrarily how to organize the recursion and how to use the bounded space. The conceptual takeaway is the "intermediate powers barrier", which indicates that fully derandomizing BPL cannot rely on a natural, recursive use of approximated intermediate powers.

ISSN 1433-8092

Contents

1 Introduction **3**
- 1.1 The query-with-sketch model . . . 3
  - 1.1.1 Query-with-sketch and related models . . . 4
  - 1.1.2 Examples of algorithms in the query-with-sketch . . . 5
  - 1.1.3 The limitations of the query-with-sketch model . . . 5
- 1.2 The recursive model and intermediate powers . . . 7
  - 1.2.1 The space-bounded recursive model . . . 8
  - 1.2.2 Barriers in complexity theory and space-bounded derandomization . . . 8
  - 1.2.3 The limitations of the intermediate powers method . . . 9

2 Outline of the argument and high-level presentation of our tools (lower bounds – main result) **10**
- 2.1 Computing $\mathbf{M}^2$ from $\mathbf{M}$ is hard in the query model (without sketch) . . . 11
- 2.2 Computing $\mathbf{M}^n$ from $\mathbf{M}^{n/2}$ is hard in the query model (without sketch): Bound the probability by a volume-embedding argument . . . 11
- 2.3 Computing any one element of $\mathbf{M}^2$ from $\mathbf{M}$ is hard with short sketch: min-entropy theorem and perturbation analysis . . . 12
- 2.4 Computing any polylog elements in $\mathbf{M}^n$ via $\mathbf{M}^{n/2}$ is hard with short sketch: putting everything together . . . 13
- 2.5 Approximate matrix powering in the recursive model . . . 14

3 Warm up: algorithms in query-with-sketch **15**

4 The min-entropy theorem **17**

5 Definition and examples of the recursive model **19**

6 Casting the Saks-Zhou's framework in the recursive model **21**

7 Main Results **23**
- 7.1 Formalization and input distribution . . . 23
- 7.2 Main results . . . 23

8 Reducing the min-entropy bound by the analysis of perturbations **26**
- 8.1 Roadmap to the argument . . . 26
- 8.2 Partial derivatives bound: from $\mathbf{M}$ to $\mathbf{M}^k$ . . . 27
- 8.3 Partial derivatives bound: from $\mathbf{M}^k$ to $\mathbf{M}^{k'}$ . . . 32
- 8.4 Perturbation analysis . . . 35

9 Conditional probability bounds from volume and embeddings **40**
- 9.1 Geometric decomposition . . . 41
- 9.2 Volume argument . . . 43
- 9.3 Proof of the main theorem . . . 48

Acknowledgements **50**

1 Introduction

It is believed that it is possible to turn every terminating probabilistic logarithmic space algorithm into a deterministic one by paying only a constant factor in the algorithm's space. Despite the significant research activity, this question has no general answer. Understanding one aspect of the limitations of the known techniques for the derandomization of randomized logarithmic space (BPL) is the motivation of this work.

Our contribution is twofold. First, we introduce the query-with-sketch model and develop tools for proving lower bounds. The lower bounds are for the approximate matrix powering (AMP) problem, which is ultimately related to the derandomization of BPL. Both the query-with-sketch and AMP are interesting in their own right, and the developed techniques might also be of independent interest. By generalizing the query-with-sketch model, we get to our second contribution. This contribution is the conceptual message of this work. It regards the limits of one current approach in derandomizing BPL. Importantly, we note that almost all known derandomizations of BPL fall within this barrier.

Here is a brief description of the barrier. After the seminal work[1] of Nisan [Nis92], which derandomized RL in polynomial time and $O(\log^2 n)$ space, the followup works studied the use of pseudorandom generators in a recursive framework. Our barrier is a super-polynomial query lower bound for a class of space-bounded recursive algorithms solving AMP. These algorithms have $\omega(1)$ recursion levels, with each level approximating elements of an intermediate power where the exponents are divisors of one another (e.g., $\mathbf{M}^n, \mathbf{M}^{n/3}, \mathbf{M}^{n/12}, \ldots, \mathbf{M}$). Each level can only query the approximated elements of the immediately lower level. A query to one or more elements whose values are not determined yet, triggers a recursive call. Furthermore, each level is given a local space of size polylog(n), which it can use arbitrarily. Syntactically, each level of the recursion is associated with a distinct query-with-sketch model with the added feature that the algorithm can update its space (sketch) during its computation. These algorithms are allowed to organize the recursive calls arbitrarily and also make arbitrary use of the bounded space.

The remainder of the Introduction is structured as follows. First, we define the query-with-sketch model and present our main results. Next, we discuss the barrier in the derandomization of BPL.

1.1 The query-with-sketch model

An algorithm in the query-with-sketch model aims to compute a multi-bit function $f : \{0,1\}^N \to \{0,1\}^M$ with the help of a short piece of information that we call "sketch". The sketch size should be smaller than N and M for the model to make sense. Instead of a multi-bit function, more generally, a problem $f : \mathcal{X}^n \to \mathcal{Y}$ is a function we want to compute, where $\mathcal{X}, \mathcal{Y}$ are finite sets. An algorithm in the query-with-sketch consists of the agent and the query algorithm. These two are computationally unbounded. The resources we care to bound are: (i) the sketch size that we think of as space, and (ii) the number of queries that we think of as time.

[1] Nisan's work uses intermediate powers of matrices but in a non-recursive way. This is because the derandomization algorithm stores in its memory enough many hash functions that can be efficiently used to reconstruct an approximation of the entire intermediate power (see pp.621–622 [Nis92]). In this way there is no need for recursion. Recursion is used when instead of storing a succinct representation of the intermediate matrix power you calculate it on-demand.

Informal definition. An algorithm in this model works as follows: first, send the agent S to the input $\mathbf{x}$ and get the sketch. Then, proceed by making queries until you specify the output. A more detailed definition and some notation follows.

Definition 1 (Query-with-sketch: algorithms and resources)**.** *Let $f : \mathcal{X}^n \to \mathcal{Y}$ be a problem.*

[Algorithm] *A* query-with-sketch algorithm Π^S *is defined through the functions S, Π as follows. Fix any input $\mathbf{x} \in \mathcal{X}^n$. The function $S : \mathcal{X}^n \to \{0,1\}^s$ is the sketch of the input, which outputs a string $S(\mathbf{x})$ of fixed length s. The function Π gets as input a sequence of length $\ell + 1$: $C = \langle S(\mathbf{x}), (q_1, a_1), (q_2, a_2), \ldots, (q_\ell, a_\ell)\rangle$, and Π is defined for every $\ell \in \mathbb{Z}^+$. A query $q_i \in \{1, \ldots, n\}$ is an input location, whereas the answer to the query $a_i = \mathbf{x}_{q_i}$. On a sequence C, Π outputs the next query $\Pi(C) = q_{\ell+1} \in \{1, \ldots, n\}$ or it produces an output from $\mathcal{Y}$ and halts.*

[Computation] *The computation of an algorithm Π^S on input $\mathbf{x}$ is a sequence of Cs: $C_0 \vdash \cdots \vdash C_m$, where the transition from C_i to C_{i+1} is consistent with Π; $C_0 = \langle S(\mathbf{x})\rangle$ and $C_m \in \mathcal{Y}$.*

[Complexity] *We define two types of query complexity:* $\mathrm{Cost}(\Pi^S)$ *the worst-case cost of Π^S, and* $\text{Avg-Cost}_{\mu_{\mathcal{X}^n}}(\Pi^S)$ *the average-case cost of Π^S for an input distribution $\mu_{\mathcal{X}^n}$ on $\mathcal{X}^n$. Let $\mathsf{T}_{\Pi^S}(\mathbf{x})$ be the number of queries of the algorithm Π^S on input $\mathbf{x}$.*

$$\mathrm{Cost}(\Pi^S) \stackrel{\text{def}}{=} \max_{\mathbf{x}\in\mathcal{X}^n} \mathsf{T}_{\Pi^S}(\mathbf{x}) \quad \text{and} \quad \text{Avg-Cost}_{\mu_{\mathcal{X}^n}}(\Pi^S) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{x}\leftarrow\mu_{\mathcal{X}^n}}\left[\mathsf{T}_{\Pi^S}(\mathbf{x})\right]$$

1.1.1 Query-with-sketch and related models

Before we explain what the model can and, importantly, cannot do, we will locate the query-with-sketch on the map of information-theoretic models.

In query and communication complexity [Yao79, KN97, RY20] the goal is to determine the output without knowing the whole $\mathbf{x}$. In such models "computation" corresponds to "information" revealed to us in a specific order. In contrast, in space-bounded models such as space-bounded Turing Machines, branching programs [Sha38, Sha49, Weg00], and streaming algorithms [AMS96, M$^+$05], the computational device gets to see the entire input, but during its computation can only remember a limited number of things. In the query-with-sketch model, in some sense, we do both. We get to see a sketch, a compression of the entire input, and our goal is to determine the output of the function by querying the input.

In the most common, 2-party communication complexity model each player computes by typically knowing half of the input. The space-bounded versions of the communication complexity model [BCP$^+$13, PSS14] also fundamentally differ from the query-with-sketch since in that case the space-bounded players maintain non-uniform space, but they do not get to see the entire input. The communication model studied in [DVM14] is loosely reminiscent of ours. However, it is used for the computationally bounded case and has a completely different application.

The models that are more closely related to the query-with-sketch are the decision tree model with help bits and the succinct cell-probe model. The most related one is the decision tree model with help bits [NRS98, BH98], which can be seen as a special case of our model. In that model, algorithms compute the direct sum of k instances of a boolean function using $k-1$ help bits, which can be viewed as a sketch of length $k-1$. The succinct cell probe model [PV10] also uses sketches. However, it is important to note that this model fundamentally differs from our query-with-sketch model since the cell probe model, being a data structure model, primarily focuses on managing the memory of inputs rather than computing a single function.

1.1.2 Examples of algorithms in the query-with-sketch

The query-with-sketch generalizes the query model [BDW02]. Without the sketch, Definition 1 above becomes the standard query model. This new model is strictly more powerful than the query model, as shown in the examples below.

Two examples. We present two problems and algorithms for these problems that make polynomially fewer queries than any algorithm in the standard query model. Both algorithms work by first approximating the output and then using the sketch to correct the (additive) error.

First, consider the Hamming Weight Problem where the input is a string from $\{0,1\}^n$ and the output is the number of 1s. In the query model, n queries are needed to compute, with probability more than $\frac{1}{2}$, the exact answer to this problem. But, in the query-with-sketch we can do the following: first sample n^ε-many bits to approximate the higher-order digits of the hamming weight and use a small sketch to get the lower-order digits. The intuition is that the lower-order (least significant) digits are harder to compute than the higher-order ones.

Our second example is the problem of Counting the Connected Components of an undirected graph in the adjacency list model. Our query-with-sketch algorithm simulates the sublinear approximation algorithm [CRT05] to approximate the answer with a subquadratic number of queries and then use the sketch to correct the approximation error.

The above toy-examples regard what the model can do.[2] For completeness, the easy details of these algorithms are given in Section 3. Algorithms in the query-with-sketch is a digression from the main topic of this work, which regards lower bounds.

1.1.3 The limitations of the query-with-sketch model

The main technical result for the query-with-sketch model is Theorem 4 (see below). This is a lower bound not just for the Approximate Matrix Powering (AMP) problem, but also for a generalization of the AMP. The reason we prove the lower bound for the Generalized AMP is because this is the version we need when we later on extend our results in the recursive model.

Let us now define the AMP problem and explain its relationship to our model. The AMP problem is determined by a rounding constant $\alpha > 0$. For a substochastic matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ – i.e., a non-negative matrix whose rows sum up to at most 1 – we denote by $[\mathbf{M}]_\alpha$ the matrix whose entries are the entries of $\mathbf{M}$ rounded to the closest multiple of $\frac{1}{n^\alpha}$.

Problem AMP_α
Input: the input is a rounded substochastic matrix $[\mathbf{M}]_\alpha$.
Output: the output is the rounded $[\mathbf{M}^n]_\alpha$ (or any $[(\mathbf{M}')^n]_\alpha$ where $[\mathbf{M}']_\alpha = [\mathbf{M}]_\alpha$)[3].

[2] To be precise, in this discussion we compared the randomized versions of the query-with-sketch and standard query algorithms.

[3] This is the appropriate definition of the problem. Two remarks are in order. First, why the definition does not allow an arbitrary matrix $\mathbf{A}$ which is close (in normed sense) to $\mathbf{M}^n$? This is because one can use the lower-order bits of the precision of the entries of $\mathbf{A}$ to encode arbitrary information which would have made any non-trivial lower bound in the recursive model (see below) impossible. In other words, the rounding of the entries is necessary in the "correct" definition of the problem. Second, the definition is more restricted (but cleaner) than the problem for which we have proved the lower bound. If instead of $[\mathbf{M}^n]_\alpha$ we allow to output $[\mathbf{A}]_\alpha$ for an arbitrary matrix $\mathbf{A}$ which is close to $[\mathbf{M}^n]_\alpha$, then this $[\mathbf{A}]_\alpha$ is still among the valid outputs. This is because every $[\mathbf{A}]_\alpha$ has preimages $\mathbf{M}'$ whose n-th power that is arbitrarily close to $\mathbf{A}$ and thus $[\mathbf{M}'^n]_\alpha = [\mathbf{A}]_\alpha$ (see Lemma 15 and 19, pp. 44 and 47).

Algorithms in this work are given query access to elements of rounded matrices.

As mentioned for the application to the recursive model we need a more general AMP. The more general problem besides α is determined by two parameters k and k', where k' is a multiple of k. In $\mathsf{Generalized-AMP}$ the query-with-sketch algorithm is given query access of $[\mathbf{M}^k]_\alpha$ and the goal is to output $[\mathbf{M}^{k'}]_\alpha$.

Problem $\mathsf{Generalized-AMP}_{\alpha,k,k',\textbf{where } k|k'}$
Input: the input is a rounded substochastic matrix $[\mathbf{M}^k]_\alpha$, for some substochastic matrix $\mathbf{M}$.
Output: the output is the rounded $[\mathbf{M}^{k'}]_\alpha$ (or any $[(\mathbf{M}')^{k'}]_\alpha$ where $[(\mathbf{M}')^k]_\alpha = [\mathbf{M}^k]_\alpha$).

Informally, the problem is given an $n \times n$ matrix $\mathbf{M}$ (or some power of $\mathbf{M}$) we would like to compute a power of this matrix, e.g., $\mathbf{M}^2$ or $\mathbf{M}^4$ or $\mathbf{M}^{n/2}$ or $\mathbf{M}^n$. $\mathsf{Generalized-AMP}$ is a family of problems; one problem for each choice of α, k, k'. For example, one problem is $\mathsf{Generalized-AMP}_{3,1,2}$, where for $n = 10$ the problem is: given a 10×10 matrix $\mathbf{M}$ each element of which is rounded to the third decimal digit ($\frac{1}{10^3}$) the goal is to compute $\mathbf{M}^2$ where the elements of the output are also rounded to the third decimal digit. As we will see, our lower bound holds simultaneously for all pairs of matrix powers $k, k' \leq n$.

Before we proceed we will clarify two things.

First, the reason that we denote the input as $\mathbf{M}^k$ (and not as $\mathbf{M}' = \mathbf{M}^k$) is to emphasize that we use one nemesis distribution for the whole family of the $\mathsf{Generalized-AMP}$ problems.

Second, due to rounding, for one input $[\mathbf{M}^k]_\alpha$ there are multiple valid outputs $[\mathbf{M}^{k'}]_\alpha$. In this work, *an algorithm is correct if it outputs any of the valid outputs.* This way the lower bound is stronger. Moreover, this is in line with the desired behavior of a successful derandomization of BPL.[4]

The main lower bound in the query-with-sketch model is informally stated as follows.

Theorem 4 (informally stated: lower bound for the query-with-sketch)**.** *Fix the matrix dimension $n \times n$. We can construct a distribution $\mathcal{M}$ over matrices that works for every k, k', where $1 \leq k < k' \leq n$ and k' is a multiple of k. This $\mathcal{M}$ is the nemesis distribution, which means that every algorithm Π^S with inputs from $\mathcal{M}$ has average query complexity $n^{\Omega(1)}$, when Π^S approximates a power $\mathbf{M}^{k'}$ given a sketch and query access to a smaller power $\mathbf{M}^k$. This is true for sketch size $s = \text{polylog}(n)$ and remains true even if Π^S approximates up to $N = \text{polylog}(n)$-many elements of $\mathbf{M}^{k'}$.*

Later on, when we state the above theorem rigorously, we will see that given s, N is a sufficiently large polylog.

For example, a corollary of the above informal theorem (after parameterizing it carefully) is that there is a distribution $\mathcal{M}$ over $n \times n$ substochastic matrices $\mathbf{M}$, where given query access to $[\mathbf{M}]_3$ every query-with-sketch algorithm that aims to compute any $\log^4 n$-many elements of $[\mathbf{M}^2]_3$ using any sketch of size $\log n$ must make at least $\sqrt{n}$ queries in average.

The above theorem says that computing the Generalized-AMP even for a small part of the output matrix is difficult. The theorem is restated on page 24, with the correct range of parameters and appropriate quantification. The formal statement and proof of this theorem is given in Sections 7, 8, and 9 (pp. 23 – 49). An outline of the argument is given in Section 2 (pp. 10 – 15).

[4]Derandomization algorithms compute a rounded output $\tilde{\mathbf{M}}^n$ that has small error $||\mathbf{M}^n - \tilde{\mathbf{M}}^n||_\infty$, where $||\cdot||_\infty$ denotes the entry-wise max norm. For derandomization, every such small-error rounded output would do.

To show this lower bound we develop new technical tools. At a high level the conceptual difficulty in proving the theorem lies in the following:

> *When proving a lower bound using a nemesis input distribution, the sketch – which is a function of the entire input – arbitrarily creates statistical dependencies in the analysis. It is not clear how existing techniques (prior to this work) can be used to prove such a lower bound.*

Here is a summary of the argument: First, we reduce the query complexity when given a sketch, to a min-entropy conditioned on certain boundary conditions. We refer to this as "restricted conditional min-entropy" or simply "min-entropy". This reduction is in Section 4 and is a general tool in the query-with-sketch model; i.e., not specific to the AMP problem. Here is an informal statement of the min-entropy theorem.[5]

Theorem 1 (Informal statement – Query-with-sketch to min-entropy reduction)**.** *Let $f : \mathcal{X}^n \to \mathcal{Y}$ be an arbitrary function, and $\mu_{\mathcal{X}^n}$ an arbitrary distribution over $\mathcal{X}^n$. Fix Π^S to be a correct algorithm for f with the help of an arbitrary sketch S, where $s = |S|$. High min-entropy of f implies a high number of queries in the query-with-sketch model.*

Now, the problem reduces to lower bounding this entropy conditioned on boundary conditions. To deal with those boundary conditions we study perturbations of the input matrix $\mathbf{M}$. We use this perturbation study when proving a lower bound in the number of queries to approximate say $\mathbf{M}^2$ with query access and sketch to $\mathbf{M}$. Now, if we want to approximate $\mathbf{M}^{k'}$ when we are given query access and sketch not to $\mathbf{M}$ but to a higher power $\mathbf{M}^k$ ($k < k'$), things are more complicated. This is because one element of $\mathbf{M}^k$ potentially reveals information of polynomial many elements of $\mathbf{M}$ due to powering the matrix ($\mathbf{M}^k$). To that end, we show that the distribution of $\mathbf{M}^k$ (which lives inside a metric space) can be embedded with low distortion into the distribution of $\mathbf{M}$. Thus, it suffices to analyze the events in the simpler $\mathbf{M}$-space. For a more detailed outline see Section 2.

1.2 The recursive model and intermediate powers

At a high level, we show that the complete derandomization of BPL ($\mathsf{BPL} = \mathsf{L}$) or even a polynomial time but space-restricted derandomization of $\mathsf{BPL} \subseteq \mathsf{TISP}(\mathrm{poly}(n), \mathrm{polylog}(n))$ cannot rely on approximating $\mathbf{M}^n$ of the transition matrix $\mathbf{M}$ by using intermediate powers, such as $\mathbf{M}^{n/2}$. We refer to this limitation as the *intermediate powers barrier* in the derandomization of BPL. The limitation is shown in an information-theoretic recursive model, where time and space are naturally formalized. In this model, algorithms are restricted to using previous powers to approximate higher matrix powers.

Below, we define the recursive model. Next, we compare our work to previous works and discuss barriers in complexity theory as well as prior works on derandomizing BPL. Finally, we state and explain the barrier result.

[5] Caution is needed. This is not a general "min-entropy" notion, but one where the conditionals are very specific (see Section 4).

1.2.1 The space-bounded recursive model

Previously, we defined the query-with-sketch model in its generality. Below, the space-bounded recursive model is defined only for AMP. This is because different problems have different recursive structures/subproblems.

Informal definition. To compute $[\mathbf{M}^n]_\alpha$ given $[\mathbf{M}]_\alpha$ as input, a *space-bounded recursive algorithm* recursively computes intermediate powers (e.g., $[\mathbf{M}^{n/2}]_\alpha, [\mathbf{M}^{n/4}]_\alpha, \dots$). Each recursion level is associated with a distinct query-with-sketch algorithm. The workspace (sketch) at each level is dynamically updated during the computation. To compute an element of $[\mathbf{M}^n[u,v]]_\alpha$, the algorithm issues a list of queries to elements of $[\mathbf{M}^{n/2}[u,v]]_\alpha$ until enough information is gathered in the workspace and becomes possible to determine the value of $[\mathbf{M}^n[u,v]]_\alpha$. Once this value is computed, the algorithm returns it to the recursion level that made the call and updates its space accordingly. We answer queries recursively. In this work, recursive algorithms are *partially adaptive*, meaning that the queries made at each recursion level are not dependent on what the algorithm has seen so far, but are instead based on an index h, which is a polylog(n)-long string computed from the input.[6] One can view the recursive algorithm as a family of non-adaptive algorithms, where each algorithm is indexed by h, a precomputed string from the input.

Formal definition. To formally define the recursive model we must specify various details. For the informal statement of the barrier Theorem 3 these details are not necessary. We defer the definition to Section 5, pp. 19.

In the remainder of this section, we first discuss the existing barriers in complexity theory to contextualize our proposed barrier. Next, we review the state-of-the-art in the derandomization of BPL. Finally, we link these to the intermediate powers barrier and state the key barrier theorem.

1.2.2 Barriers in complexity theory and space-bounded derandomization

The question of whether randomized logarithmic space or BPL, can be completely derandomized has remained open for more than four decades. In this work, we propose that a common algorithmic approach toward the derandomization of BPL cannot succeed within a structured model of computation. Meta-results of this nature are generally known as *barriers*, and we refer to our specific barrier as *the intermediate powers barrier*. Before explaining what this is we will briefly discuss barriers in complexity theory.

Barriers in computational complexity. Almost all of the major open questions in computational complexity remain unresolved. For example, we do not know if $\mathsf{P} \neq \mathsf{NP}$, or even the uniform $\mathsf{TC}^0 \neq \mathsf{NP}$. We also do not know if $\mathsf{BPP} = \mathsf{P}$ or if $\mathsf{BPL} = \mathsf{L}$, among many other important questions that are widely open. Thus, understanding the limitations of existing techniques toward answering a conjecture has become a common theme in computational complexity. These types of results are termed as "barriers". For example, when it comes to the open questions regarding class separations, most notably $\mathsf{P} \neq \mathsf{NP}$, three main barriers are known: relativization [BGS75], natural proofs [RR94], and algebrization [AW09]. Results that relate the difficulty of the derandomization of probabilistic polynomial time BPP with proving exponential circuit lower bounds [KI03] can also be

[6] String h corresponds to the "offline seed" in [SZ99].

thought of as barriers. Similarly, for improving slightly existing circuit lower bounds, e.g., [CT19]. Many other interesting results fall within the same barrier category, e.g., [LP21, Hir22, CJSW24] Most barriers are about time/circuit size lower bounds. Regarding space-bounded derandomization, back in the 1980s it was not known if the standard randomized algorithm for undirected st-connectivity [AKL$^+$79] can be derandomized. In [CR80, BBR$^+$96] it was shown that certain structured algorithms that test connectivity by placing pebbles on the given graph fail to derandomize the randomized st-connectivity algorithm. Years later, [Rei08] gives a logarithmic space algorithm that tests connectivity without moving pebbles on the original graph – thus, bypassing the [CR80, BBR$^+$96] barrier. How about the derandomization of the entire BPL?

Progress in $\mathsf{BPL} = \mathsf{L}$. Despite three decades of research [Sak96], progress in derandomization of space-bounded computation has been greatly slowed down. In the 1990s two seminal works shaped much of the field. The first is a pseudorandom generator [Nis90, Nis92], which uses $O(\log^2 n)$ space and can be computed in polynomial time. The second is a recursive algorithm [SZ99, CCvM06] that using this pseudorandom generator, or any other that uses $O(\log^{1+c} n)$ space, $c \leq 1$, can shrink the space to $O(\log^{1+c'} n)$ for $c' < c$ at the expense of the running time which becomes $n^{\omega(1)}$. As mentioned before, besides a clever deterministic algorithm [Rei08] for undirected st-connectivity, there is no major advance in the derandomization of the entire BPL in the sense that there is no known improvement on the constant c, which remains $3/2$, i.e., essentially $\mathsf{BPL} \subseteq \mathsf{DPSACE}(\log^{3/2} n)$. The best known derandomization is through a more recent work [HZ20] showing that $\mathsf{BPL} \subseteq \mathsf{DSPACE}(\frac{\log^{3/2} n}{\sqrt{\log \log n}})$; i.e., $\log^{3/2} n$ can be improved. This latter result still relies on the recursive structure of [SZ99]. In the recent years, there is massive progress in derandomizing branching programs [CDSTS23, PP23], hardness vs. randomness results on BPL [PRZ23, DPT24], L-AC^1 derandomization of BPL [CW24] and catalytic time-space $\mathsf{CTISP}(n, \log n, \log^2 n)$ derandomization of BPL [Pyn24], which are also of great interest.

In space-bounded derandomization [Nis90, Nis92, BNS92, INW94, SZ99, CCvM06, RTV06, GR14, BCG19, CL20, AKM$^+$20, HZ20, PV21, Hoz21, CDSTS23, PP23, PRZ23, CW24, Pyn24, DPT24], and in particular in the derandomization of logarithmic space, the way that most known derandomization algorithms proceed is by approximating matrix powering for stochastic matrices in small space or simultaneously small space and time. Many of these works rely on approximating higher powers of the computation matrix by first computing intermediate powers [SZ99, CCvM06, BCG19, CL20, HZ20, PV21, Hoz21, CDSTS23, PP23, Pyn24].

1.2.3 The limitations of the intermediate powers method

We give a super-polynomial query lower bound for AMP in the space-bounded recursive model, which is a natural recursive model for computing AMP and generalizes the Saks-Zhou framework [SZ99]. Algorithms in the recursive model approximate the matrix power $\mathbf{M}^n$ by recursively approximating its intermediate powers first (e.g., $\mathbf{M}^{n/2}$). We require $\omega(1)$ recursion levels and $O(\text{polylog}(n))$ space for each recursion level. We also require the algorithms to be partially adaptive. For a formal definition and examples of the recursive model, please refer to Section 5.

In the previous subsection, we informally stated a theorem showing that the generalized AMP problem is hard for the query-with-sketch model. By carefully using this theorem we can apply it to each recursion level of the recursive model, and obtain our barrier below.

Theorem 3 (informal statement of the barrier to derandomization of BPL)**.** *Fix the matrix dimen-*

sion n. We can construct a nemesis distribution $\mathcal{M}$ over substochastic matrices, such that every algorithm Π_h^* in the space-bounded recursive model correctly computes the matrix power $[\mathbf{M}^n]_\alpha$ given inputs from $\mathcal{M}$ either requires super-polylogarithmic space complexity, or super-polynomial average-case query complexity.

In Section 6 we show how to express the Saks-Zhou algorithm together with all of its variants and extensions in our recursive model. We do this because we wish to explicitly state which existing techniques our work rules out.

Most of the work following up [SZ99] falls into a recursive framework, where computing the n-th power of a matrix reduces to computing intermediate powers. In the seminal works [Nis90, Nis92], Nisan proposed a pseudorandom generator whose random seed has $O(\log^2 n)$ bits. By simulating the Turing Machine on $\text{poly}(n)$ many pseudorandom numbers generated from the random seed, Nisan showed that $\mathsf{BPL} \subseteq \mathsf{SC}$. Saks and Zhou [SZ99] significantly improved the space from $O(\log^2 n)$ to $O(\log^{1.5} n)$ by reusing Nisan's PRG. Specifically, they *approximate* $\mathbf{M}^n$ by recursively approximating intermediate powers $\mathbf{M}^{i\cdot\sqrt{\log n}}$ for each $i \in \{1, \ldots, \sqrt{\log n}\}$, where all the recursion levels share the same random seed of length $O(\log^{1.5} n)$. [CCvM06] further generalized [SZ99], and gave a smooth time-space trade-off to the derandomization of BPL. More recently and perhaps unexpectedly, [Hoz21] improved the bound to $\mathsf{BPL} \subseteq \mathsf{DSPACE}(\log^{1.5} n/\sqrt{\log\log n})$ by plugging in a better WPRG (weighted pseudorandom generator) into the Saks-Zhou framework. As mentioned our barrier holds for all those works. We defer to Section 6 a more detailed discussion and a formalization of the Saks-Zhou framework. As a corollary to the intermediate powers barrier, a lower bound to previous works is also included and stated at Theorem 2.

2 Outline of the argument and high-level presentation of our tools (lower bounds – main result)

In this section we give an outline of our main results. This work introduces the following three lower-bounding tools, which might also be of independent interest.

(I) Reduction to min-entropy.

(II) Perturbation analysis.

(III) Conditional probability bounds from volume and embeddings.

The purpose of this outline is to identify the exact points in the argument where each of the aforementioned tools is applied. The presentation is high-level, and the arguments are heuristic[7]. These heuristic arguments are rigorously formalized in Sections 3 through 9. Below we list in which sub-section each of the (I), (II), or (III) is used.

- Section 2.1: From elementary principles (no new tools).
- Section 2.2: (III).
- Section 2.3: (I) and (II).
- Section 2.4: (I), (II), and (III).
- Section 2.5: From elementary principles (application to the recursive model).

[7] "Heuristic argument" means that we make simplifying assumptions to advance the exposition.

Before the proof: dealing with a multi-valued function

The definition of AMP due to rounding permits two distinct matrices $\mathbf{M}_1 \neq \mathbf{M}_2$ to be such that $[\mathbf{M}_1]_\alpha = [\mathbf{M}_2]_\alpha$ and $[\mathbf{M}_1^n]_\alpha \neq [\mathbf{M}_2^n]_\alpha$. In our exposition in this section, we shall assume that this does not happen. We take care of this detail in Section 8.

2.1 Computing $\mathbf{M}^2$ from $\mathbf{M}$ is hard in the query model (without sketch)

When there is no sketch, $\mathbf{M}^2$ depends on all the elements of the substochastic matrix $\mathbf{M}$. The same thing holds for approximating $\mathbf{M}^2$. To see this, let $\mathbf{M}$ be everywhere $\frac{1}{n}$ and $\mathbf{M}'$ be the same as $\mathbf{M}$ but with one element equal to zero. Then, $\mathbf{M}^2 = \mathbf{M}$ but in $(\mathbf{M}')^2$ the minimum element is at most $\frac{n-1}{n^2}$. This means, that $||\mathbf{M}^2 - (\mathbf{M}')^2||_\infty \geq \frac{1}{n^2}$. Hence, every algorithm which approximates the square in the output with error smaller than $1/(2n^2)$ must query all the elements of the input matrix.

2.2 Computing $\mathbf{M}^n$ from $\mathbf{M}^{n/2}$ is hard in the query model (without sketch): Bound the probability by a volume-embedding argument

The situation becomes more complex when the input elements are dependent. For instance, consider a distribution $\mathcal{M}$ from which we sample $\mathbf{M}$. If the input to our problem is not $\mathbf{M}$, but rather $\mathbf{M}^{n/2}$, the elements of $\mathbf{M}^{n/2}$ may exhibit statistical dependencies. To address this, we need to design $\mathcal{M}$ such that knowing a few values of $[\mathbf{M}^{n/2}]_\alpha$ does not allow recovery of $[\mathbf{M}]_\alpha$ (since in that case, we can compute $[\mathbf{M}^n]_\alpha$).

We will construct a nemesis distribution[8] $\mathcal{M}$ of the form $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}(\frac{1}{n}\mathbf{J} - \mathbf{\Delta})$, which is the weighted sum of the identity matrix $\mathbf{I}$, the all-one matrix $\mathbf{J}$ and a relatively small uniform random matrix $\mathbf{\Delta}$ (i.e., $\mathbf{\Delta}$ is the only source of randomness) whose elements are i.i.d. uniformly selected from a small range. Specifically, $\mathbf{\Delta}$ is a discrete random matrix whose elements are uniformly distributed in $[0, \frac{1}{n^\gamma}] \cap \{d \cdot \frac{1}{n^\tau}\}_{d \geq 0}$, where $2 \leq \gamma < \alpha - 12.2$ and τ large enough are constants.[9] Then, $\mathbf{M}^k$ for each $k \geq 1$ can be rewritten as $\sum_{t=0}^{k} \binom{k}{t} (\frac{n^2-1}{n^2})^{k-t} n^{-3t} (\frac{1}{n}\mathbf{J} - \mathbf{\Delta})^t$, which is dominated by the first few terms of the sum; i.e., the terms for small t.

The main point of our "*volume and embeddings*" technique (bullet number (III) above) is that the mapping $\mathbf{M} \mapsto \mathbf{M}^k$ is a *low distortion* embedding when the probability space is viewed as a metric space (and now probability corresponds to volume). This is because $\mathbf{M}^k$ is dominated by $(\frac{n^2-1}{n^2})^k \mathbf{I} + k(\frac{n^2-1}{n^2})^{k-1} \frac{1}{n^3}(\frac{1}{n}\mathbf{J} - \mathbf{\Delta})$, which is linear in $\mathbf{\Delta}$ and hence also linear in $\mathbf{M}$. Specifically, using Lipschitz analysis we can show that (i) if we regard $\phi_k(\frac{1}{n}\mathbf{J} - \mathbf{\Delta}) \stackrel{\text{def}}{=} \mathbf{M}^k = (\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}(\frac{1}{n}\mathbf{J} - \mathbf{\Delta}))^k$ as a function of $\frac{1}{n}\mathbf{J} - \mathbf{\Delta}$, then ϕ_k is an invertible function[10]. (ii) ϕ_k and ϕ_k^{-1} are Lipschitz functions with constants that are not very hard to bound. Hence, ϕ_k is a bi-Lipschitz function, i.e. a low distortion embedding. These two results are stated and proved as Lemma 14 and Lemma 15, pp. 43-44. Finally, observe that the (discrete) volumes under low-distortion embeddings are somewhat preserved. It follows that (almost) preserving discrete volumes also preserves statistical dependence/independence, which is the key observation.

Since the elements of $\mathbf{M}$ are independent and identically distributed and since $\mathbf{M}^{n/2}$ is a low-distortion embedding of $\mathbf{M}$, the statistical dependencies among the elements of $\mathbf{M}^{n/2}$ are relatively

[8] The complete definition of the distribution $\mathcal{M}$ is on page 23.

[9] The parameters γ, τ will not be used until the next subsection.

[10] This is not true in general, but is always true for $\mathbf{\Delta}$ in a sufficiently small range.

weak. By "weak" we mean that even if $0.99n^2$ elements of $[\mathbf{M}^{n/2}]_\alpha$ are fixed, there remains significant uncertainty about the remaining elements (cf. Theorem 6, p. 48). With this in mind, a lower-bound argument similar to that of the previous section applies here as well.

2.3 Computing any one element of $\mathbf{M}^2$ from $\mathbf{M}$ is hard with short sketch: min-entropy theorem and perturbation analysis

We will now incorporate the sketch into our model of computation and analysis. Consider the function whose input is $\mathbf{M}$ and the output is $\mathbf{M}^2$ both rounded to the nearest multiple of $\frac{1}{n^\alpha}$ where $\alpha > 14.2$ is a constant. Then, each element can be described with $\log(n^\alpha) = \lceil \alpha \log_2 n \rceil$ bits. We claim that computing an arbitrary element $\mathbf{M}^2[u, v]$ with the help of an arbitrary sketch S of length $|S| = \log n$ requires average query complexity $\Omega(n)$ over the same nemesis distribution $\mathcal{M}$ in the previous section.

Obviously, the sketch might reduce the number of queries but makes lower bounds harder to get. Our intuition about the lower bound is that when the sketch is short it cannot bring enough information to save queries for all possible inputs. To quantify this we introduce *restricted conditional min-entropy* or *min-entropy* for short, and relate it to the query-with-sketch model. According to our min-entropy theorem (cf. Theorem 1, p. 17), a lower bound in the number of queries of a query-with-sketch algorithm can be obtained by lower bounding this min-entropy notion. This way the analysis focuses on the entropy and is not concerned with queries and sketches.

Specifically, fix an integer $q \geq 0$, and fix a distribution μ over the input $\mathbf{M}$. Fix also an arbitrary event over the inputs $\sigma = ([\mathbf{M}[i_1, j_1]]_\alpha = x_1, ..., [\mathbf{M}[i_q, j_q]]_\alpha = x_q)$ for every $(i_1, j_1, x_1), ..., (i_q, j_q, x_q)$. The min-entropy of this one element $[\mathbf{M}^2[u, v]]_\alpha$ conditioned on σ is

$$H_{\min}([\mathbf{M}^2[u, v]]_\alpha | \mu, \sigma) \stackrel{\text{def}}{=} \min_{y \in \mathbb{R}} \log \left(\Pr_{\mathbf{M} \sim \mu} [[\mathbf{M}^2[u, v]]_\alpha = y | \sigma] \right)^{-1}$$

Note that this conditional min-entropy can be viewed as the entropy of the function $\mathbf{M} \mapsto \mathbf{M}^2[u, v]$, i.e., the entropy measured on the output where the randomness is in the input of this function. The min-entropy theorem states that, if $H_{\min}([\mathbf{M}^2[u, v]]_\alpha | \mu, \sigma) > 2|S|$ for every σ, then every query-with-sketch algorithm Π^S that computes $[\mathbf{M}^2[u, v]]_\alpha$ has $\text{Avg-Cost}_\mu(\Pi^S) > q/3$.

This theorem is not hard to prove, but we will not outline its proof here to continue with the flow of the outline.

We use the above theorem to establish an $\Omega(n)$ query lower bound as follows. Let $q = 2n - 2$ be the desired query lower bound. Note that $2n - 1$ are all the elements of the u-th row and v-th column. Given the nemesis distribution $\mathcal{M}$ as constructed in Section 2.2, the query lower bound is reduced to upper bound $\Pr_{\mathbf{M} \sim \mathcal{M}}[[\mathbf{M}^2[u, v]]_\alpha = y | \sigma]$ *for every* y and σ (equivalently to lower bound the conditional min-entropy of $[\mathbf{M}^2[u, v]]_\alpha$ *for every* y and σ). Now, we will bound this probability in three steps.

Step 1: First, observe that $[\mathbf{M}^2[u, v]]_\alpha$ is a constant function in each $[\mathbf{M}[i, j]]_\alpha$ when $i \neq u$ and $j \neq v$. There are $n^2 - 2n + 1$ many such indices (i, j). We want to show that, even if σ fixes $q = 2n - 2$ many input elements $[\mathbf{M}[i, j]]_\alpha$ for which $u = i$ or $v = j$, the element $[\mathbf{M}^2[u, v]]_\alpha$ still heavily depends on the one remaining input element (i', j'). Let us arbitrarily choose the remaining (i', j') where $i' = u$ or $j' = v$ and call this input element $\mathbf{M}[i', j']$ the *core element*.

Step 2: To upper bound the probability, it suffices to focus on the core element and disregard other elements of $\mathbf{M}$. Specifically, upper bounding the conditional probability $\Pr_{\mathbf{M} \sim \mathcal{M}}[[\mathbf{M}^2[u, v]]_\alpha = y | \sigma]$

can be reduced to upper bounding $\Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma']$, since $\Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma] \leq \max_{\sigma'} \Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma']$ where σ is a restriction to $2n-2$ elements, and σ' restricts n^2-1 elements (other than the core element) and is consistent to σ. This is because the previous probability is a weighted sum of the latter one.[11]. This probability is also analytically easier to handle.

Step 3: Recall that our goal is to bound the conditional probability for the value of $[\mathbf{M}^2[u,v]]_\alpha$, i.e., we want to show that this probability is not too large even when the only source of randomness is the core element $\mathbf{M}[i',j']$. In other words, we want to show that enough uncertainty from $\mathbf{M}[i',j']$ is preserved in $[\mathbf{M}^2[u,v]]_\alpha$. To that end, we observe that $\mathbf{M}^2[u,v]$ is an increasing function of $\mathbf{M}[i',j']$. Intuitively, this means that the randomness from $\mathbf{M}[i',j']$ makes it to the output $\mathbf{M}^2[u,v]$. The problem is that the output is $[\mathbf{M}^2[u,v]]_\alpha$ not $\mathbf{M}^2[u,v]$. This means that as a function of the input $\mathbf{M}[i',j']$ the output $[\mathbf{M}^2[u,v]]_\alpha$ is a step-wise increasing function, due to rounding. However, if we could show that the function $\mathbf{M}[i',j'] \mapsto [\mathbf{M}^2[u,v]]_\alpha$ increases not too slow, then this would imply that the number of discrete inputs that make the same (rounded) output is small. Given a uniform input distribution, which is also over equidistant points, the upper bound of the desired probability $\Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma']$ follows from carefully counting the distinct possible values for $[\mathbf{M}^2[u,v]]_\alpha$. Everything holds true when the input elements are also rounded, i.e., the function becomes $[\mathbf{M}[i',j']]_\alpha \mapsto [\mathbf{M}^2[u,v]]_\alpha$.

Let us give a little more details about how fast $\mathbf{M}^2[u,v]$ grows. Observe that the partial derivative $\frac{\partial \mathbf{M}^2[u,v]}{\partial \mathbf{M}[i',j']} = \Omega(\frac{1}{n^4})$, which means that the speed in which $\mathbf{M}^2[u,v]$ grows at least linearly. It turns out this to be sufficient for the purpose of upper bounding $\Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma']$.

To bound the probability and ignoring that the input elements $[\mathbf{M}[i,j]]_\alpha$ are also rounded, we argue as follows. To increase $\mathbf{M}^2[u,v]$ by $n^{-\alpha}$, one only need to increase $\mathbf{M}[i',j']$ by at most $n^{-\alpha}/\frac{\partial \mathbf{M}^2[u,v]}{\partial \mathbf{M}[i',j']} = O(n^{4-\alpha})$. Given that each element of $\mathbf{M}$ is taken uniformly and equidistantly from a range of length $1/n^{3+\gamma}$, the probability can be bounded

$$\Pr_{\mathbf{M}\sim\mathcal{M}}[[\mathbf{M}^2[u,v]]_\alpha = y|\sigma'] \leq \frac{O(n^{4-\alpha})}{n^{-3-\gamma}} = O(n^{7-\alpha+\gamma})$$

Recall that $\alpha > 14.2$ and $2 \leq \gamma < \alpha - 12.2$, the probability is bounded by $O(n^{-5.2}) \ll 2^{-2|S|} = n^{-2}$. Therefore, the restricted conditional min-entropy is at least $2|S|$. By the theorem connecting this entropy and the query lower bound in the query-with-sketch model, we get an $\Omega(n)$ query lower bound.

The *Perturbation Analysis* method (Section 8) generalizes the above high-level argument in two ways: (i) computing elements of $[\mathbf{M}^{k'}]_\alpha$ as a function of $[\mathbf{M}^k]_\alpha$ for every $k|k'$, and (ii) computing polylog(n) elements of $[\mathbf{M}^{k'}]_\alpha$, instead of a single element. This generalization is crucial for establishing our main result, which is outlined in the next section.

2.4 Computing any polylog elements in $\mathbf{M}^n$ via $\mathbf{M}^{n/2}$ is hard with short sketch: putting everything together

We are now ready to give an outline of the argument of our main query-with-sketch result. Here, we are given a sketch of length $|S| = \text{polylog}(n)$, and as many as $N = (\frac{|S|}{\log n})^2$ indices of $\mathbf{M}^n$ that

[11] For a discrete probability space and two events A, B, $\Pr[A|B] = \sum_{B'} \Pr[A|B'\cap B]\cdot\Pr[B'|B]$, where the summation runs over a partition of B space; and thus, $\Pr[A|B] \leq \Pr[A|B']$, for some $B' \subseteq B$.

we wish to approximate their values. We will show that computing any set of N elements of $[\mathbf{M}^n]_\alpha$ with the help of sketch S, and given $[\mathbf{M}^{n/2}]_\alpha$ as input, requires $\Omega(\sqrt{n})$ queries on the average when the underlying distribution is the same $\mathcal{M}$ as before; see Theorem 4 (page 24) for the formal statement. This result is non-trivial as the sketch length can be as long as polylog(n) (not too large polylogarithmic, but still polylogarithmic).

We still use the min-entropy theorem to reduce the query complexity with sketch to the min entropy of the output conditioned on partial restrictions of the input elements. Formally, let $\text{out} = ((u_1, v_1), \ldots, (u_N, v_N))$ be the N indices of $[\mathbf{M}^n]_\alpha$ we want to evaluate. To show an $\Omega(\sqrt{n})$ query lower bound, it suffices to show that

$$\Pr_{\mathbf{M}\sim\mathcal{M}}[[(\mathbf{M}^n)_{\text{out}}]_\alpha = \mathbf{y}|\sigma] < 2^{-2|S|}$$

for every output $\mathbf{y}$ and an arbitrary partial assignment σ to any $q = \sqrt{n}$ elements of $[\mathbf{M}^{n/2}]_\alpha$.

The technical problem with choosing more than one elements of $\mathbf{M}^n$ (in this case polylogarithmic many) is that we have to carefully "match" each of them with an element of $\mathbf{M}^{n/2}$. This is because, same as in Section 2.3, we want to transfer simultaneously for all elements entropy from the input (i.e., from $\mathbf{M}^{n/2}$) to the output (i.e., to $\mathbf{M}^n$). To that end, instead of looking to identify a single core element (as in Section 2.3) we are somehow able to show that polylogarithmically many core elements of $\mathbf{M}^{n/2}$ are "matched" to a same number of co-core elements of $\mathbf{M}^n$. The correct way to generalize Section 2.3 is through the notion of a diagonally dominant Jacobian matrix. The details are left for actual argument, but the point is that we can find such core and co-core elements.

Now, recall that in Section 2.3 the reason we found the core element was to use the restricted conditional min-entropy notion and its connection to the number of queries. Same here with the core and co-core elements. A key step in upper bounding the corresponding conditional probability is to express the event about the co-core elements of $[\mathbf{M}^n]_\alpha$ as an event about the elements of $[\mathbf{M}^{n/2}]_\alpha$. We do this by relying on the Jacobian. Now, the conditional event we wish to bound is expressed in a language where certain matrix elements are contained in a hypercube (or hyper-rectangle). This view of the event is rather useful, because we can further decompose the hyper-rectangle into smaller hypercubes and by the union bound the probabilistic upper bound reduces to upper bound a sum of probabilities. Finally, to upper bound each such probability we need to invoke a generalization of our "embeddings and volume" technique we discussed in Section 2.2. This summary does not provide any details. The details matter a lot in this argument, and are given in Sections 8 and 9.

We will use a generalization of this result to obtain the lower bound for the AMP problem in the recursive model that follows.

2.5 Approximate matrix powering in the recursive model

The main conceptual/meta-result result of our work is a query lower bound for AMP in the recursive model, which generalizes the Saks-Zhou framework. Algorithms in the recursive model approximate the matrix power $\mathbf{M}^n$ by recursively computing its intermediate powers first (e.g., $\mathbf{M}^{n/2}$). Each recursion level has $O(\text{polylog}(n))$ space. The model is partially adaptive in the sense instead of one recursive algorithm we have a family of quasi-polynomially many algorithms $\{\Pi^*_h\}_h$. Each Π^*_h is non-adaptive, but before the computation starts given the input a short string

$$h = h(\mathbf{M})$$

with $|h| = O(\text{polylog}(n))$ is determined. A formal definition of the model is in Section 5.

To explain how the recursive model works let us look at a natural recursive algorithm (conceptually simpler than Saks-Zhou). This algorithm recursively computes the power by realizing the following equation (for simplicity assume that n is a power of 2).

$$\mathbf{M}^k[u,v] = \sum_{i=1}^{n} \mathbf{M}^{k/2}[u,i] \cdot \mathbf{M}^{k/2}[i,v]$$

The algorithm runs in space $O(\log^2 n)$ space and $O((2n)^{\log n})$ time (for the same reason that Savitch's algorithm [Sav70] runs in this space and time) and solves the AMP problem. To compute each element of the matrix power $\mathbf{M}^k[u,v]$, we recursively query the elements of smaller powers $\mathbf{M}^{k/2}[u,i]$ and $\mathbf{M}^{k/2}[i,v]$, and sum up their products. Since there are only $\log n$ recursion levels, and in each levels of the recursion, we only maintain a summation variable which costs $O(\log n)$ space, the total space complexity is $O(\log^2 n)$. Note that the algorithm is non-adaptive; i.e., the order of queries made to the lower powers is predetermined.

In addition to the Saks-Zhou framework, several other known algorithms can be expressed in the recursive model. Saks-Zhou framework (Section 6) in comparison to the above algorithm, has an extra string h produced by the Nisan's pseudorandom generator (the precomputed hash function on the input) and has a smaller number of recursion levels.

Theorem 3 says that, to compute AMP in the recursive model with a piece of extra advice h of length $O(\text{polylog}(n))$, either super-polylogarithmic space or super-polynomial running time is necessary.

The proof goes by grouping successive $O(\text{polylog}(n))$ queries from the upper recursion level to the lower recursion level. By applying the previously obtained lower-bound for the query-with-sketch, we know that to answer every $\text{polylog}(n)$ queries one must make on the average $\Omega(\sqrt{n})$-many queries to its succeeding recursion level. Then, the lowest recursion level will receive $(\frac{\Omega(\sqrt{n})}{O(\text{polylog}(n))})^{\omega(1)} = n^{\omega(1)}$ queries in expectation.

3 Warm up: algorithms in query-with-sketch

The query-with-sketch model is a non-trivial model when the length of the sketch is shorter than the input and the output. To motivate our lower bounding studies, and to showcase how the query-with-sketch model works, we give non-trivial sublinear algorithms to two classic problems, the *Hamming weight problem* and the *counting connected components problem.*

Hamming weight problem. Given n input bits $X_1, ..., X_n \in \{0,1\}$. Output the Hamming weight of the n bits, i.e., the sum of the n bits $T = \sum_{i=1}^{n} X_i$.

Fact 1. *There exist constants $c < c' < 1$ and a probabilistic query-with-sketch algorithm with query complexity $O(n^{c'})$ and sketch length $c \log n$, such that for every distribution of the input μ over $\{0,1\}^n$, the algorithm computes the exact Hamming weight with high probability.*

Proof of Fact 1. Given any fixed input $X_1, ..., X_n$. We let their average $p := \frac{\sum_{i=1}^{n} X_i}{n}$. The Hamming weight is exactly $p \cdot n$. The standard estimation algorithm goes by estimating p using a sublinear number of queries. Our idea of computing $p \cdot n$ is simply to use the sketch to correct the error of the estimation.

Formally, the query-with-sketch algorithm runs as follows. We set the sketch S to be the last $\frac{2}{3}\log n$ bits of the Hamming weight pn, which is the hardest part to estimate. The query-with-sketch algorithm will uniformly and independently query $n^{3/4}$ bits from the input. Denote by $\tilde{p}$ the average of the $n^{3/4}$ bits. The algorithm will output the closest number to $\tilde{p} \cdot n$ such that the last $\frac{2}{3}\log n$ bits are exactly S.

Notice that each query evaluates to 1 with probability p. We will show by Chernoff bound that, $n^{3/4}$ number of queries suffice to estimate p within error $n^{-1/3}$ with high probability.

Let $\sigma = (\sigma_1, ..., \sigma_{n^{3/4}})$ denote the list of queries to the input. By Chernoff bound, we have

$$\Pr_{\sigma}[|n^{1/4} \cdot \sum_{i=1}^{n^{3/4}} X_{\sigma_i} - pn| \geq \varepsilon \cdot n^{1/4}] \leq 2 \cdot e^{-\frac{\varepsilon^2}{3n^{3/4}}}$$

for every possible input $(X_1, ..., X_n) \in \{0,1\}^n$ and every $\varepsilon > 0$. By letting $\varepsilon = \frac{1}{2}n^{5/12}$ and introducing the input distribution μ, we have

$$\Pr_{\sigma,(X_1,...,X_n)\sim\mu}[|n^{1/4} \cdot \sum_{i=1}^{n^{3/4}} X_{\sigma_i} - pn| \geq \frac{1}{2}n^{2/3}] \leq 2e^{-\frac{1}{12}n^{1/12}}$$

The inequality above says that if the queries are made uniformly at random, we can always approximate the Hamming weight within error $\frac{1}{2}n^{2/3}$ with high probability. By setting the sketch S to be the last $\frac{2}{3}\log n$ bits of $f(X_1, ..., X_n)$, we can recover the exact answer with high probability. The query complexity of the algorithm is $O(n^{3/4})$. □

The algorithm for the *counting connected components problem* goes in a similar way as above and relies on the following result.

Lemma 2 ([CRT05])**.** *Fix n the number of vertices, and $\varepsilon, \delta \in (0,1)$ functions of n. One can construct a probabilistic algorithm Π in the adjacency list query model with query complexity $O(\varepsilon^{-4} \cdot \ln(1/\delta))$ estimating the number of connected components of a graph. Specifically, fix an arbitrary input graph G, let $\#CC(G)$ denote the number of connected components of G, then we have*

$$\Pr_{r}(|\Pi(G, r) - \#CC(G)| \leq \varepsilon \cdot n) \geq 1 - \delta$$

where r is the randomness of the algorithm.

Fact 3. *For every constant $c \in (0,1)$, there exists a probabilistic query-with-sketch algorithm running in $O(n^{4-4c} \cdot \log n)$ queries computing the exact number of connected components in the adjacency list query model, with sketch length $c \cdot \log n$, where n is the number of vertices.*

When $c > 3/4$, we obtain a sublinear algorithm.

Proof of Fact 3. Set $\delta = 1/\mathrm{poly}(n)$ and $\varepsilon = \frac{n^{c-1}}{2}$. We simply run the algorithm in Lemma 2 which gives the first $\log n - \log(2\varepsilon \cdot n) = (1-c)\log n$ bits of the answer with high probability. Combined with the $c \log n$-bit long sketch we can obtain the correct answer with high probability. □

4 The min-entropy theorem

Here we give a formal statement of the min-entropy theorem, which is the key to reducing the query-sketch lower bound to lower-bounding the entropy of the output, which is amenable to analysis.

Definition 2 (restricted conditional min-entropy)**.** *Let $f : \mathcal{X}^n \to \mathcal{Y}$ be a function, $\mu_{\mathcal{X}^n}$ a distribution over $\mathcal{X}^n$, and two parameters: the hard instance set $\mathcal{E} \subseteq \mathcal{X}^n$ (optional) and $q \in \{0, ..., n\}$ many fixed elements of the input $\sigma = (X_{i_1} = x_{i_1}, ..., X_{i_q} = x_{i_q})$ (corresponding to q queries and their answers) for some $i_1, ..., i_q \in [n]$. The* restricted conditional min-entropy[12] *of f given $\mu_{\mathcal{X}^n}, \mathcal{E}, \sigma$ is*

$$H_{\min}(f|\mu_{\mathcal{X}^n}, \mathcal{E}, \sigma) \stackrel{\text{def}}{=} \min_{y \in \mathcal{Y}} \log \left(\Pr_{\mathbf{x} \leftarrow \mu_{\mathcal{X}^n}} [f(\mathbf{x}) = y \text{ and } \mathbf{x} \in \mathcal{E} | \sigma] \right)^{-1}$$

If the probability is zero then the $H_{\min}(f|\mu_{\mathcal{X}^n}, \mathcal{E}, \sigma)$ is defined to be ∞.

When it is clear from the context, we use the notation $H_{\min}(f)$ to denote the restricted conditional min-entropy.

We note that σ here denotes the condition that $X_{i_1} = x_{i_1}, ..., X_{i_q} = x_{i_q}$ holds true. The restricted conditional min-entropy does not involve queries and the sketch.

Our restricted conditional min-entropy is a special form of the min-entropy of the output of a function conditioned on a partial assignment to the input. This is the only notion of entropy we will use in this work. We also call it min-entropy without ambiguity.

Throughout the paper $q = \sqrt{n}$ and the min-entropy is supposed to be lower-bounded by polylog(n). Note that while $\mathcal{E} = \mathcal{X}^n$ and $q = 0$ the min-entropy collapses to the (unconditional) min-entropy of f. However, we observe that the min-entropy alone is not a good measure of query complexity. For example, the parity of n uniformly random bits requires linear query complexity but has a min-entropy of only 1. In contrast, if we show that the min-entropy of a function is always high conditioned on an arbitrary assignment to $\Omega(n)$ elements of the input, then a linear query lower bound follows.

The set $\mathcal{E}$ is important in our proof to restrict the set of inputs where bounding $\Pr[f(\mathbf{x}) = y|\sigma]$ is feasible. When optional, we set $\mathcal{E} = \mathcal{X}^n$. The role $\mathcal{E}$ plays will be clear in Section 9.

Intuitively, the observed entropy in the output is related to the sketch and queries as follows. Without any precomputed information the observed entropy on the output of f is the (unconditional) min-entropy. Given the precomputed sketch S, which is a function of $\mathbf{x}$, the observed entropy in the output $f(\mathbf{x})$ reduces, and every subsequent query makes the entropy smaller until it becomes zero. At this point, an algorithm can make a correct decision.

Theorem 1 (min-entropy theorem: main tool for lower-bounding the query-with-sketch model)**.** *Let $f : \mathcal{X}^n \to \mathcal{Y}$ be an arbitrary function, and $\mu_{\mathcal{X}^n}$ an arbitrary distribution over $\mathcal{X}^n$. Fix Π^S to be a correct algorithm for f with the help of an arbitrary sketch S of length $s \geq 3$, and fix a set $\mathcal{E} \subseteq \mathcal{X}^n$ (optional) with $\Pr_{\mu_{\mathcal{X}^n}}[\mathcal{E}] \geq 1/2$ (where $\Pr_{\mu_{\mathcal{X}^n}}[\mathcal{E}] = \Pr_{\mathbf{x} \leftarrow \mu_{\mathcal{X}^n}}[\mathbf{x} \in \mathcal{E}]$). If for every $q, i_1, i_2, ..., i_q \in [n]$ and every partial assignment $\sigma = (X_{i_1} = x_{i_1}, ..., X_{i_q} = x_{i_q}) \in \mathcal{X}^q$, we have*

$$H_{\min}(f) > 2s$$

[12] Min-entropy as a fundamental concept in information theory, has been extensively used in coding theory, cryptography, and various complexity-theoretic applications [HILL99, Vad19]. It is important to note that the notion of min-entropy in our work is defined in a form with a specific purpose and application and bears little resemblance to the traditional notion of min-entropy used in previous works.

then

$$\text{Avg-Cost}_{\mu_{\mathcal{X}^n}}(\Pi^S) > \frac{q}{3}$$

To clarify, we note that the min-entropy theorem does not require the algorithms to be non-adaptive. The non-adaptivity emerges at the reduction from the lower bound of the recursive model to the lower bound of the query-with-sketch model.

The intuition of the theorem is that, if for every fixed q queries we need an at-least-s-length sketch to figure out the answer, then conversely the query complexity is also $\Omega(q)$ for sketches of fixed length s. The proof goes by a counting argument that regards the query-with-sketch model as a family of decision trees, where each sketch corresponds to a decision tree.

Proof of Theorem 1. Assume for sake of contradiction that $\text{Avg-Cost}_{\mu_{\mathcal{X}^n}}(\Pi) \leq \frac{q}{3}$. Recall that $\mathsf{T}_{\Pi^S}(\mathbf{x})$ denotes the query complexity of Π^S on input $\mathbf{x}$. Consider the subset $\mathcal{E}' \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathcal{X}^n | \mathsf{T}_{\Pi^S}(\mathbf{x}) \leq q\}$. By the definition of the average-case cost and Markov's inequality, we know that $\Pr_{\mu_{\mathcal{X}^n}}[\mathcal{E}'] \geq \frac{2}{3}$. Hence $\Pr_{\mu_{\mathcal{X}^n}}[\mathcal{E} \cap \mathcal{E}'] \geq \frac{1}{6}$.

However, we will show that the probability mass of $\mathcal{E} \cap \mathcal{E}'$ is actually small. To that end, we partition $\mathcal{E} \cap \mathcal{E}'$ into depth-q decision trees and bound the overall probability of their leaves. Note that the length-s sketch of the input partitions Π^S into 2^s decision trees. Each leaf of the decision trees corresponds to an output $y \in \mathcal{Y}$ and a list of queries and answers $\sigma' = (X_{i_1} = x_{i_1}, ..., X_{i_{q'}} = x_{i_{q'}}) \in \mathcal{X}^{q'}$ for some $q' \leq q$.

Another condition $H_{\min}(f) > 2s$ implies that, for every $y \in \mathcal{Y}$, every $i_1, i_2, ..., i_q \in [n]$, and every $\sigma = (x_{i_1}, ..., x_{i_q}) \in \mathcal{X}^q$,

$$\Pr_{\mathbf{x} \leftarrow \mu_{\mathcal{X}^n}}[f(\mathbf{x}) = y \text{ and } \mathbf{x} \in \mathcal{E} \cap \mathcal{E}' | \sigma] < 2^{-2s}$$

For the case the decision tree is not full, i.e., there are some $\sigma' = (x_{i_1}, ..., x_{i_{q'}})$ where $q' < q$ is a shorter partial assignment to $\mathbf{x}$. The above inequality still holds simply because we allow redundant queries, and it is equivalent to making $q - q'$ additional redundant queries $i_{q'+1} = ... = i_q = i_1$.

Fix an arbitrary sketch S, we denote by T the corresponding decision tree, and $\mathcal{E}_T$ the set of inputs assigned to this decision tree. We note that every leaf of the tree can be uniquely characterized by its evaluation $f(\mathbf{x}) = y$ and its partial assignment σ' to q elements of the input. The probability mass of the first q layers of the tree can be bounded as follows.

$$\begin{aligned}
\Pr_{\mu_{\mathcal{X}^n}}(\mathcal{E}_T \cap \mathcal{E} \cap \mathcal{E}') &= \sum_{(y,\sigma') \text{ is a leaf of T}} \Pr[\mathbf{x} \in \mathcal{E} \cap \mathcal{E}', f(\mathbf{x}) = y, \sigma'] \\
&= \sum_{(y,\sigma') \text{ is a leaf of T}} \Pr[\sigma'] \cdot \Pr[f(\mathbf{x}) = y \text{ and } \mathbf{x} \in \mathcal{E} \cap \mathcal{E}' | \sigma'] \\
&< \sum_{(y,\sigma') \text{ is a leaf of T}} \Pr[\sigma'] \cdot 2^{-2s} \\
&\leq \frac{1}{2^{2s}}
\end{aligned}$$

The last inequality holds because the partial assignments σ'_1, σ'_2 of every two leaves $(y_1, \sigma'_1), (y_2, \sigma'_2)$ are disjoint events.

By summing up all the 2^s decision trees we have

$$\Pr_{\mu_{\mathcal{X}^n}}(\mathcal{E} \cap \mathcal{E}') < \frac{2^s}{2^{2s}} = \frac{1}{2^s} \leq \frac{1}{8}$$

a contradiction with $\Pr_{\mu_{\mathcal{X}^n}}(\mathcal{E} \cap \mathcal{E}') \geq \frac{1}{6}$.

□

We apply this min-entropy theorem to show that a function f has *average case* query complexity greater than $q/3$ given a precomputed string of length s. For this we will (i) construct an input distribution $\mu_{\mathcal{X}^n}$, (ii) find a large subset $\mathcal{E}$ of $\mathcal{X}^n$, and (iii) show that for every q indices and restriction $\sigma = (X_{i_1} = x_{i_1}, ..., X_{i_q} = x_{i_q})$, $H_{\min}(f) > 2s$.

5 Definition and examples of the recursive model

In this section, we formally define the *space-bounded recursive model*, which operates as a *partially adaptive* model. For clarity in our definition, we begin by introducing the fully adaptive space-bounded recursive model. We then impose certain restrictions on this model to define the partially adaptive version.

The following notation is necessary before giving the definition. Given a list of indices $\text{ind} = ((u_1, v_1), \ldots, (u_N, v_N))$. We define the vector $(\mathbf{M})_{\text{ind}} \stackrel{\text{def}}{=} (\mathbf{M}[u_1, v_1], \ldots, \mathbf{M}[u_N, v_N])$ to be the elements of corresponding indices. We write $[(\mathbf{M})_{\text{ind}}]_\alpha$ to denote the vector whose elements are rounded to precision $\frac{1}{n^\alpha}$.

Definition 3 (adaptive space-bounded recursive model)**.** *Fix the matrix dimension $n \in \mathbb{Z}^+$.*

[Algorithm] *A* space-bounded recursive algorithm for matrix powering*, or* recursive algorithm *for short, is an algorithm that evaluates $[\mathbf{M}^n]_\alpha$ given as input $\mathbf{M}$ for a fixed rounding constant α. The algorithm is determined by $\alpha > 0$ and $\Pi^* = \langle \Pi_1^{S_1}, \ldots, \Pi_{\mathcal{L}}^{S_{\mathcal{L}}} \rangle$, for a number of recursion levels $\mathcal{L} = \omega(1)$, and parameters $1 = k_0 < k_1 < \cdots < k_{\mathcal{L}} = n$, where $k_i | k_{i+1}$. Each recursion level is associated with a distinct query-with-sketch algorithm $\Pi_i^{S_i}$, which computes $[\mathbf{M}^{k_i}]_\alpha$ when given query access to $[\mathbf{M}^{k_{i-1}}]_\alpha$. Each query-with-sketch algorithm $\Pi_i^{S_i}$ has its own private memory S_i, where for all i, $|S_i| = s$, and is restricted in the following syntactic form. $\Pi_i^{S_i}$ has two "states":* clear *and* wait*.*

$$\Pi_i([u,v], S_i, \mathsf{clear}) = \begin{cases} \textit{update } S_i\textit{; return } \mathsf{clear}\textit{; return } g \\ \textit{update } S_i\textit{; return } \mathsf{wait}\textit{; return "query } \mathbf{M}^{k_{i-1}}[u', v']\textit{"} \end{cases}$$

and

$$\Pi_i([u,v], S_i, \mathsf{wait}, g') = \begin{cases} \textit{update } S_i\textit{; return } \mathsf{clear}\textit{; return } g \\ \textit{update } S_i\textit{; return } \mathsf{wait}\textit{; return "query } \mathbf{M}^{k_{i-1}}[u', v']\textit{"} \end{cases}$$

In a correct recursive it must be the case that $g = [\mathbf{M}^{k_i}[u,v]]_\alpha$ and $g' = [\mathbf{M}^{k_{i-1}}[u',v']]_\alpha$.

The meaning of $\Pi_i([u,v], S_i, \mathsf{clear})$ is: determine and return the value of $[\mathbf{M}^{k_i}[u,v]]_\alpha$ in case the internal state of the information-theoretic computing machine (transition system) is not currently performing recursion.

The meaning of $\Pi_i([u,v], S_i, \mathsf{wait}, [\mathbf{M}^{k_{i-1}}[u',v']]_\alpha)$ is: in case the internal state of the machine is to perform a recursive call to determine the value of $[\mathbf{M}^{k_i}[u,v]]_\alpha$, feed the current recursive step with a returned value $[\mathbf{M}^{k_{i-1}}[u',v']]_\alpha$ from the previous recursion level and check if this terminates the current recursive call or whether a new recursive call has to be made.

[Computation] *The computation of Π^* on input $[\mathbf{M}]_\alpha$ consists of n^2 sub-computations each for the computation of $[\mathbf{M}^n[u,v]]_\alpha$ for every $u, v \in \{1, \ldots, n\}$.*

A configuration $C = \langle S_{\mathcal{L}}, \mathsf{clear}\ \textit{or}\ \mathsf{wait}, S_{\mathcal{L}-1}, \mathsf{clear}\ \textit{or}\ \mathsf{wait}, \ldots, S_1, \mathsf{clear}\ \textit{or}\ \mathsf{wait}\rangle$*. It can be inductively shown that a valid configuration will have states of the form* $(\mathsf{wait}, \ldots, \mathsf{wait}, \mathsf{clear}, \ldots, \mathsf{clear})$*, i.e., consecutive* wait *(if there is any* wait *state) followed by consecutive* clear*. Each* S_i *is the private memory of each recursion level. In a valid success of configurations* $C \vdash C'$*, it must be the case that* C' *has been derived from* C *as follows: let* i *be the first* clear *in the sequence; apply* Π_i*; if* Π_i *returns* clear *together with the value* $[\mathbf{M}^{k_i}[u,v]]_\alpha$ *then mark as appropriately the* $\langle S_{i-1}, \text{state}\rangle$ *pair in the configuration and update appropriately by using* Π_{i-1} *(which must be* wait*) with input* $[\mathbf{M}^{k_i}[u,v]]_\alpha$*, which is the output of the previous recursion level. A subcomputation for the element* $[u,v]$*:* $C_1 \vdash C_2 \vdash \cdots \vdash C_\ell$ *is valid if all states are* clear *in* C_1, C_ℓ*, every succession of configurations* $C_j \vdash C_{j+1}$ *is valid and* C_ℓ *outputs the correct value* $[\mathbf{M}^n[u,v]]_\alpha$*.*

[Complexity] *The space of* Π^* *is* $\mathit{Space}(\Pi_h^*) = \mathcal{L} \cdot s$*. Given a computation* $C_1 \vdash \cdots \vdash C_\ell$ *the number of queries* $[u,v]$ *is the cost of the computation, whereas* $\mathsf{T}_{\Pi^*}(\mathbf{M})$ *is the total cost of the computation of* Π^* *on input* $\mathbf{M}$ *is the total number of queries made for computing each* $[u,v]$*. We denote by* $\text{Cost}(\Pi_h^*)$ *the worst-case cost of* Π_h^* *and* $\text{Avg-Cost}_\mu(\Pi^*)$ *the average-case cost of* Π_h^* *for an input distribution* μ *over substochastic matrices.*

$$\text{Cost}(\Pi_h^*) \stackrel{\text{def}}{=} \max_{\mathbf{M}} \mathsf{T}_{\Pi^*}(\mathbf{M}) \quad \text{and} \quad \text{Avg-Cost}_\mu(\Pi_h^*) \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{M} \leftarrow \mu}[\mathsf{T}_{\Pi^*}(\mathbf{M})]$$

where the maximum is taken over all substochastic matrices.

Figure 1 shows an example of the first few steps of a run in this model.

Now, we modify the model to introduce some technically necessary restrictions and make it partially adaptive.

Analogous to the AMP problem in the query-with-sketch model, we accept both $[\mathbf{M}^n]_\alpha$ and $[(\mathbf{M}')^n]_\alpha$ as valid outputs if the input matrices satisfy $[\mathbf{M}]_\alpha = [\mathbf{M}']_\alpha$. Since each recursion level $\Pi_i^{S_i}$ functions as a query-with-sketch algorithm, we extend this consistency requirement to the entire recursive model. Specifically, if the input matrix to the recursive algorithm is $\mathbf{M}$, then for each recursion level i, (i) the input to level i is fixed to be $[\mathbf{M}^{k_{i-1}}]_\alpha$, and (ii) there exists a matrix $\mathbf{M}_i'$ such that $\Pi_i^{S_i}$ always returns elements from $[(\mathbf{M}')^{k_i}]_\alpha$. These restrictions imply that (i) each recursion level is assumed to "repair" its intermediate powers in some way,[13] and (ii) any valid output is acceptable.

In this work, we restrict our attention to *partially adaptive* recursive algorithms. A partially adaptive recursive algorithm is a pair $\langle h, \{\Pi_\iota^*\}_\iota\rangle$, where $h : \{\text{substochastic } n \times n \text{ matrices}\} \to \{0,1\}^{\text{polylog}(n)}$ is a function that given a matrix $[\mathbf{M}]_\alpha$ outputs a polylog(n)-bit string used as an index $\iota = h(\mathbf{M})$. This way we choose an algorithm Π_ι^* from the family of algorithms. Each Π_ι^* is non-adaptive, in the sense that it makes the same queries and in the same order for every input $[\mathbf{M}]_\alpha$, but it is allowed to make arbitrary (intelligent) use of its space at each recursion level to answer these queries. Henceforth, the term recursive algorithm in this work means partially adaptive recursive algorithm.

[13] This assumption does not necessarily strengthen the algorithms, nor does it undermine our core conceptual message: previous works following the Saks-Zhou framework remain valid under this assumption. Several other assumptions are also made without loss of generality, e.g., we assume that the elements of the intermediate powers $[\mathbf{M}^k]_\alpha$ are rounded to its nearest multiple of $n^{-\alpha}$, whereas in [SZ99] elements are rounded down.

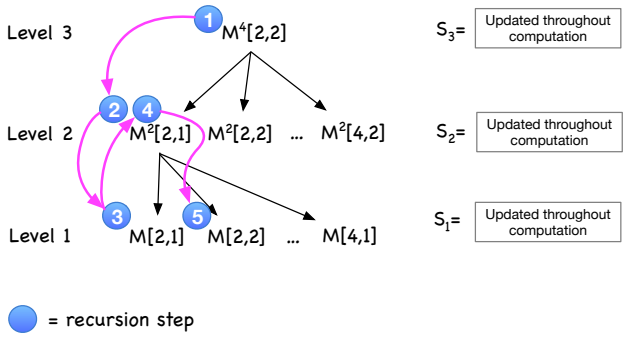

Figure 1: This is an example of a non-adaptive recursive algorithm for calculating M^4 of a 4×4 matrix whose entries are all $1/4$. Say that the workspace at each recursion level is empty (filled with zeros) and each recursion level is "clear". $S_1 = S_2 = S_3 = 00\ldots00$. Suppose that this is the query to determine $M^4[2,2]$.
#1: $\Pi_3([2,2], S_3, clear) = (wait,$ "query $M^2[2,1]$")
Update S_3: $S_3 =$"when the call returns the next query will be $M^2[2,2]$"
#2: $\Pi_2([2,1], S_2, clear) = (wait,$ "query $M[2,1]$")
Update S_2: $S_2 =$"when the call returns the next query will be $M[2,2]$"
#3: $\Pi_1([2,1], S_1, clear) = (clear, 1/4)$
Update S_1: $S_1 =$"still empty"
#4: $\Pi_2([2,1], S_2, wait, 1/4) = (wait,$ "query $M[2,2]$")
Update S_2: $S_2 =$"$M[2,1] = 1/4$ and also when the call returns the next query will be $M[2,3]$"
#5: $\Pi_1([2,2], S_1, clear) = (clear, 1/4)$
Update S_1: $S_1 =$"still empty"

6 Casting the Saks-Zhou's framework in the recursive model

In this section, we summarize previous works that build on [SZ99], collectively referred to as the Saks-Zhou framework. We demonstrate that the recursive model is a generalization of the Saks-Zhou framework. Consequently, these prior works are subject to the lower bounds established by our results.

Our barrier applies to a family of recursive algorithms, encompassing nearly all the previous works based on Saks and Zhou's framework [SZ99, CCvM06, Hoz21, CDSTS23, PP23]. Notably, two exceptions are running in polynomial time: Nisan's pseudorandom generator [Nis90, Nis92], which is not a recursive algorithm, and the catalytic time-space $\mathsf{CTISP}(n, \log n, \log^2 n)$ derandomization [Pyn24], which employs a variant of the Saks-Zhou framework with a constant recursion tree depth. In this section, we summarize the Saks-Zhou framework and compare the differences of previous works in this framework.

Derandomizing BPL naturally reduces the problem of approximating the matrix power of the transition matrix of a space-bounded probabilistic Turing machine. In the seminal work [Nis90, Nis92], Nisan designed a pseudorandom generator whose random seed has $O(\log^2 n)$ bits. By simulating the Turing machine on poly(n) many pseudorandom numbers generated from the random seed, Nisan proved that $BPL \subseteq SC$.

Saks and Zhou [SZ99] utilized the pseudorandom generator in a recursive way, and improved

Algorithm 1 The Saks-Zhou framework

Input: A substochastic matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$.
Output: Approximation of $\mathbf{M}^n$.

1: Choose an "offline" seed $h \in \{0,1\}^{\text{polylog}(n)}$ of the Nisan generator.
2: $\mathbf{M}_0 = \mathbf{M}$ // $\mathbf{M}_i$ *is the approximation of* $\mathbf{M}^{(2^{\sqrt{\log n}})^i}$*, for* $i = 0, ..., \sqrt{\log n}$
3: **for** $i \leftarrow 1$ to $\sqrt{\log n}$ **do**
4: **for** "online" seed $s \in \{0,1\}^{\text{polylog}(n)}$ **do**
5: Simulate $\mathbf{M}_{i-1}^{2^{\sqrt{\log n}}}$ on pseudorandom bits generated from h and s.
6: If the simulation maps x to y, count it at $\mathbf{M}_i[x, y]$.
7: **end for**
8: Adjust $\mathbf{M}_i$ locally.
9: **end for**
10: **return** $\mathbf{M}_{\sqrt{\log n}}$

the space from $O(\log^2 n)$ to $O(\log^{1.5} n)$. As Algorithm 1 shows, they generate an offline seed by using Nisan generator. Then they break down the problem of approximating the n-th power of the matrix into $\sqrt{\log n}$ iterations of computing the intermediate powers. They still approximate the matrix power by simulating the random choices on Nisan's pseudorandom generator produced by the offline seed and the online seed, in which the offline seed is reused in each recursion layer. At line 8 of Algorithm 1, they randomly shift and round $\mathbf{M}_i$ to make sure that the offline seed h looks random to each of the $\sqrt{\log n}$ recursion levels. The whole process can be done in low space.

[CCvM06] is a generalization of [SZ99], where they save the running time by refining the step searching for the offline seed h.

Based on a series of works [Arm98, BCG19, CL20, CDR+21, PV21], [Hoz21] shows that a slightly better derandomization can be obtained by applying weighted pseudorandom generators to the Saks-Zhou framework, where the structure of the algorithm remains unchanged.

In the special case where the size of the matrix is much smaller[14] than n, [CDSTS23, PP23] significantly improved the previous results by truncating and recovering the precision of the matrix powers by using Richardson iteration. In their work, line 8 is replaced by applying the space-efficient Richardson iteration, which can be done locally and non-adaptively.

Here we are ready to give our main result in the context of the Saks-Zhou framework.

Theorem 2 (barrier in the context of the Saks-Zhou framework)**.** *There exists a constant* $c > 0$*. For every algorithm* Π *that lies in the Saks-Zhou framework described in Algorithm 1. If the length of both the offline seed and the online seed are* $O(\text{polylog}(n))$*, and* Π *approximates each intermediate power* $\mathbf{M}^{(2^{\sqrt{\log n}})^i}$ *for* $i = 1, ..., \sqrt{\log n}$ *to within precision* n^{-c} *(or higher precision), then the worst-case running time of* Π *is* $n^{\omega(1)}$*.*

Theorem 2 directly follows from Theorem 3 the intermediate powers barrier, and a reduction from the AMP in the Saks-Zhou framework to the AMP in the recursive model.

We note that the recursive model is generalized from the Saks-Zhou framework above, where some conditions are relaxed. For example, instead of $\sqrt{\log n}$, the lower bound applies to every

[14]That is, approximating the n-th power of a $w \times w$-dimensional matrix, where $w \ll n$.

recursive algorithm with $\omega(1)$ recursion levels. Variants of Theorem 2 can be obtained by applying the same reduction from analogous frameworks.

Proof. We show that the space-bounded recursive model strictly generalizes the Saks-Zhou framework in Algorithm 1. Recall that we defined the recursive model in Section 5 at page 19.

We skip lines 1-2 which are outside the recursive algorithm. The outside for-loop can be implemented as a $\sqrt{\log n}$-level recursive algorithm, where $\mathbf{M}_i$ queries elements from $\mathbf{M}_{i-1}$ for each i. To output $\mathbf{M}_{\sqrt{\log n}}$ we just query and output the n^2 elements $\mathbf{M}_{\sqrt{\log n}}[1,1], ..., \mathbf{M}_{\sqrt{\log n}}[n,n]$ in order. To compute $\mathbf{M}_i[x,y]$, at the i-th recursion level, we enumerate the "online" seeds s using polylog(n) space. We simulate $\mathbf{M}_{i-1}^{2^{\sqrt{\log n}}}[x,y]$ by taking a walk starting from x, and following the pseudorandom choices of PRG(h,s). Nisan's generator helps generate pseudorandom bits on the fly. Finally, we count the number of walks from x to y and obtain the approximation to $\mathbf{M}_i[x,y]$. And we adjust the matrix locally and non-adaptively.

As we can see, the recursive algorithm above has $\omega(1)$ recursion levels. The list of queries made by each recursion level is only depending on the "offline" seed h. Each recursion level i only has query access to elements of its successive recursion level $i-1$. While the algorithm goes in a counting argument, the space needed for each recursion level is $O(\log n)$. In conclusion, the Saks-Zhou framework falls in the space-bounded recursive model. □

7 Main Results

In this section, we present the necessary notation and restate our primary results: the lower bounds for the AMP problem in both the query-with-sketch model and the recursive model.

7.1 Formalization and input distribution

Denote by $\text{out} = ([u_1, v_1], ..., [u_N, v_N])$ the list of indices of the output matrix we want to compute. Recall that we use the vector $(\mathbf{M})_{\text{out}} \stackrel{\text{def}}{=} (\mathbf{M}[u_1, v_1], ..., \mathbf{M}[u_N, v_N])$ to denote the elements of corresponding indices. And $[(\mathbf{M})_{\text{out}}]_\alpha$ denotes the vector whose elements are rounded to precision $\frac{1}{n^\alpha}$.

Here we formally give the nemesis distribution of the underlying matrix $\mathbf{M}$ that is hard for the AMP problem. Fix $n \in \mathbb{N}^{\geq 0}$, and $\gamma, \tau \geq 2$ constant parameters to be determined later. Let $\mathcal{M}_{\gamma,\tau}$ be a distribution of $n \times n$ matrices $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M}_\mathbf{U}$, where $\mathbf{M}_\mathbf{U}$ is a random matrix whose elements are independent and identically distributed (i.i.d.) uniformly from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}] \cap \{d\frac{1}{n^\tau} | d \in \mathbb{Z}^+\}$. We write $\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}$ for $\mathbf{M}$ sampled from distribution $\mathcal{M}_{\gamma,\tau}$. We pick a discrete distribution for the convenience of counting arguments in the proofs.

7.2 Main results

Theorem 3 (main theorem)**.** *Let $n \in \mathbb{N}^{\geq 0}$ and the number of recursion levels $\mathcal{L} = \omega(1)$. Let α, γ, τ to be constants such that $2 \leq \gamma < \alpha - 12.2$, and $\tau > \alpha$. Fix $\mathcal{M}_{\gamma,\tau}$ to be the input distribution constructed above. Let $h = h(\mathbf{M})$ be a function of the input, where $|h| = O(\text{polylog}(n))$. Fix $\Pi_h^* = (\Pi_1^{(\cdot)}, ..., \Pi_\mathcal{L}^{(\cdot)})$ to be an algorithm in the space-bounded recursive model, with input and output precision $\frac{1}{n^\alpha}$. If space size of each recursion level of Π_h^* is $s = O(\text{polylog}(n))$, then $\text{Cost}(\Pi^*) = n^{\omega(1)}$. Moreover, for $\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}$ we have $\text{Avg-Cost}_{\mathcal{M}_{\gamma,\tau}}(\Pi^*) = n^{\omega(1)}$.*

Briefly, any partially-adaptive space-bounded recursive algorithm with $\omega(1)$ recursion levels, space size polylog(n) each level, and a string h with polylog(n) length requires superpolynomial query complexity to approximate the matrix power $\mathbf{M}^n$. We note that α is the rounding parameter, and γ, τ are constants specifying the input distribution. The following corollary captures our main result with fewer parameters.

Corollary 4 (restated theorem 3)**.** *For every large enough constant α, to approximate $\mathbf{M}^n$ given $\mathbf{M}$ as input, where both the input and the output matrices are rounded to precision $n^{-\alpha}$. Any space-bounded recursive algorithm Π^* specified above either requires $\log^{\omega(1)}(n)$ space or $n^{\omega(1)}$ query complexity.*

Theorem 3 follows from the theorem below.

Theorem 4 (main result for query-with-sketch)**.** *Let $n \in \mathbb{N}^{\geq 0}$, and $s = O(\text{polylog}(n))$ be the length of the sketch S. Let $N = \lceil (\frac{s}{\log n})^2 \rceil$. Fix α, γ, τ to be constants such that $2 \leq \gamma < \alpha - 12.2$, and $\tau > \alpha$. Fix $\mathcal{M}_{\gamma,\tau}$ to be the input distribution constructed above. And fix $k < k'$ to be two positive integers such that $k | k'$, $k' \leq n$. Then, for every N many distinct indices $out = ([u_1, v_1], ..., [u_N, v_N])$ of $\mathbf{M}^{k'}$, any algorithm Π^S that correctly computes the output $[(\mathbf{M}^{k'})_{out}]_\alpha = ([\mathbf{M}^{k'}[u_1, v_1]]_\alpha, ..., [\mathbf{M}^{k'}[u_N, v_N]]_\alpha)$ with sketch S has query complexity $\text{Cost}(\Pi^S) = \Omega(\sqrt{n})$ and in particular $\text{Avg-Cost}_{\mathcal{M}_{\gamma,\tau}}(\Pi^S) = \Omega(\sqrt{n})$.*

Here the sketch S corresponds to the private memory S_i in the recursive model. As we assume, S can be an arbitrary function of the input. Intuitively, at least one query is needed when $N > \frac{s}{\log n}$. The above theorem says that to approximate arbitrary $N = \lceil (\frac{s}{\log n})^2 \rceil$ many elements of a matrix power, one needs to make at least $\Omega(\sqrt{n})$ queries. To put things in the proper context, we fix an assignment to the above parameters and restate the theorem in the following.

Corollary 5 (restated theorem 4)**.** *To approximate $\mathbf{M}^n$ given $\mathbf{M}^{n/2}$ as input, where both the input and the output matrices are rounded to precision n^{-20}. Any query with sketch algorithm Π^S with sketch length $\log^2 n$ that approximates arbitrary $\log^2 n$ fixed elements of $\mathbf{M}^n$ requires worst-case query complexity $\Omega(\sqrt{n})$ and average-case query complexity $\Omega(\sqrt{n})$ given the above nemesis distribution $\mathcal{M}_{\gamma,\tau}$.*

Proof of Theorem 3. We first show the special case where h is always a empty string. Then we deal with h later.

Given a space-bounded recursive algorithm Π^* that correctly computes $[\mathbf{M}^n]_\alpha$, the algorithm is non-adaptive given h to be empty.

At level $\mathcal{L}$, the protocol makes n^2 queries to $\Pi_{\mathcal{L}}^{S_{\mathcal{L}}}$ for the whole $[\mathbf{M}^n]_\alpha$. To compute each element, $\Pi_{\mathcal{L}}^{S_{\mathcal{L}}}$ recursively makes queries to its lower recursion level $\mathcal{L} - 1$. By grouping every $N = \lceil (\frac{s}{\log n})^2 \rceil$ successive queries to $\Pi_{\mathcal{L}}^{S_{\mathcal{L}}}$, Theorem 4 tells us that to compute these N elements at least $\Omega(\sqrt{n})$ queries to level $\mathcal{L} - 1$ are necessary in average. We consider the content of the local space $S_{\mathcal{L}}$ before computing the N elements as the sketch. Since Π^* is non-adaptive, $\Omega(\sqrt{n})$ queries must be made on every input.

The above argument applies to each pair of successive recursion levels. As a result, the number of queries blow up by a polynomial factor at each recursion level. To answer the n^2 queries from level $\mathcal{L}$ we need to make at least

$$\sum_{i=1}^{\mathcal{L}} n^2 \cdot \left(\frac{\Omega(\sqrt{n})}{N} \right)^{i-1} = n^{\omega(1)}$$

queries in total, since $N = O(\text{polylog}(n))$ and $\mathcal{L} = \omega(1)$.

Now consider the extra string h, where $|h| = O(\text{polylog}(n))$. The algorithm Π^* becomes a family of $2^{|h|}$ non-adaptive recursive algorithms Π^*_h. For every h' an instance of h, we denote by $\mathcal{M}_{\gamma,\tau}(h')$ (or $\mathcal{M}(h')$ for simplicity) the input distribution conditioned on $h(\mathbf{M}) = h'$, for $\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}$. That is, for every matrix $\mathbf{M}'$,

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}(h')}[\mathbf{M} = \mathbf{M}'] \stackrel{\text{def}}{=} \Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[\mathbf{M} = \mathbf{M}'|h(\mathbf{M}) = h']$$

We will show that Theorem 4 still holds given input distribution $\mathcal{M}(h')$ if $\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M}) = h'] = \frac{1}{2^{O(\text{polylog}(n))}}$, and if we slightly increase N. The key observation is that the min-entropy of the output $[(\mathbf{M}^{k'})_{\text{out}}]_\alpha$ in Theorem 4 is still high. This is because $\mathcal{M}(h')$ takes a part from $\mathcal{M}_{\gamma,\tau}$, and zooms in by a factor of $2^{O(\text{polylog}(n))}$. By increasing N, we can still get the desired bound.

We call $h' \in \{0,1\}^{|h|}$ *typical* if $\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M}) = h'] \geq \frac{1}{n^2\cdot 2^{|h|}}$. We will increase N to $\lceil(\frac{s+|h|+2\log n}{\log n})^2\rceil$, which is still polylog$(n)$. Notice that Theorem 4 follows directly from combining Theorem 1 (pp. 17), Theorem 5 (pp. 37) and Theorem 6 (pp. 48). We plug this increased N into Theorem 5 and Theorem 6, and get

$$H_{\min}([(\mathbf{M}^{k'})_{\text{out}}]_\alpha|\mathcal{M}_{\gamma,\tau},\mathcal{E},\sigma) \geq 2(s + |h| + 2\log n)$$

for every partial assignment σ and a set $\mathcal{E}$ whose definition is not important for now. When we replace $\mathcal{M}_{\gamma,\tau}$ by $\mathcal{M}(h')$ for an arbitrary typical h', observe that for every vector $\mathbf{y}$ and partial assignment σ,

$$\begin{aligned}
&\Pr_{\mathbf{M}\leftarrow\mathcal{M}(h')}[[(\mathbf{M}^{k'})_{\text{out}}]_\alpha = \mathbf{y} \text{ and } \mathbf{M}\in\mathcal{E}|\sigma]\\
\leq&\frac{\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[[(\mathbf{M}^{k'})_{\text{out}}]_\alpha = \mathbf{y} \text{ and } \mathbf{M}\in\mathcal{E}|\sigma]}{\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M}) = h']}\\
\leq&n^2\cdot 2^{|h|}\cdot\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[[(\mathbf{M}^{k'})_{\text{out}}]_\alpha = \mathbf{y} \text{ and } \mathbf{M}\in\mathcal{E}|\sigma]
\end{aligned}$$

which means that

$$H_{\min}([(\mathbf{M}^{k'})_{\text{out}}]_\alpha|\mathcal{M}(h'),\mathcal{E},\sigma) \geq 2s$$

By Theorem 1, the query lower bound follows. And we get an edited version of Theorem 4 where (i) the input distribution is replaced by $\mathcal{M}(h')$ for arbitrary typical h' (ii) N is increased to $\lceil(\frac{s+|h|+2\log n}{\log n})^2\rceil$. Combined with the proof above, it is now clear that when $h' = h(\mathbf{M})$ is typical, the average-case query complexity of Π^* is still $n^{\omega(1)}$.

We get the lower bound for the case $h' = h(\mathbf{M})$ is typical. Now we show that the probability that $h(\mathbf{M})$ is typical is high.

$$\begin{aligned}
&\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M}) \text{ is typical}]\\
=&\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[\Pr_{\mathbf{M}'\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M}) = h(\mathbf{M}')] \geq \frac{1}{n^2\cdot 2^{|h|}}\right]\\
\geq&1 - n^{-2}
\end{aligned}$$

This is because the probability mass of non-typical h' is upper bounded by $2^{|h|}\cdot\frac{1}{n^2\cdot 2^{|h|}}$ in total.

Therefore, the average-case query complexity of Π^* given input distribution $\mathcal{M}_{\gamma,\tau}$ and extra advice bits h is lower-bounded by:

$$\begin{aligned}
&\text{Avg-Cost}_{\mathcal{M}_{\gamma,\tau}}(\Pi^*)\\
\geq &\sum_{h' \text{ is typical}} \Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}[h(\mathbf{M})=h'] \cdot \text{Avg-Cost}_{\mathcal{M}(h')}(\Pi^*)\\
\geq &(1-n^{-2})\cdot n^{\omega(1)}\\
= &n^{\omega(1)}
\end{aligned}$$

□

The remaining part of this paper will focus on proving Theorem 4.

8 Reducing the min-entropy bound by the analysis of perturbations

In this section, we reduce the min-entropy bound of the AMP to an upper bound on a more manageable conditional probability. We note that a significant portion of this section involves calculating the partial derivatives of matrix powers, while some of these calculations are straightforward and repetitive, but necessary. Bounding these partial derivatives is essential for establishing the reductions.

8.1 Roadmap to the argument

Recall that the lower bound to the multi-valued AMP problem naturally reduces to the lower bound of each fixed-valued AMP problem. We denote by $f_{\text{fix}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ the fixed-valued AMP function that maps $[\mathbf{M}^k]_\alpha$ to $[\mathbf{M}'^{k'}]_\alpha$ for some $\mathbf{M}'$ such that $[\mathbf{M}^k]_\alpha = [(\mathbf{M}')^k]_\alpha$. That is, given $[\mathbf{M}^k]_\alpha$ as input, we fix $f_{\text{fix}}([\mathbf{M}^k]_\alpha)$ as the output matrix. For simplicity, we write it as $f_{\text{fix}}(\mathbf{M})$. We will simultaneously lower bound the min-entropy of each $f_{\text{fix}}(\mathbf{M})$. By the end of this section, bounding the min-entropy of $(f_{\text{fix}}(\mathbf{M}))_{\text{out}}$ will be reduced to bounding the probability of an event of $\mathbf{M}^k$, while getting rid of the function f_{fix}.

By Theorem 1 it suffices to show that for every $q = \sqrt{n}$ many elements $((i_1,j_1),\ldots,(i_q,j_q))$, every vector $\mathbf{w}\in\mathbb{R}^q$ and event $\sigma = \big(([\mathbf{M}^k[i_1,j_1]]_\alpha,\ldots,[\mathbf{M}^k[i_q,j_q]]_\alpha)=\mathbf{w}\big)$, the min-entropy of the output $(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = (f_{\text{fix}}(\mathbf{M})[u_1,v_1],\ldots,f_{\text{fix}}(\mathbf{M})[u_N,v_N])$ is high. That is, if you do not know a proportion of the elements of $\mathbf{M}^k$, then we will know little about $\mathbf{M}^{k'}$. This is not true in general, but is true given our nemesis distribution. Recall that our input distribution is $\mathcal{M}_{\gamma,\tau} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U}$,[15] where $\mathbf{M_U}$ is a matrix whose elements are independent and identically distributed (i.i.d.) uniformly from $[\frac{1}{n}-\frac{1}{n^\gamma},\frac{1}{n}]\cap\{d\frac{1}{n^\tau}|d\in\mathbb{Z}^+\}$. Recall that α,τ,γ are constants satisfying $2\leq\gamma<\alpha-12.2$ and $\tau>\alpha$. Matrices are drawn from a small range since γ is lower bounded.

We will demonstrate that the matrix power, given this distribution, maintains the desired high min-entropy. Roughly speaking, this result can be interpreted as the power of the transition matrix mixing quite slowly, thereby preserving high entropy.

[15] For simplicity, we denote by the input distribution $\mathcal{M}$ the distribution of the underlying matrix $\mathbf{M}$ instead of the input matrix $[\mathbf{M}^k]_\alpha$, where the previous one is more manageable.

To achieve this, we first quantify the effect of a perturbation[16] on $\mathbf{M}^k$ to $\mathbf{M}^{k'}$, i.e., the partial derivatives $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]}$, $\frac{\partial \mathbf{M}^{k'}[u,v]}{\partial \mathbf{M}^k[i,j]}$ for each $u, v, i, j \in [n]$.

Based on that, we reduce the task of lower bounding the entropy of $(f_{\text{fix}}(\mathbf{M}))_{\text{out}}$ to upper bounding the conditional probability $\lceil\sqrt{N}\rceil$ elements of $\mathbf{M}^k$ falling in some small range, during which we get rid of both the entropy and the fixed-valued function $f_{\text{fix}}(\cdot)$. This will be presented in Theorem 5, the main result of this section.

8.2 Partial derivatives bound: from M to $\mathbf{M}^k$

In the following two subsections, we will show how larger matrix powers will change when we make small perturbations to small powers given input distribution $\mathcal{M} : \mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M}_{\mathbf{U}}$. We regard each element of the larger matrix power as a polynomial of elements of the smaller matrix power while applying derivations. To be clear, the matrices are differentiable variables on $\mathbb{R}^{n\times n}$ for now, instead of discrete random matrices.

Before delving into the partial derivations, we give the following inequality which will be used numerous times to bound the partial derivatives.

Lemma 6. *For every integer $n > 1$,*

$$(1-\frac{1}{n})^n \le e^{-1} \le (1-\frac{1}{n})^{n-1}$$

Proof. For simplicity, we let $f(n) = (1-\frac{1}{n})^n$, $g(n) = (1-\frac{1}{n})^{n-1}$ for now. It is known that

$$\lim_{n\to\infty} f(n) = \lim_{n\to\infty} g(n) = e^{-1}$$

Given this fact, to show the inequalities hold whenever $n > 1$, we show that (i) the inequalities hold true when $n = 2$, and (ii) $f(n)$ is an increasing function and $g(n)$ is a decreasing function.

One can easily verify that the inequalities hold for the case $n = 2$. To show the monotonicity of $f(n)$ and $g(n)$,

$$\begin{aligned} f'(n) &= (e^{n\ln(1-1/n)})' = (1-\frac{1}{n})^n\left(\ln(1-1/n) + \frac{1}{n-1}\right) \\ &= (1-\frac{1}{n})^n\left(\left(-\frac{1}{n} - \frac{1}{2n^2} - \frac{1}{3n^3} - ...\right) + \frac{1}{n-1}\right) > 0 \end{aligned}$$

when $n \ge 2$.

$$\begin{aligned} g'(n) &= (e^{(n-1)\ln(1-1/n)})' = (1-\frac{1}{n})^{n-1}\left(\ln(1-1/n) + \frac{1}{n}\right) \\ &= (1-\frac{1}{n})^n\left(\left(-\frac{1}{n} - \frac{1}{2n^2} - \frac{1}{3n^3} - ...\right) + \frac{1}{n}\right) < 0 \end{aligned}$$

when $n \ge 2$. □

[16] We note that the perturbation here comes with a different purpose to the perturbation in [SZ99]. In [SZ99], small perturbations are applied to intermediate matrix powers to ensure the correctness of reusing Nisan's pseudorandom generator.

Lemma 7. *Let $\gamma \geq 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every positive integer $k \leq n$, for every element $[i, j]$ of $\mathbf{M_U}$ and every element $[u, v]$ of ${\mathbf{M_U}}^k$. If $k \geq 2$, we have:*

$$\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \begin{cases} \Theta\left(\frac{2n+k-2}{n^2}\right), & u = i \text{ and } v = j \\ \Theta\left(\frac{n+k-2}{n^2}\right), & u = i \text{ or } v = j, \text{ but not both} \\ \Theta\left(\frac{k-2}{n^2}\right), & \text{otherwise} \end{cases}$$

If $k = 1$, we have:

$$\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \begin{cases} 1, & u = i \text{ and } v = j \\ 0, & \text{otherwise} \end{cases}$$

When $k = o(n)$ where $k > 1$, we can see that each element in ${\mathbf{M_U}}^k$ depends more heavily on the $2n - 1$ out of n^2 elements of $\mathbf{M_U}$.

Proof of Lemma 7. The partial derivation of the case $k = 1$ is trivial.

For $k \geq 2$, note that

$${\mathbf{M_U}}^k[u,v] = \sum_{u=h_0,h_1,\ldots,h_{k-1},h_k=v} \prod_{t=0}^{k-1} \mathbf{M_U}[h_t, h_{t+1}]$$

So the partial derivative is

$$\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \sum_{g=0}^{k-1} \left(\sum_{u=h_0,h_1,\ldots,h_g=i,h_{g+1}=j,\ldots,h_{k-1},h_k=v} \left(\prod_{t=0,t\neq g}^{k-1} \mathbf{M_U}[h_t, h_{t+1}] \right) \right)$$

In this equation the appearance of $\mathbf{M_U}[i, j]$ is canceled at position g in the summation of the product. Since for all $i, j \in \{1, \ldots, n\}$, $\mathbf{M_U}[i, j] \in [\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$, which is a quite small range, $\prod_{t=0,t\neq g}^{k-1} \mathbf{M_U}[h_t, h_{t+1}]$ can be lower-bounded by setting all elements in $\mathbf{M_U}$ equal to $\frac{1}{n} - \frac{1}{n^\gamma}$ and upper bounded by setting all elements in $\mathbf{M_U}$ equal to $\frac{1}{n}$. We only need to count the number of sequences $u = h_0, h_1, ..., h_{k-1}, h_k = v$ in which two adjacent indices are i and j, while in both the lower and upper bound all elements are equivalent. If $u \neq i, v \neq j$, g can only be picked from $\{1, 2, ..., k-2\}$, since $h_g = i$, $h_{g+1} = j$. Then, all the rest of $k - 3$ indices can be chosen from $[n]$. Therefore, there are $(k-2)n^{k-3}$ distinct sequences. Of course, at the edge case $k = 2$, g cannot be from $\{1, 2, ..., k-2\}$, and the count becomes 0. For the case $u = i$ or $v = j$ but not both, g can be 0 or $k - 1$, which gives additional n^{k-2} sequences. Hence, the total number of distinct sequences is $n^{k-2} + (k-2)n^{k-3}$. Similarly, when $u = i, v = j$ the total number of sequences is $2n^{k-2} + (k-2)n^{k-3}$.

If $u \neq i$, $v \neq j$,

$$(k-2)n^{-2} = (k-2)n^{k-3}\left(\frac{1}{n}\right)^{k-1} \geq \frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} \geq (k-2)n^{k-3}\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)^{k-1} \geq e^{-\frac{k-1}{n^{\gamma-1}-1}}(k-2)n^{-2}$$

Recall that $\gamma \geq 2$, so $\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \Theta(\frac{k-2}{n^2})$.

Similarly, if $u = i$ or $v = j$ but not both happens, then

$$n^{-1} + (k-2)n^{-2} = (n^{k-2} + (k-2)n^{k-3})\left(\frac{1}{n}\right)^{k-1} \geq \frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]} \geq (n^{k-2} + (k-2)n^{k-3})\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)^{k-1}$$
$$\geq e^{-\frac{k-1}{n^{\gamma-1}-1}}\left(\frac{1}{n} + \frac{k-2}{n^2}\right)$$

namely $\frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \Theta(\frac{n+k-2}{n^2})$.

And finally, if $u = i$, $v = j$,

$$\frac{2}{n} + \frac{k-2}{n^2} = (2n^{k-2} + (k-2)n^{k-3})\left(\frac{1}{n}\right)^{k-1} \geq \frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]} \geq (2n^{k-2} + (k-2)n^{k-3})\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)^{k-1}$$
$$\geq e^{-\frac{k-1}{n^{\gamma-1}-1}}\left(\frac{2}{n} + \frac{k-2}{n^2}\right)$$

namely $\frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \Theta(\frac{2n+k-2}{n^2})$.

Note that the $k-2$ term becomes 0 when $k = 2$. So the above bounds also apply to the case $k = 2$. □

Furthermore, we need the following corollary. We will not use it in this section, but we will use it later on to obtain the Lipschitz constant of the matrix power when we introduce our geometric machinery in section 9.2. We use $||\mathbf{M}||$ to denote the entry-wise max norm of a matrix throughout this paper.

Corollary 8. *Let $\gamma \geq 2$ and $2 \leq k \leq n$. Let $\mathbf{M_U}, \mathbf{M_U}'$ be two matrices whose elements range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. We have*

$$||\mathbf{M_U}^k - \mathbf{M_U}'^k|| \leq k||\mathbf{M_U} - \mathbf{M_U}'||$$

Proof. For convenience, let $x_{1,1}, ..., x_{n,n}$ be the elements of $\mathbf{M_U}$, and $y_{1,1}, ..., y_{n,n}$ be the elements of $\mathbf{M_U}'$. And let $d = ||\mathbf{M_U} - \mathbf{M_U}'||$. For each $(u,v) \in \{1,\dots,n\} \times \{1,\dots,n\}$, we rewrite $(\mathbf{M_U}^k)[u,v]$ as $f_{k,u,v}(x_{1,1}, ..., x_{n,n}) = (\mathbf{M_U} = (x_{1,1}, ..., x_{n,n}))^k[u,v]$ a function of n^2 elements. Then

$$\begin{aligned}
&||\mathbf{M_U}^k - \mathbf{M_U}'^k|| \\
=&\max_{u,v\in[n]} |f_{k,u,v}(x_{1,1}, ..., x_{n,n}) - f_{k,u,v}(y_{1,1}, ..., y_{n,n})| \\
=&\max_{u,v\in[n]} \left| \sum_{i,j\in[n]} \int_{x_{i,j}}^{y_{i,j}} \frac{\partial f_{k,u,v}(x_{1,1}, ..., x_{i,j-1}, z, y_{i,j+1}, ..., y_{n,n})}{\partial z} dz \right| \\
\leq&\frac{2n+k-2}{n^2} \cdot d + \frac{n+k-2}{n^2} \cdot (2n-2) \cdot d + \frac{k-2}{n^2} \cdot (n^2 - 2n + 1) \cdot d \\
=&k \cdot d = k \cdot ||\mathbf{M_U} - \mathbf{M_U}'||
\end{aligned}$$

where the inequality follows from the upper bound to the partial derivatives $\frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]}$ in Lemma 7. □

Given Lemma 7 and the fact that $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U}$ which is linear to $\mathbf{M_U}$, observe the following

$$\mathbf{M}^k = \sum_{t=0}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\mathbf{M_U}\right)^t$$

We have

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \sum_{t=0}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]} \tag{1}$$

Now, we are ready to prove the following lemma.

Lemma 9. *Let $\gamma \geq 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every positive integer $k \leq n$, every element (i,j) of $\mathbf{M}$ and every element (u,v) of $\mathbf{M}^k$, we have:*

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \begin{cases} \Theta(k), & u = i \text{ and } v = j \\ \Theta\left(\frac{k^2}{n^4}\right), & u = i \text{ or } v = j, \text{ but not both; and } k \geq 2 \\ \Theta\left(\frac{k^3}{n^8}\right), & u \neq i, v \neq j, \text{ and } k \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

Proof. It is not hard to see that at the edge case, the partial derivation becomes 0.

Note that in the equation (1), $\binom{k}{t}\left(\frac{n^2-1}{n^2}\right)^{k-t}\left(\frac{1}{n^3}\right)^{t-1}$ will decrease by a factor of at least n^2 when t increases by 1. For fixed u, v, i, j and $t < n$, $\frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]}$ is of the same order for two successive t. Thus, the summation asymptotically forms a geometric series. We only need to find the smallest t such that $\frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]}$ is non-zero, which dominates the summation.

If $u = i$, $v = j$, then

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \left(\frac{n^2-1}{n^2}\right)^{k-1} k + \sum_{t=2}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]} = \Theta(k) \tag{2}$$

If $u = i$ or $v = j$ but not both happens, then

$$\begin{aligned} \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} &= \left(\frac{n^2-1}{n^2}\right)^{k-2} \binom{k}{2} \left(\frac{1}{n^3}\right) \frac{\partial {\mathbf{M_U}}^2[u,v]}{\partial \mathbf{M_U}[i,j]} + \sum_{t=3}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]} \\ &= \Theta\left(\frac{k^2}{n^4}\right) \end{aligned} \tag{3}$$

If $u \neq i$, $v \neq j$, then

$$\begin{aligned} \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} &= \left(\frac{n^2-1}{n^2}\right)^{k-3} \binom{k}{3} \left(\frac{1}{n^3}\right)^2 \frac{\partial {\mathbf{M_U}}^3[u,v]}{\partial \mathbf{M_U}[i,j]} + \sum_{t=4}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]} \\ &= \Theta\left(\frac{k^3}{n^8}\right) \end{aligned} \tag{4}$$

□

Also, the same as in Corollary 8, we need the following bound precise to constant, which will be used to obtain the Lipschitz constant in Section 9.2.

Corollary 10. *Suppose n is sufficiently large. Let $\gamma \geq 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every integer $k \leq n$, for every element $[i, j]$ of $\mathbf{M}$ and every element $[u, v]$ of $\mathbf{M}^k$, we have the upper bound of $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]}$:*

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} \leq \begin{cases} k, & u = i \text{ and } v = j \\ \dfrac{k^2}{2n^4}, & u = i \text{ or } v = j, \text{ but not both; and } k \geq 2 \\ \dfrac{k^3}{6n^8}, & u \neq i, v \neq j, \text{ and } k \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

Proof. This corollary follows from the intermediate results of the proof of the above two lemmas. We only need to bound the first three cases, since the edge case follows directly from Lemma 9.

We first know from the proof of Lemma 7 that when $k \geq 2$,

$$\frac{\partial {\mathbf{M}_\mathbf{U}}^k[u,v]}{\partial \mathbf{M}_\mathbf{U}[i,j]} \leq \begin{cases} \dfrac{2}{n} + \dfrac{k-2}{n^2}, & u = i \text{ and } v = j \\ \dfrac{1}{n} + \dfrac{k-2}{n^2}, & u = i \text{ or } v = j, \text{ but not both} \\ \dfrac{k-2}{n^2}, & \text{otherwise} \end{cases}$$

When $u = i$, $v = j$, by (2) we have

$$\begin{aligned} \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} &\leq \left(\frac{n^2-1}{n^2}\right)^{k-1} k + \sum_{t=2}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \left(\frac{2}{n} + \frac{t-2}{n^2}\right) \\ &< \left(k - \frac{k(k-1)}{n^2} + \frac{k^3}{2n^4}\right) + O\left(\frac{k^2}{n^4}\right) \\ &< k \end{aligned}$$

When $u = i$ or $v = j$ but not both by (3) we have

$$\begin{aligned} \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} &\leq \left(\frac{n^2-1}{n^2}\right)^{k-2} \frac{k(k-1)}{2n^4} + \sum_{t=3}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \left(\frac{1}{n} + \frac{t-2}{n^2}\right) \\ &< \left(1 - \frac{k-2}{n^2} + \frac{k^2}{n^4}\right) \frac{k^2}{2n^4} + O\left(\frac{k^3}{n^7}\right) \\ &< \frac{k^2}{2n^4} \end{aligned}$$

When $u \neq i$, $v \neq j$, by (4) we have

$$\begin{aligned}\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} &\leq \left(\frac{n^2-1}{n^2}\right)^{k-3} \frac{k(k-1)(k-2)}{6n^8} + \sum_{t=4}^{k} \binom{k}{t} \left(\frac{n^2-1}{n^2}\right)^{k-t} \left(\frac{1}{n^3}\right)^{t-1} \frac{t-2}{n^2} \\ &< \left(1 - \frac{k-3}{n^2} + \frac{k^2}{2n^4}\right) \frac{k^3}{6n^8} + O\left(\frac{k^4}{n^{11}}\right) \\ &< \frac{k^3}{6n^8}\end{aligned}$$

□

8.3 Partial derivatives bound: from $\mathbf{M}^k$ to $\mathbf{M}^{k'}$

The previous section bounds to precision the partial derivatives of each $\mathbf{M}^k[u,v]$ to each $\mathbf{M}[i,j]$. Things become more involved when we consider the relationship between $\mathbf{M}^k$ and $\mathbf{M}^{k'}$, where the dependencies of elements in $\mathbf{M}^k$ are non-trivial. More precisely, we will compute $\frac{\partial \mathbf{M}^{k'}[u,v]}{\partial \mathbf{M}^k[i,j]}$ for every $u,v,i,j \in \{1,\ldots,n\}$.[17]

Lemma 11. *Let $\gamma \geq 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. Let $1 \leq k < k' \leq n$ be integers such that $k|k'$. For every element $[i,j]$ of $\mathbf{M}^k$ and every element $[u,v]$ of $\mathbf{M}^{k'}$, we have:*

$$\frac{\partial \mathbf{M}^{k'}[u,v]}{\partial \mathbf{M}^k[i,j]} = \begin{cases} \Theta\left(\frac{k'}{k}\right), & u = i \text{ and } v = j \\ \Theta\left(\frac{k'^2}{n^4 k}\right), & u = i \text{ or } v = j, \text{ but not both; and } \frac{k'}{k} \geq 2 \\ \Theta\left(\frac{k'^3}{n^8 k}\right), & u \neq i, v \neq j, \text{ and } \frac{k'}{k} \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

For simplicity, let us denote $t = \frac{k'}{k}$ and in this way the above becomes:

$$\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]} = \begin{cases} \Theta(t), & u = i \text{ and } v = j \\ \Theta\left(\frac{kt^2}{n^4}\right), & u = i \text{ or } v = j, \text{ but not both; and } t \geq 2 \\ \Theta\left(\frac{k^2 t^3}{n^8}\right), & u \neq i, v \neq j, \text{ and } t \geq 3 \\ 0, & \text{otherwise} \end{cases}$$

Since $k, t \leq n$, we can see that each $\mathbf{M}^{kt}[i,j]$ depends more heavily on $2n-1$ out of n^2 elements of $\mathbf{M}^k$.

The key observation is the following

$$\mathbf{M}^k = \left(\frac{n^2-1}{n^2}\right)^k \mathbf{I} + \sum_{r=1}^{k} \binom{k}{r} \left(\frac{n^2-1}{n^2}\right)^{k-r} \left(\frac{1}{n^3}\mathbf{M_U}\right)^r$$

[17] Notice that here we ignore the effect of $\partial \mathbf{M}^k[i,j]$ on other elements in $\partial \mathbf{M}^k$. Because for the purpose of this section to reduce the entropy of $\mathbf{M}^{k'}$ to the entropy of $\mathbf{M}^k$, we could for now regard $\mathbf{M}^{k'}$ as a function of $\mathbf{M}^k$, and to show that the function preserves entropy. Hence discard the distribution for now.

Let $C \stackrel{\text{def}}{=} \left(\frac{n^2-1}{n^2}\right)^k$ and $\bar{\mathbf{M}} \stackrel{\text{def}}{=} \sum_{r=1}^{k} \binom{k}{r} \left(\frac{n^2-1}{n^2}\right)^{k-r} (\frac{1}{n^3}\mathbf{M_U})^r$, then $\mathbf{M}^k = C\mathbf{I} + \bar{\mathbf{M}}$. The matrix $\bar{\mathbf{M}}$ is defined with the sole purpose of reducing notational clutter.

Observe that $C = \Theta(1)$, and every element of $\bar{\mathbf{M}}$ is also in a small range. In this way, we rewrite $\mathbf{M}^k$ in a similar form as $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U}$. Thus we can expect similar bounds to still hold.

Proof of Lemma 11. The edge case where $\mathbf{M}^{kt}[u,v]$ does not depend on $\mathbf{M}^k[i,j]$ obviously gives a 0 partial derivative. Also, we may assume that $k \geq 2$ since we have already shown the case $k = 1$ in Lemma 9.

Given $\mathbf{M}^k = C\mathbf{I} + \bar{\mathbf{M}}$, we have

$$\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]} = \sum_{t'=0}^{t} \binom{t}{t'} C^{t-t'} \frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]}$$

We need to bound the value of $\bar{\mathbf{M}}[i,j]$ before bounding the partial derivatives.

From the lower bound side, when n is large enough

$$\bar{\mathbf{M}}[i,j] \geq k\left(\frac{n^2-1}{n^2}\right)^{k-1} \frac{1}{n^3}\mathbf{M_U}[i,j] \geq \frac{e^{-\frac{k-1}{n^2-1}} \cdot k}{n^3}\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)$$

From the upper bound side, when n is large enough

$$\begin{aligned}
\bar{\mathbf{M}}[i,j] &\leq \frac{k}{n^4} \cdot \left(\frac{n^2-1}{n^2}\right)^{k-1} + \sum_{r=2}^{k} \binom{k}{r} \left(\frac{n^2-1}{n^2}\right)^{k-r} \left(\frac{1}{n^3}\right)^r \frac{1}{n} \\
&\leq \frac{k}{n^4} \cdot \left(\frac{n^2-1}{n^2}\right) + \frac{k^2}{n^7} \sum_{r=0}^{k-2} \binom{k-2}{r} \left(\frac{n^2-1}{n^2}\right)^{k-2-r} \left(\frac{1}{n^3}\right)^r \\
&\leq \frac{k}{n^4} - \frac{k}{n^6} + \frac{k^2}{n^7} \leq \frac{k}{n^4}
\end{aligned}$$

Clearly, when $t' = 1$, $\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \begin{cases} 1, & \text{if } u = i, v = j \\ 0, & \text{otherwise} \end{cases}$

Now, we proceed with the case where $t' \geq 2$. We apply the same proof as in Lemma 7, considering $\bar{\mathbf{M}}$ in place of $\mathbf{M_U}$.

If $u = i$ and $v = j$,

$$\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} \leq \left(2n^{t'-2} + (t'-2)n^{t'-3}\right)\left(\frac{k}{n^4}\right)^{t'-1} \leq \frac{k^{t'-1}}{n^{3t'-1}}(2n + t' - 2)$$

$$\begin{aligned}
\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} &\geq \left(2n^{t'-2} + (t'-2)n^{t'-3}\right)\left(\frac{e^{-\frac{k-1}{n^2-1}} \cdot k}{n^3}\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)\right)^{t'-1} \\
&\geq (2n + t' - 2)\frac{k^{t'-1}}{n^{3t'-1}} \cdot e^{-\frac{(k-1)(t'-1)}{n^2-1}} \cdot e^{-\frac{t'-1}{n^{\gamma-1}-1}} \geq e^{-O(1)} \cdot (2n + t' - 2)\frac{k^{t'-1}}{n^{3t'-1}}
\end{aligned}$$

where the inequalities follows from Lemma 6 and the fact that $\gamma \geq 2$. Thus, $\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta\left(\frac{k^{t'-1}}{n^{3t'-2}}\right)$.

Similarly, if $u = i$ or $v = j$ but not both happens, then

$$\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} \le \left(n^{t'-2} + (t'-2)n^{t'-3}\right)\left(\frac{k}{n^4}\right)^{t'-1} \le \frac{k^{t'-1}}{n^{3t'-1}}(n+t'-2)$$

$$\begin{aligned}\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} &\ge \left(n^{t'-2} + (t'-2)n^{t'-3}\right)\left(\frac{e^{-\frac{k-1}{n^2-1}}\cdot k}{n^3}\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)\right)^{t'-1} \\ &\ge (n+t'-2)\frac{k^{t'-1}}{n^{3t'-1}} \cdot e^{-\frac{(k-1)(t'-1)}{n^2-1}} \cdot e^{-\frac{t'-1}{n^{\gamma-1}-1}} \ge e^{-O(1)} \cdot (n+t'-2)\frac{k^{t'-1}}{n^{3t'-1}}\end{aligned}$$

namely $\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta\left(\frac{k^{t'-1}}{n^{3t'-2}}\right)$.

If $u \neq i, v \neq j$,

$$\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} \le (t'-2)n^{t'-3}\left(\frac{k}{n^4}\right)^{t'-1} \le (t'-2)\frac{k^{t'-1}}{n^{3t'-1}}$$

$$\begin{aligned}\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} &\ge (t'-2)n^{t'-3}\left(\frac{e^{-\frac{k-1}{n^2-1}}\cdot k}{n^3}\left(\frac{1}{n} - \frac{1}{n^\gamma}\right)\right)^{t'-1} \\ &\ge (t'-2)\frac{k^{t'-1}}{n^{3t'-1}} \cdot e^{-\frac{(k-1)(t'-1)}{n^2-1}} \cdot e^{-\frac{t'-1}{n^{\gamma-1}-1}} \ge e^{-O(1)} \cdot (t'-2)\frac{k^{t'-1}}{n^{3t'-1}}\end{aligned}$$

namely $\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta\left((t'-2)\frac{k^{t'-1}}{n^{3t'-1}}\right)$.

Now we can bound $\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]}$. Recall that $\mathbf{M}^k = C\mathbf{I} + \bar{\mathbf{M}}$ and $C = \left(\frac{n^2-1}{n^2}\right)^k$.

If $u = i$, $v = j$, then

$$\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]} = t \cdot \left(\frac{n^2-1}{n^2}\right)^{k(t-1)} + \sum_{t'=2}^{t}\binom{t}{t'}C^{t-t'}\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta(t)$$

If $u = i$ or $v = j$ but not both happens, then

$$\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]} = \binom{t}{2}\left(\frac{n^2-1}{n^2}\right)^{k(t-2)}\frac{\partial \bar{\mathbf{M}}^2[u,v]}{\partial \bar{\mathbf{M}}[i,j]} + \sum_{t'=3}^{t}\binom{t}{t'}C^{t-t'}\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta(\frac{kt^2}{n^4})$$

If $u \neq i$, $v \neq j$, then

$$\frac{\partial \mathbf{M}^{kt}[u,v]}{\partial \mathbf{M}^k[i,j]} = \binom{t}{3}\left(\frac{n^2-1}{n^2}\right)^{k(t-3)}\frac{\partial \bar{\mathbf{M}}^3[u,v]}{\partial \bar{\mathbf{M}}[i,j]} + \sum_{t'=4}^{t}\binom{t}{t'}C^{t-t'}\frac{\partial \bar{\mathbf{M}}^{t'}[u,v]}{\partial \bar{\mathbf{M}}[i,j]} = \Theta(\frac{k^2t^3}{n^8})$$

□

8.4 Perturbation analysis

In this section, we reduce lower bounding the min-entropy of the output $(f_{\text{fix}}(\mathbf{M}))_{\text{out}}$ to upper bounding the probability some elements of $\mathbf{M}^k$ fall into some small range, so that it depends only on $\mathbf{M}^k$. Recall that $f_{\text{fix}}(\mathbf{M}) = [\mathbf{M}'^{k'}]_\alpha$ for some $[\mathbf{M}'^k]_\alpha = [\mathbf{M}^k]_\alpha$.

One can see that $f_{\text{fix}}(\mathbf{M})$ is close to $[\mathbf{M}^{k'}]_\alpha$, where $\mathbf{M}^{k'}$ is more manageable. Let us discard f_{fix} but focus on $[\mathbf{M}^{k'}]_\alpha$ for now, to better figure out the intuition of this section. We will bring f_{fix} back in our formal argument.

We remove the dependence from $\mathbf{M}^{k'}$ by using the partial derivative bounds from the previous two sections. This section deals with the following issue. There are many dependencies in the elements of $\mathbf{M}^{k'}$. If we only consider one element, it is clear that it will have a conditional min-entropy of $\Omega(\log n)$. However, if we put N elements together it is not clear at all that the min-entropy of them will be much larger than one element. To that end, we manage to find $\lceil\sqrt{N}\rceil$ elements of $\mathbf{M}^k$ such that they affect corresponding $\lceil\sqrt{N}\rceil$ out of N elements of $\mathbf{M}^{k'}$ orthogonally. Here, by "orthogonally" we mean fixing all the other $n - \lceil\sqrt{N}\rceil$ elements of $\mathbf{M}^k$, each of the $\lceil\sqrt{N}\rceil$ elements of $\mathbf{M}^{k'}$ is dominated by its corresponding element in $\mathbf{M}^k$. In this way, we are able to reduce the conditional min-entropy of the N output elements to the uncertainty of the $\lceil\sqrt{N}\rceil$ elements of $\mathbf{M}^k$.

For given N elements of $\mathbf{M}^{k'}$: $\mathbf{M}^{k'}[u_1, v_1], \ldots, \mathbf{M}^{k'}[u_N, v_N]$, after fixing any q elements of $\mathbf{M}^k$ we can find $N_{\text{core}} = \lceil\sqrt{N}\rceil$ elements (we call them core elements) from the unfixed ones of $\mathbf{M}^k$ and correspondingly N_{core} elements of $\mathbf{M}^{k'}$ (we call them co-core elements), with the following notation:

- the N_{core} many *core elements* in $\mathbf{M}^k$ are denoted by $\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]$
- the N_{core} many *co-core elements* elements in $\mathbf{M}^{k'}$ are denoted by $\mathbf{M}^{k'}[u'_1, v'_1], \ldots, \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]$

and the following holds true

- any large perturbation (we will define it later) in $\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]$ will cause observable change (larger than the accuracy) in $\mathbf{M}^{k'}[u'_1, v'_1], \ldots, \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]$

The above informal statement is the key to lower bound the min-entropy. There are many ways to perturb $\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]$, but only when the perturbation is very small can the value of $\mathbf{M}^{k'}[u'_1, v'_1], \ldots, \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]$ be approximately the same. Therefore, the probability that $\mathbf{M}^{k'}[u'_1, v'_1], \ldots, \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]$ approximately equal to some fixed values can never be too large.

Here let us introduce some notations. We write $||\cdot||_\infty$ or simply $||\cdot||$ to indicate the entry-wise max norm. For every vector $\mathbf{w} \in \mathbb{R}^d$, define $\mathcal{Q}_d(\mathbf{w}, r)$ as the hypercube $\mathcal{Q}_d(\mathbf{w}, r) \stackrel{\text{def}}{=} \left\{ \mathbf{w}' \in \mathbb{R}^d \middle| \, ||\mathbf{w}' - \mathbf{w}||_\infty \le r \right\}$. So we rewrite the events in a more manageable form.

Let σ be the partial assignment, i.e., the queried (fixed) elements of $\mathbf{M}^k$ with their values. Note that the query is given with input precision α. So if we denote fixed $= ([i_1, j_1], \ldots, [i_q, j_q])$ as the queried indices[18] and $(\mathbf{M}^k)_{\text{fixed}} = (\mathbf{M}^k[i_1, j_1], \ldots, \mathbf{M}^k[i_q, j_q])$ as the vector representation of the queried elements, then σ can be rewritten as the event $(\mathbf{M}^k)_{\text{fixed}} \in \mathcal{Q}_q(\mathbf{w}_{\text{fixed}}, \frac{1}{2n^\alpha})$ for some vector

[18]Note that the vector *fixed* here denotes the indices of the elements of $[\mathbf{M}^k]_\alpha$ that has been fixed by the queries made to the input, whereas f_{fix} denotes the matrix power function whose rounded output has been fixed.

$\mathbf{w}_{\text{fixed}}$. Then by the definition of the restricted conditional min-entropy, the probability we want to bound can be rewritten as

$$\Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = (y_1, \ldots, y_N), \mathbf{M} \in \mathcal{E} \,\middle|\, (\mathbf{M}^k)_{\text{fixed}} \in \mathcal{Q}_q\left(\mathbf{w}_{\text{fixed}}, \frac{1}{2n^\alpha}\right)\right]$$

for every vectors *out*, *fixed*, $\mathbf{w}_{\text{fixed}}$, and $(y_1, ..., y_N)$.

Given the vector *fixed*, we will show how to choose the core elements and the co-core elements. By the standard averaging argument, we further reduce the above probability to the probability conditioning on a partial assignment on all but the core elements instead of σ on $q = \sqrt{n}$ elements. We use $(\mathbf{M}^k)_{\overline{\text{core}}}$ to denote the $N_{\overline{\text{core}}} \stackrel{\text{def}}{=} n^2 - N_{\text{core}}$ elements other than the core elements. And we use $\mathbf{w}_{\overline{\text{core}}}$ to denote a vector of $N_{\overline{\text{core}}}$ values.

$$\begin{aligned}
&\Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = (y_1, \ldots, y_N), \mathbf{M} \in \mathcal{E} \,\middle|\, (\mathbf{M}^k)_{\text{fixed}} \in \mathcal{Q}_q\left(\mathbf{w}_{\text{fixed}}, \frac{1}{2n^\alpha}\right)\right] \\
= &\sum_{\mathbf{w}_{\overline{\text{core}}} \text{ consistent with } \mathbf{w}_{\text{fixed}}} \Pr\left[\left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right) \,\middle|\, (\mathbf{M}^k)_{\text{fixed}} \in \mathcal{Q}_q\left(\mathbf{w}_{\text{fixed}}, \frac{1}{2n^\alpha}\right)\right] \times \\
&\Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = (y_1, \ldots, y_N), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\le &\max_{\mathbf{w}_{\overline{\text{core}}}} \Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = (y_1, \ldots, y_N), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]
\end{aligned}$$

We will show by Theorem 5 that, even after we have $N_{\overline{\text{core}}}$ many elements fixed, the min-entropy of the output is still high.

Now we show how we choose the core elements in $\mathbf{M}^k$, and the corresponding N_{core} co-core elements in $\mathbf{M}^{k'}$. This builds upon the partial derivatives developed so far. Observe that an element of $\mathbf{M}^k$ affects more heavily the elements of $\mathbf{M}^{k'}$ on the same row or the same column. But two or more elements of $\mathbf{M}^k$ may cancel with each other when we consider their effect on $\mathbf{M}^{k'}$. However, we show below that we can find $N_{\text{core}} = \lceil\sqrt{N}\rceil$ many elements of $\mathbf{M}^k$ that are not fixed by the previous $q = \sqrt{n}$ queries, and they affect "orthogonally" to the output.

Lemma 12. *Let $\gamma \ge 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every $N = O(\text{polylog}(n))$, and every $1 \le k < k' \le n$ such that $k|k'$. Let $q = \sqrt{n}$, $N_{\text{core}} = \lceil\sqrt{N}\rceil$. For every* $\text{out} = ([u_t, v_t])_{t=1}^N$ *output indices of $\mathbf{M}^{k'}$ and for every q indices* $\text{fixed} = ([i_t, j_t])_{t=1}^q$ *of $\mathbf{M}^k$, there exists N_{core} core elements $\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]$ where $[i'_1, j'_1], ..., [i'_{N_{\text{core}}}, j'_{N_{\text{core}}}] \notin$ fixed, and corresponding N_{core} co-core elements* $\text{core} = ([u'_t, v'_t])_{t=1}^{N_{\text{core}}}$ *of $\mathbf{M}^{k'}$ where each $[u'_t, v'_t] \in$ out such that*

$$\begin{pmatrix}
\frac{\partial \mathbf{M}^{k'}[u'_1, v'_1]}{\partial \mathbf{M}^k[i'_1, j'_1]} & \cdots & \cdots & \frac{\partial \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]}{\partial \mathbf{M}^k[i'_1, j'_1]} \\
\vdots & \ddots & & \vdots \\
\vdots & & \ddots & \vdots \\
\frac{\partial \mathbf{M}^{k'}[u'_1, v'_1]}{\partial \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]} & \cdots & \cdots & \frac{\partial \mathbf{M}^{k'}[u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]}{\partial \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]}
\end{pmatrix} = \begin{pmatrix}
\Omega(\frac{k'^2}{kn^4}) & O(\frac{k'^3}{kn^8}) & \ldots & O(\frac{k'^3}{kn^8}) \\
O(\frac{k'^3}{kn^8}) & \Omega(\frac{k'^2}{kn^4}) & & \vdots \\
\vdots & & \ddots & \vdots \\
O(\frac{k'^3}{kn^8}) & \ldots & \ldots & \Omega(\frac{k'^2}{kn^4})
\end{pmatrix}$$

We observe that when $k' \le n$ this matrix is diagonally dominant.

Proof. By Lemma 11, the core elements and the co-core elements have a Jacobian matrix of the above form if and only if (i) for every $t = 1, \ldots, N_{\text{core}}$, $u'_t = i'_t$ or $v'_t = i'_t$, and (ii) for $t' \neq t$, $u'_{t'} \neq i'_t$ and $v'_{t'} \neq i'_t$. Fix arbitrary co-core elements $[u'_1, v'_1], ..., [u'_{N_{\text{core}}}, v'_{N_{\text{core}}}]$, notice the following fact:

- When $u'_1 = u'_2 = \cdots = u'_{N_{\text{core}}}$, we just let each $j'_t = v'_t$ and each i'_t to be such that (i) $[i'_t, j'_t] \notin$ fixed (ii) $i'_t \neq u'_1$ (iii) for each $t' \neq t$, $i'_t \neq i'_{t'}$. There always exist such core elements because there are at least $n - q = n - \sqrt{n}$ many rows of $\mathbf{M}^k$ that have no fixed elements. This way we get the desired core and co-core elements.

- If $u'_t \neq u'_{t'}$ for every pair of $t \neq t'$, then we can also find core elements. We let each $i'_t = u'_t$. Then we find a column v of $\mathbf{M}^k$ that has no fixed elements, and let $v'_t = v$ for every $t = 1, ..., N_{\text{core}}$.

Since we are selecting the co-core elements from $N \geq (N_{\text{core}} - 1)^2 + 1$ output elements, one of the above cases must occur. Because otherwise we will have less than N_{core} different rows, and each row has less than N_{core} different elements, no more than $(N_{\text{core}} - 1)^2$ different elements in total. Therefore, the desired core and co-core elements always exist. □

Now we are ready to give the main result of this section. Note that the hard instance set $\mathcal{E}$ will not be specified below until the next section.

Theorem 5 (main result for the perturbation analysis part)**.** *For every $N = O(\text{polylog}(n))$, and every $1 \leq k < k' \leq n$ such that $k|k'$. Let $N_{\text{core}} = \lceil \sqrt{N} \rceil$, $N_{\overline{\text{core}}} = n^2 - N_{\text{core}}$, $q = \sqrt{n}$. Let $\tau > \alpha$, $2 \leq \gamma < \alpha - 12.2$, $\eta = \alpha - 7.1$ be constants. Fix $\text{out} = ([u_t, v_t])_{t=1}^N$ to be arbitrary N indices of $\mathbf{M}^{k'}$, and $\text{fixed} = ([i_t, j_t])_{t=1}^q$ to be the indices of $\mathbf{M}^k$ that has been queried and hence fixed. Let $(\mathbf{M}^k)_{\text{core}} = (\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}])$ be the core elements so the corresponding Jacobian matrix is of the form in Lemma 12. $(\mathbf{M}^k)_{\overline{core}}$ denotes the remaining $N_{\overline{\text{core}}}$ elements of $\mathbf{M}^k$. For sufficiently large n, for every $Y_{\text{out}} \in \mathbb{R}^N, \mathbf{w}_{\overline{core}} \in \mathbb{R}^{N_{\overline{\text{core}}}}$, there exists $\mathbf{w}_{\text{core}} \in \mathbb{R}^{N_{\text{core}}}$ such that*

$$\begin{aligned}
&\Pr_{\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}} \left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = Y_{\text{out}}, \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{core}} \in \mathcal{Q}_{N_{\overline{\text{core}}}} \left(\mathbf{w}_{\overline{core}}, \frac{1}{2n^\alpha} \right) \right] \\
\leq &\Pr_{\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}} \left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}} \left(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta} \right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{core}} \in \mathcal{Q}_{N_{\overline{\text{core}}}} \left(\mathbf{w}_{\overline{core}}, \frac{1}{2n^\alpha} \right) \right]
\end{aligned}$$

To prove Theorem 5 we introduce the following lemma. This will also formalize what we mean by "large perturbation". Here we use $\text{Supp}(\mathcal{M}_{\gamma,\tau})$ to denote the support of the distribution $\mathcal{M}_{\gamma,\tau}$.

Lemma 13 (perturbation)**.** *For every $N = O(\text{polylog}(n))$, and every $1 \leq k < k' \leq n$ such that $k|k'$. Let $N_{\text{core}} = \lceil \sqrt{N} \rceil$, $N_{\overline{\text{core}}} = n^2 - N_{\text{core}}$. Let $\tau > \alpha$, $2 \leq \gamma < \alpha - 12.2$, $\eta = \alpha - 7.1$ be constants. Fix the indices of the core elements $([i'_t, j'_t])_{t=1}^{N_{\text{core}}}$ of $\mathbf{M}^k$, and $([u'_t, v'_t])_{t=1}^{N_{\text{core}}}$ the indices of the co-core elements of $\mathbf{M}^{k'}$, so the corresponding Jacobian matrix is of the form in Lemma 12. For sufficiently large n, and for two arbitrary assignments to the core elements $(\mathbf{M}_1^k)_{\text{core}} = (\mathbf{M}_1^k[i'_1, j'_1], \ldots, \mathbf{M}_1^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}])$ and $(\mathbf{M}_2^k)_{\text{core}} = (\mathbf{M}_2^k[i'_1, j'_1], \ldots, \mathbf{M}_2^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}])$, corresponding to the power of two matrices $\mathbf{M}_1^k$ and $\mathbf{M}_2^k$, if the rest $N_{\overline{\text{core}}}$ elements $[\mathbf{M}_1^k]_{\overline{core}}$ and $[\mathbf{M}_2^k]_{\overline{core}}$ fall in $\mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\overline{core}}, \frac{1}{2n^\alpha})$ for arbitrary $\mathbf{w}_{\overline{core}} \in \mathbb{R}^{n^2 - N_{\text{core}}}$ such that (i) $\mathbf{M}_1, \mathbf{M}_2 \in \text{Supp}(\mathcal{M}_{\gamma,\tau})$, and (ii)*

$$\max_{t=1,\ldots,N_{\text{core}}} |\mathbf{M}_1^k[i'_t, j'_t] - \mathbf{M}_2^k[i'_t, j'_t]| \geq \frac{1}{2n^{\eta+0.05}}$$

then

$$\max_{t=1,\ldots,N_{\text{core}}} |\mathbf{M}_1^{k'}[u'_t, v'_t] - \mathbf{M}_2^{k'}[u'_t, v'_t]| \geq \frac{1}{n^\alpha}$$

The main idea is that if the Jacobian matrix between the core elements and the co-core elements is diagonally dominant, then when the p-th element of $\mathbf{M}^k[i_1, j_1], \ldots, \mathbf{M}^k[i_{N_{\text{core}}}, j_{N_{\text{core}}}]$ has the maximum change over the rest, we should be able to observe the change in the corresponding p-th element of $\mathbf{M}^{k'}[u_1, v_1], \ldots, \mathbf{M}^{k'}[u_{N_{\text{core}}}, v_{N_{\text{core}}}]$.

Proof. The proof is similar to Corollary 8. We consider matrices as vectors of dimension n^2. For simplicity, we write the elements $\mathbf{M}^k[i'_1, j'_1], \ldots, \mathbf{M}^k[i'_{N_{\text{core}}}, j'_{N_{\text{core}}}]$ as $x_1, \ldots, x_{N_{\text{core}}}$ and the rest elements as $y_{N_{\text{core}}+1}, \ldots, y_{n^2}$. We arrange the order of the elements such that the core elements come first and write $\mathbf{M}_1^k = (x_1, \ldots, x_{N_{\text{core}}}, y_{N_{\text{core}}+1}, \ldots, y_{n^2})$, $\mathbf{M}_2^k = (x'_1, \ldots, x'_{N_{\text{core}}}, y'_{N_{\text{core}}+1}, \ldots, y'_{n^2})$.

Let $t = \arg\max_{i \in [N_{\text{core}}]} |x_i - x'_i|$. Then we have $\sum_{i \neq t}^{N_{\text{core}}} |x_i - x'_i| \leq N_{\text{core}} |x_t - x'_t|$ and $|x_t - x'_t| \geq \frac{1}{n^{\eta+0.05}}$. Moreover, for every $N_{\text{core}} < j \leq n^2$, $|y_j - y'_j| \leq \frac{1}{n^\alpha}$.

It is much more convenient if we rewrite each $\mathbf{M}^{k'}[u'_t, v'_t]$ as a function f_t of $\mathbf{M}^k$. We have

$$\begin{aligned}
|\mathbf{M}_1^{k'}[u'_t, v'_t] - \mathbf{M}_2^{k'}[u'_t, v'_t]| &= |f_t(x_1, \ldots, x_{N_{\text{core}}}, y_{N_{\text{core}}+1}, \ldots, y_{n^2}) - f_t(x'_1, \ldots, x'_{N_{\text{core}}}, y'_{N_{\text{core}}+1}, \ldots, y'_{n^2})| \\
&= \left| \sum_{t'=1}^{N_{\text{core}}} X_{t'} + \sum_{t'=N_{\text{core}}+1}^{n^2} Y_{t'} \right| \\
&\geq |X_t| - \sum_{t' \neq t}^{N_{\text{core}}} |X_{t'}| - \sum_{t'=N_{\text{core}}+1}^{n^2} |Y_{t'}|
\end{aligned}$$

where

$$X_{t'} = f_t(x'_1, ..., x'_{t'-1}, x_{t'}, x_{t'+1}, ..., x_{N_{\text{core}}}, y_{N_{\text{core}}+1}, ..., y_{n^2}) - f_t(x'_1, ..., x'_{t'-1}, x'_{t'}, x_{t'+1}, ..., x_{N_{\text{core}}}, y_{N_{\text{core}}+1}, ..., y_{n^2})$$

$$Y_{t'} = f_t(x'_1, ..., x'_{N_{\text{core}}}, y'_{N_{\text{core}}+1}, ..., y'_{t'-1}, y_{t'}, y_{t'+1}, ..., y_{n^2}) - f_t(x'_1, ..., x'_{N_{\text{core}}}, y'_{N_{\text{core}}+1}, ..., y'_{t'-1}, y'_{t'}, y_{t'+1}, ..., y_{n^2})$$

Here the difference is divided into the difference of the n^2 dimensions. We can first use integral to rewrite each $X_{t'}$ and $Y_{t'}$, and then bound it by using the partial derivation results we established above.

$$|X_{t'}| = \left| \int_{x'_{t'}}^{x_{t'}} \frac{\partial f_t}{\partial \mathbf{M}^k[i'_{t'}, j'_{t'}]} d\mathbf{M}^k[i'_{t'}, j'_{t'}] \right|$$

Note that $\frac{\partial f_t}{\partial \mathbf{M}^k[i'_{t'}, j'_{t'}]}$ is always positive, so we have the following bound:

$$|\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]| \min_{\mathbf{M}} \frac{\partial f_t}{\partial \mathbf{M}^k[i_{t'}, j_{t'}]} \leq |X_{t'}| \leq |\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]| \max_{\mathbf{M}} \frac{\partial f_t}{\partial \mathbf{M}^k[i_{t'}, j_{t'}]}$$

By Lemma 12, there exists some constants c_1, c_2, c_3 that for term t

$$\frac{c_1 k'^2}{k n^4} |\mathbf{M}_1^k[i'_t, j'_t] - \mathbf{M}_2^k[i'_t, j'_t]| \leq |X_t|$$

and for $t' \neq t$

$$\frac{c_3 k'^3}{k n^8} |\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]| \leq |X_{t'}| \leq \frac{c_2 k'^3}{k n^8} |\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]|$$

The same argument also applies to $Y_{t'}$. Actually we only need $|Y_{t'}| \le c_4 \cdot \frac{k'}{k}|\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]|$, for some $c_4 > 0$, by Lemma 11.

Thus, we obtain

$$\begin{aligned}
|\mathbf{M}_1^{k'}[i'_t, j'_t] - \mathbf{M}_2^{k'}[i'_t, j_t]'|| \ge & |X_t| - \sum_{t' \ne t}^{N_{\text{core}}} |X_{t'}| - \sum_{t'=N_{\text{core}}+1}^{n^2} |Y_{t'}| \\
\ge & \frac{1}{k}\left(\frac{c_1 k'^2}{n^4}|\mathbf{M}_1^k[i'_t, j'_t] - \mathbf{M}_2^k[i'_t, j'_t]| - \sum_{t' \ne t}^{N_{\text{core}}} \frac{c_2 k'^3}{n^8}|\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]| \right. \\
& \left. - \sum_{t'=N_{\text{core}}+1}^{n^2} c_4 k' |\mathbf{M}_1^k[i'_{t'}, j'_{t'}] - \mathbf{M}_2^k[i'_{t'}, j'_{t'}]| \right) \\
\ge & \frac{1}{k}\left(|x_t - x'_t| \left(\frac{c_1 k'^2}{n^4} - \sum_{t' \ne t}^{N_{\text{core}}} \frac{c_2 k'^3}{n^8} \right) - (n^2 - N_{\text{core}}) \frac{c_4 k'}{n^\alpha} \right) \\
\ge & \frac{1}{k}\left(\frac{c_1 k'^2}{2n^{\eta+4.05}} - \frac{c_2 k'^3 N_{\text{core}}}{2n^{\eta+8.05}} - \frac{c_4 k'(n^2 - N_{\text{core}})}{n^\alpha} \right) \\
\ge & \frac{c_1 k'^2}{2kn^{\alpha-3.05}} - \frac{c_2 k'^3 N_{\text{core}}}{2kn^{\alpha+0.95}} - \frac{c_4 k'}{kn^{\alpha-2}} \\
\ge & \frac{c_1}{2n^{\alpha-3.05}} - \frac{c_2 N_{\text{core}}}{2n^{\alpha+0.95}} - \frac{c_4}{n^{\alpha-2}} \ge \frac{1}{n^\alpha}
\end{aligned}$$

The last inequalities follow from that $\eta = \alpha - 7.1$ and $1 \le \frac{k'}{k}, k \le n$. Therefore, for sufficiently large n, $|\mathbf{M}_1^{k'}[i'_t, j'_t] - \mathbf{M}_2^{k'}[i'_t, j'_t]| \ge \frac{1}{n^\alpha}$. □

Now, we can conclude the proof of Theorem 5.

Proof of Theorem 5 by Lemma 13. We select the indices of co-core elements $\text{core}' = ([u'_t, v'_t])_{t=1}^{N_{\text{core}}}$ and the indices of core elements $\text{core} = ([i'_t, j'_t])_{t=1}^{N_{\text{core}}}$ as in Lemma 13.

For every $\mathbf{w} \in \mathbb{R}^{N_{\text{core}}}$ such that

$$\Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = \mathbf{w}, \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha} \right) \right] > 0$$

there must be a feasible $\mathbf{M}_0 \in \text{Supp}(\mathcal{M}_{\gamma,\tau})$ such that $[(\mathbf{M}_0^{k'})_{\text{out}}]_\alpha = \mathbf{w}$ approximates $\mathbf{M}_0^{k'}$, and above condition in the probability is satisfied. Let $\mathbf{w}_{\text{core}} = (\mathbf{M}_0^k)_{\text{core}}$. We will show that conditioning on $(\mathbf{M}^k)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha})$, the event that $(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = \mathbf{w}$, $\mathbf{M} \in \mathcal{E}$ occurs only if $(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta})$.

By the definition of the fixed-valued function f_{fix}, $(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = \mathbf{w}$ only if there exists another matrix $\mathbf{M}'$ where $[\mathbf{M}'^k]_\alpha = [\mathbf{M}^k]_\alpha$ and $[(\mathbf{M}'^{k'})_{\text{out}}]_\alpha = \mathbf{w}$. This happens only if $(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}((\mathbf{M}'^k)_{\text{core}}, \frac{1}{n^\alpha})$. Here $\mathbf{M}'$ might be or might not be the same as $\mathbf{M}_0$. But by our observation from Lemma 13, we know that the two matrices must have low max-norm distance.

Specifically, if $(\mathbf{M}'^k)_{\text{core}} \notin \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^{\eta+0.05}})$, then by Lemma 13 we can get that there exists $t \in [N_{\text{core}}]$ such that for sufficiently large n, $|\mathbf{M}'^{k'}[u'_t, v'_t] - \mathbf{M}_0^{k'}[u'_t, v'_t]| \ge \frac{1}{n^\alpha}$. A contradiction to the fact that $[(\mathbf{M}_0^{k'})_{\text{out}}]_\alpha = [(\mathbf{M}'^{k'})_{\text{out}}]_\alpha = \mathbf{w}$. This way, $||\mathbf{w}_{\text{core}} - (\mathbf{M}^k)_{\text{core}}|| \le ||\mathbf{w}_{\text{core}} - (\mathbf{M}'^k)_{\text{core}}|| + ||(\mathbf{M}'^k)_{\text{core}} - (\mathbf{M}^k)_{\text{core}}|| \le \frac{1}{2n^{\eta+0.05}} + \frac{1}{n^\alpha} \le \frac{1}{2n^\eta}$ is a necessary condition for $(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = \mathbf{w}$.

Formally, we have

$$\begin{aligned}
&\Pr\left[(f_{\text{fix}}(\mathbf{M}))_{\text{out}} = \mathbf{w}, \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\leq &\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left((\mathbf{M}'^k)_{\text{core}}, \frac{1}{n^\alpha}\right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\leq &\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^{\eta+0.05}} + \frac{1}{n^\alpha}), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\leq &\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^{\eta}}), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]
\end{aligned}$$

Because η is a smaller constant than α.

□

In this way, we get rid of f_{fix} and $\mathbf{M}^{k'}$. The lower bound reduces to upper-bounding the above conditional probability.

9 Conditional probability bounds from volume and embeddings

The previous Theorem 5 reduces the problem of lower-bounding the restricted conditional min-entropy to the problem of upper-bounding the following probability:

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]$$

In this section, we will show that for every $\mathbf{w}_{\text{core}} \in \mathbb{R}^{N_{\text{core}}}, \mathbf{w}_{\overline{\text{core}}} \in \mathbb{R}^{n^2-N_{\text{core}}}$,

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \leq 55 \cdot n^{(\gamma-\eta+3.1)N_{\text{core}}}$$

in which γ, η are constants such that $\gamma - \eta + 3.1 < -2$ and $\eta = \alpha - 7.1$. Recall that $\mathbf{w}_{\text{core}}$ and $\mathbf{w}_{\overline{\text{core}}}$ are respectively assignments to the core elements and the remaining $n^2 - N_{\text{core}}$ elements.

We show this in three steps: (i) introduce the hard instance set $\mathcal{E}$. (ii) rewrite the above conditional probability as the ratio of unconditional probabilities (i.e., $\Pr[A|B] = \frac{\Pr[A,B]}{\Pr[B]}$). Moreover, we partition the events into small hypercubes of the same size, in the probability space. (iii) by Lipschitz analysis, we show that the metric space of the matrix power $\mathbf{M}^k = (\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M}_{\mathbf{U}})^k$ has low distortion from the space of $\mathbf{M}_{\mathbf{U}}$. Since $\mathbf{M}_{\mathbf{U}}$ follows a uniform distribution, we can bound the probability measure of the small hypercubes. Given the upper bound and the lower bound of the hypercubes, we just need to count the number of the small hypercubes in the denominator and numerator.

We begin with the definition of $\mathcal{E}$.[19]

Definition 4 (construction of $\mathcal{E}$)**.** *Let $\mathcal{E}_U$ be the set of matrices $\mathbf{M}_{\mathbf{U}}$ such that (i) each element of $\mathbf{M}_{\mathbf{U}}$ is in the interval $[\frac{1}{n} - \frac{1}{n^\gamma} + \frac{1}{n^{\gamma+3}}, \frac{1}{n} - \frac{1}{n^{\gamma+3}}]$. (ii) each element of $\mathbf{M}_{\mathbf{U}}$ is a multiple of $\frac{1}{n^\tau}$. Define*

$$\mathcal{E} = \left\{\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M}_{\mathbf{U}} : \mathbf{M}_{\mathbf{U}} \in \mathcal{E}_U\right\}$$

[19] We are very careful on the following definitions and decompositions to avoid boundary problems in lower bounding and upper bounding the volumes of the hypercubes.

We use $\mathcal{E}$ to consider all the $\mathbf{M}$ that are at a distance from the boundary of $\mathcal{M}$, so the small hypercubes centered at $\mathbf{M}^k$ have preimages always contained inside the boundary. To formalize what we referred to as the boundary, we consider a cover of the matrix space of $\mathbf{M}$ and $\mathbf{M}^k$:

Define $\phi_k(\mathbf{M_U}) \stackrel{\text{def}}{=} \mathbf{M}^k = (\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U})^k$. Let $\mathbf{M}^k_{\max}$ be the matrix $\mathbf{M}^k_{\max} = \phi_k(\mathbf{M_U})$ where every element of $\mathbf{M_U}$ is $\frac{1}{n}$. Let $\mathbf{M}^k_{\min}$ be the matrix $\mathbf{M}^k_{\min} = \phi_k(\mathbf{M_U}')$ where every element of $\mathbf{M_U}'$ is $\frac{1}{n} - \frac{1}{n^\gamma}$. Then for each elements $(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\}$,

$$\mathbf{M}^k_{\min}[i,j] \le \mathbf{M}^k[i,j] \le \mathbf{M}^k_{\max}[i,j]$$

Similarly, let $\mathbf{M}^k_{\mathcal{E}\max}$ and $\mathbf{M}^k_{\mathcal{E}\min}$ be the corresponding matrix powers where $\mathbf{M_U} \in \mathcal{E}$. Specifically, let $\mathbf{M}^k_{\mathcal{E}\max} = \phi_k(\mathbf{M_U}'')$ where every element of $\mathbf{M_U}''$ is $\frac{1}{n} - \frac{1}{n^{\gamma+3}}$, and $\mathbf{M}^k_{\mathcal{E}\min} = \phi_k(\mathbf{M_U}''')$ where every element of $\mathbf{M_U}'''$ is $\frac{1}{n} - \frac{1}{n^\gamma} + \frac{1}{n^{\gamma+3}}$.

9.1 Geometric decomposition

In this subsection we decompose the conditional probability into more tractable hypercubes in the probability space of $\mathbf{M}^k$. This step is intuitive from a geometric view. When we rewrite the conditional probability

$$\begin{aligned}
&\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
=&\frac{\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}), \mathbf{M} \in \mathcal{E}, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]}{\Pr_{\mathbf{M}\leftarrow\mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]}
\end{aligned}$$

both the numerator and the denominator are exactly the volume of two hyperrectangles in the probability space of $\mathbf{M}^k$. The decomposition would also trivially be cutting the hyperrectangles along each dimension.

We will first decompose the denominator into hyperrectangles of the same size as the numerator.

Definition 5 (grid point)**.** *For every $\mathbf{w} \in \mathbb{R}^{N_{\text{core}}}$, we call a vector $\mathbf{w}_{\text{grid}} \in \mathbb{R}^{N_{\text{core}}}$ a* grid point *with respect to $\mathbf{w}$ if (i) each element of $\mathbf{w} - \mathbf{w}_{\text{grid}}$ is a multiple of $\frac{1}{n^\eta}$ (ii) for $l = 1,\dots,N_{\text{core}}$, $\mathbf{M}^k_{\min}[i_l, j_l] + \frac{1}{2n^\eta} \le \mathbf{w}_{\text{grid}}[l] \le \mathbf{M}^k_{\max}[i_l, j_l] - \frac{1}{2n^\eta}$, where (i_l, j_l) are the core elements.*

We note that the grid points will all be with respect to $\mathbf{w}_{\text{core}}$. For simplicity, we sometimes just call the grid points "$\mathbf{w}_{\text{grid}}$" instead of "$\mathbf{w}_{\text{grid}}$ with respect to $\mathbf{w}_{\text{core}}$".

We use grid points to partition the space of $(\mathbf{M}^k)_{\text{core}}$ into hypercubes of edge length $\frac{1}{n^\eta}$ centered at grid points

$$\mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{grid}}, \frac{1}{2n^\eta}\right)$$

Note that the hypercubes intersect at boundaries, which brings non-negligible discrete probabilities in counting. To that end, we instead use $\mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{grid}}, \frac{1}{2n^\eta+1}\right)$ in giving lower bounds to the volume of the hypercubes.

Given the partition above, the probability becomes

$$\begin{aligned}
&\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}\right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\le&\frac{\Pr[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}), \mathbf{M} \in \mathcal{E}, (\mathbf{M}^k)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha})]}{\sum_{\mathbf{w}_{\text{grid}}} \Pr[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{grid}}, \frac{1}{2n^\eta+1}), (\mathbf{M}^k)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha})]}
\end{aligned}$$

Now, we further decompose the regions $(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}(\mathbf{w}_{\text{grid}}, \frac{1}{2n^\eta+1})$, $(\mathbf{M}^k)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha})$ into many small hypercube of edge length $\frac{2}{2n^\alpha+1}$ (here the edge length is $\frac{2}{2n^\alpha+1}$ for the same reason as above), dimension n^2. The hypercubes are smaller since $\eta < \alpha$, and thus $\frac{1}{n^\eta} \gg \frac{1}{n^\alpha}$.

We achieve this decomposition by considering hypercubes with the following two types of centers respectively used for upper bounding the numerator and lower bounding the denominator:

Definition 6 (local center - global center)**.** *For every* $\mathbf{w}_{\text{core}} \in \mathbb{R}^{N_{\text{core}}}$,

We call a vector $\mathbf{w}_{\text{local}} \in \mathbb{R}^{n^2}$ local *with respect to* $(\mathbf{w}_{\text{core}}, \mathbf{w}_{\overline{\text{core}}})$*, if (i) Each element of* $\mathbf{w}_{\text{local}} - (\mathbf{w}_{\text{core}}, \mathbf{w}_{\overline{\text{core}}})$ *is a multiple of* $\frac{1}{n^\alpha}$*. (ii) each element of* $\mathbf{w}_{\text{local}}$ *satisfies:* $|\mathbf{w}_{\text{local}}[i,j] - \mathbf{w}_{\text{core}}[i,j]| < \frac{1}{2n^\eta} + \frac{1}{2n^\alpha}$ *for core elements* i, j *and* $|\mathbf{w}_{\text{local}}[i,j] - \mathbf{w}_{\overline{\text{core}}}[i,j]| \le \frac{1}{2n^\alpha}$ *for the remaining* $n^2 - N_{\text{core}}$ *elements.*

We call $\mathbf{w}_{\text{global}} \in \mathbb{R}^{n^2}$ global *with respect to* $(\mathbf{w}_{\text{core}}, \mathbf{w}_{\overline{\text{core}}})$ *if there exists some* $\mathbf{w}_{\text{grid}}$ *with respect to* $\mathbf{w}_{\text{core}}$ *such that (i) Each element of* $\mathbf{w}_{\text{global}} - (\mathbf{w}_{\text{grid}}, \mathbf{w}_{\overline{\text{core}}})$ *is a multiple of* $\frac{1}{n^\alpha}$*. (ii) each element of* $\mathbf{w}_{\text{global}}$ *satisfies:* $|\mathbf{w}_{\text{global}}[i,j] - \mathbf{w}_{\text{grid}}[i,j]| \le \frac{1}{2n^\eta+1} - \frac{1}{2n^\alpha}$ *for core elements* i, j *and* $|\mathbf{w}_{\text{global}}[i,j] - \mathbf{w}_{\overline{\text{core}}}[i,j]| \le \frac{1}{2n^\alpha}$ *for the remaining* $n^2 - N_{\text{core}}$ *elements.*

By the definition, all the $\mathbf{w}_{\text{global}}[i,j]$ and $\mathbf{w}_{\text{local}}[i,j]$ for non-core elements i, j should be equal to $\mathbf{w}_{\overline{\text{core}}}[i,j]$. Now, we can further decompose the probability as:

$$\begin{aligned}
&\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}\right), \mathbf{M} \in \mathcal{E}, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\le &\sum_{\mathbf{w}_{\text{local}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{local}}, \frac{1}{2n^\alpha}\right), \mathbf{M} \in \mathcal{E}\right]
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{\mathbf{w}_{\text{grid}}} \Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{grid}}, \frac{1}{2n^\eta+1}\right), \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\ge &\sum_{\mathbf{w}_{\text{global}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{global}}, \frac{1}{2n^\alpha+1}\right)\right]
\end{aligned}$$

So we can see that the probability measure of each hypercube $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})]$ and $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha+1})]$ for $\mathbf{w} \in \mathbb{R}^{n^2}$ inside the boundary is the key to our problem.

However, as we can see the probabilities of the hypercubes with local centers above still have $\mathcal{E}$ inside. We can remove it by further focusing on centers contained in $\mathcal{E}$.

Definition 7 (valid center)**.** *Fix* $\mathbf{w}_{\text{core}} \in \mathbb{R}^{N_{\text{core}}}$ *as the value of the core elements and* $\mathbf{w}_{\overline{\text{core}}} \in \mathbb{R}^{n^2 - N_{\text{core}}}$ *as the value of the rest of the elements. Denote* $\mathbf{w} = (\mathbf{w}_{\text{core}}, \mathbf{w}_{\overline{\text{core}}})$ *as the vector description of the whole (rounded) matrix.*

We call $\mathbf{w}_{\text{valid,local}} \in \mathbb{R}^{n^2}$ *a* valid local center *if (i)* $\mathbf{w}_{\text{valid,local}}$ *is local with respect to* $\mathbf{w}$ *(ii) for every* $(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\}$, $\mathbf{M}^k_{\mathcal{E}\min}[i,j] - \frac{1}{2n^\alpha} < \mathbf{w}_{\text{valid,local}}[i,j] < \mathbf{M}^k_{\mathcal{E}\max}[i,j] + \frac{1}{2n^\alpha}$.

Similarly, we call $\mathbf{w}_{\text{valid,global}} \in \mathbb{R}^{n^2}$ *a* valid global center *if (i)* $\mathbf{w}_{\text{valid,global}}$ *is global with respect to* $\mathbf{w}$ *(ii) for every* $(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\}$, $\mathbf{M}^k_{\mathcal{E}\min}[i,j] + \frac{1}{2n^\alpha} \le \mathbf{w}_{\text{valid,global}}[i,j] \le \mathbf{M}^k_{\mathcal{E}\max}[i,j] - \frac{1}{2n^\alpha}$.

Given the above definition, we can see that

$$\sum_{\mathbf{w}_{\text{local}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{local}}, \frac{1}{2n^\alpha}\right), \mathbf{M} \in \mathcal{E}\right]$$
$$\leq \sum_{\mathbf{w}_{\text{valid,local}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{valid,local}}, \frac{1}{2n^\alpha}\right)\right]$$

and

$$\sum_{\mathbf{w}_{\text{global}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{global}}, \frac{1}{2n^\alpha+1}\right)\right]$$
$$\geq \sum_{\mathbf{w}_{\text{valid,global}}} \Pr\left[\mathbf{M}^k \in \mathcal{Q}_{n^2}\left(\mathbf{w}_{\text{valid,global}}, \frac{1}{2n^\alpha+1}\right)\right]$$

Therefore,

$$\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}\right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right]$$
$$\leq \frac{\sum_{\mathbf{w}_{\text{valid,local}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,local}}, \frac{1}{2n^\alpha})]}{\sum_{\mathbf{w}_{\text{valid,global}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,global}}, \frac{1}{2n^\alpha+1})]}$$

9.2 Volume argument

Now we just need to (i) count the number of valid centers, and to (ii) upper bound $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})]$ and lower bound $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha+1})]$ for every valid $\mathbf{w}$. We focus on part (ii) in this section, which is also the technical part. We achieve this by equipping tools from Lipschitz analysis. More specifically, given $\phi_k(\mathbf{M_U}) = \mathbf{M}^k$, we show that the Lipschitz constant of ϕ_k and the inverse of the Lipchitz constant of ϕ_k^{-1} are close to each other, which means that the probability space of $\mathbf{M}^k$ has low distortion from the (uniform) space of $\mathbf{M_U}$. Formally, we will prove that each hypercube in the space of $\mathbf{M}^k$ has a preimage that is inside a small hypercube of the space of $\mathbf{M_U}$, and also contains a slightly smaller hypercube. In this way, we obtain the desired upper bound and lower bound of its volume.

To begin with, we show that ϕ_k is an injective function. This is not true in general. But in the special case where $\mathbf{M}$ is limited in a small range, it is true.

Lemma 14. *Given $\mathbf{M_U} \in \mathbb{R}^{n\times n}$ in which each element lies in range $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. Fix an integer $k \in [n]$. For sufficiently large n, $\phi_k = \mathbf{M}^k = (\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U})^k$ is an injective function.*

Proof. The proof goes similarly as in Lemma 8.

Suppose for sake of contradiction that ϕ_k is not an injective function, which means that there exists $\mathbf{M_U} \neq \mathbf{M_U}' \in [\frac{1}{n}-\frac{1}{n^\gamma}, \frac{1}{n}]^{n\times n}$ such that $\phi_k(\mathbf{M_U}) = \phi_k(\mathbf{M_U}')$. Let $(i,j) = \arg\max_{(i,j)} |\mathbf{M_U}[i,j] - \mathbf{M_U}'[i,j]|$.

For simplicity, let $x_{1,1}, ..., x_{n,n}$ be the elements of $\mathbf{M_U}$, and $y_{1,1}, ..., y_{n,n}$ be the elements of $\mathbf{M_U}'$. We rewrite $\phi_k(\mathbf{M_U})[i,j]$ as $f_{k,i,j}(x_{1,1}, ..., x_{n,n}) \stackrel{\text{def}}{=} (\phi_k(\mathbf{M_U} = (x_{1,1}, ..., x_{n,n})))[i,j]$ a function of n^2

elements. Then

$$\begin{aligned}
&|(\phi_k(\mathbf{M_U}) - \phi_k(\mathbf{M_U}'))[i,j]| \\
=&\left|\sum_{u,v\in[n]}\int_{x_{u,v}}^{y_{u,v}} \frac{\partial f_{k,i,j}(x_{1,1},...,x_{u,v-1},z,y_{u,v+1},...,y_{n,n})}{\partial z}dz\right| \\
\geq&\left|\frac{\Theta(k)}{n^3}\cdot(x_{i,j}-y_{i,j})\right| - (2n-2)\cdot\left|O\left(\frac{k^2}{n^7}\right)\cdot(x_{i,j}-y_{i,j})\right| \\
&-(n^2-2n+1)\cdot\left|O\left(\frac{k^3}{n^{11}}\right)\cdot(x_{i,j}-y_{i,j})\right| \\
\geq&\left(\Theta\left(\frac{k}{n^3}\right) - O\left(\frac{k^2}{n^6}\right) - O\left(\frac{k^3}{n^9}\right)\right)\cdot|x_{i,j}-y_{i,j}| \\
>&0
\end{aligned}$$

where the first inequality comes from Lemma 9 and the fact that $\frac{\partial \mathbf{M}[i,j]}{\partial \mathbf{M_U}[i,j]} = \frac{1}{n^3}$. Then $\phi_k(\mathbf{M_U}) \neq \phi_k(\mathbf{M_U}')$, a contradiction. □

Given that ϕ_k is injective, the inverse function ϕ_k^{-1} exists[20]. Recall that $||\mathbf{M}||$ is the entry-wise max norm of $\mathbf{M}$. For every function $\phi : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$, we define its *Lipschitz constant* as $||\phi||_{\text{Lip}} \stackrel{\text{def}}{=} \sup_{\mathbf{M_U}\neq\mathbf{M_U}'} \frac{||\phi(\mathbf{M_U})-\phi(\mathbf{M_U}')||}{||\mathbf{M_U}-\mathbf{M_U}'||}$. An invertible function ϕ is called a *bi-Lipschitz function* if it has bounded $||\phi||_{\text{Lip}}$ and $||\phi^{-1}||_{\text{Lip}}$.

Lemma 15. *For every $k \leq n$ and two matrices $\mathbf{M_U}, \mathbf{M_U}'$ in the space of $\mathbf{M_U}$. For every lare enough n,*

$$||\phi_k^{-1}||_{\text{Lip}}^{-1}||\mathbf{M_U}-\mathbf{M_U}'|| \leq ||\phi_k(\mathbf{M_U})-\phi_k(\mathbf{M_U}')|| \leq ||\phi_k||_{\text{Lip}}||\mathbf{M_U}-\mathbf{M_U}'||$$

where $||\phi_k^{-1}||_{\text{Lip}}^{-1} \geq ((\frac{n^2-1}{n^2})^{k-1}\frac{k}{n^3} - \frac{k^2}{n^6}), ||\phi_k||_{\text{Lip}} \leq ((\frac{n^2-1}{n^2})^{k-1}\frac{k}{n^3} + \frac{k^2}{n^6})$

Proof. The case $k = 1$ can be verified easily, where $||\phi_k(\mathbf{M_U}) - \phi_k(\mathbf{M_U}')|| = \frac{1}{n^3}||\mathbf{M_U} - \mathbf{M_U}'||$.

Assuming $k \geq 2$. We note that

$$\begin{aligned}
\phi_k(\mathbf{M_U}) - \phi_k(\mathbf{M_U}') &= \left(\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U}\right)^k - \left(\frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^3}\mathbf{M_U}'\right)^k \\
&= \sum_{k'=1}^{k}\binom{k}{k'}\left(\frac{n^2-1}{n^2}\right)^{k-k'}\frac{1}{n^{3k'}}(\mathbf{M_U}^{k'} - \mathbf{M_U}'^{k'})
\end{aligned}$$

By triangle inequality we have

$$||\phi_k(\mathbf{M_U})-\phi_k(\mathbf{M_U}')|| \leq ||k\left(\frac{n^2-1}{n^2}\right)^{k-1}\frac{1}{n^3}(\mathbf{M_U}-\mathbf{M_U}')|| + \sum_{k'=2}^{k}||\binom{k}{k'}\left(\frac{n^2-1}{n^2}\right)^{k-k'}\frac{1}{n^{3k'}}(\mathbf{M_U}^{k'}-\mathbf{M_U}'^{k'})||$$

and

$$||\phi_k(\mathbf{M_U})-\phi_k(\mathbf{M_U}')|| \geq ||k\left(\frac{n^2-1}{n^2}\right)^{k-1}\frac{1}{n^3}(\mathbf{M_U}-\mathbf{M_U}')|| - \sum_{k'=2}^{k}||\binom{k}{k'}\left(\frac{n^2-1}{n^2}\right)^{k-k'}\frac{1}{n^{3k'}}(\mathbf{M_U}^{k'}-\mathbf{M_U}'^{k'})||$$

[20]We ignore the matrices with no preimages for now.

By Corollary 8, we only need to bound

$$\sum_{k'=2}^{k}\left|\left|\binom{k}{k'}\left(\frac{n^2-1}{n^2}\right)^{k-k'}\frac{1}{n^{3k'}}(\mathbf{M_U}^{k'}-\mathbf{M_U}'^{k'})\right|\right| \le \sum_{k'=2}^{k}\binom{k}{k'}\left(\frac{n^2-1}{n^2}\right)^{k-k'}\frac{k'}{n^{3k'}}||\mathbf{M_U}-\mathbf{M_U}'||$$
$$= \frac{k}{n^3}\left(\left(\frac{n^2-1}{n^2}+\frac{1}{n^3}\right)^{k-1}-\left(\frac{n^2-1}{n^2}\right)^{k-1}\right)||\mathbf{M_U}-\mathbf{M_U}'||$$
$$\le \frac{k^2}{n^6}||\mathbf{M_U}-\mathbf{M_U}'||$$

Therefore,

$$\left(\left(\frac{n^2-1}{n^2}\right)^{k-1}\frac{k}{n^3}-\frac{k^2}{n^6}\right)||\mathbf{M_U}-\mathbf{M_U}'|| \le ||\phi_k(\mathbf{M_U})-\phi_k(\mathbf{M_U}')|| \le \left(\left(\frac{n^2-1}{n^2}\right)^{k-1}\frac{k}{n^3}+\frac{k^2}{n^6}\right)||\mathbf{M_U}-\mathbf{M_U}'||$$

□

Note that the above shows that each dimension of $\mathbf{M}^k[i,j]$ has a range of length $\Theta((\frac{n^2-1}{n^2})^{k-1}\cdot\frac{k}{n^3}\cdot\frac{1}{n^\gamma})$. Because the range of each dimension of $\mathbf{M_U}$ is $\frac{1}{n^\gamma}$.

For simplicity, we refer to $||\phi_k^{-1}||_{\text{Lip}}^{-1}$ as its lower bound $(\frac{n^2-1}{n^2})^{k-1}\frac{k}{n^3}-\frac{k^2}{n^6}$. And we denote $||\phi_k||_{\text{Lip}}$ as its upper bound $(\frac{n^2-1}{n^2})^{k-1}\frac{k}{n^3}+\frac{k^2}{n^6}$ (instead of its precise value). We have the following two corollaries which will be used later on.

Corollary 16. *For every elements $(i,j)\in\{1,\dots,n\}\times\{1,\dots,n\}$,*

$$||\phi_k^{-1}||_{\text{Lip}}^{-1}\left(\frac{1}{n^\gamma}-\frac{2}{n^{\gamma+3}}\right) \le |\mathbf{M}^k_{\mathcal{E}\,\max}[i,j]-\mathbf{M}^k_{\mathcal{E}\,\min}[i,j]| \le ||\phi_k||_{\text{Lip}}\left(\frac{1}{n^\gamma}-\frac{2}{n^{\gamma+3}}\right)$$

Corollary 17.

$$\frac{||\phi_k||_{\text{Lip}}}{||\phi_k^{-1}||_{\text{Lip}}^{-1}} \le 1+\frac{4}{n^2}$$

Given that ϕ_k is an invertible function, we define the preimage as follows.

Definition 8 (preimage)**.** *Let $\mathbf{w}\in\mathbb{R}^{n^2}$ be the vector representation of a matrix. We say $\mathbf{w}$ has a* preimage $\mathbf{w}_{\text{pre}}\in\mathbb{R}^{n^2}$*, if there exists a $\mathbf{M_U}$ such that $\mathbf{M_U}=\mathbf{w}_{\text{pre}}$ and $\phi_k(\mathbf{M_U})=\mathbf{w}$. Recall that our $\mathbf{M_U}$ has the following property: (i) each element is a multiple of $\frac{1}{n^\tau}$ (ii) each element falls in the interval $[\frac{1}{n}-\frac{1}{n^\gamma},\frac{1}{n}]$.*

With the fact that ϕ_k is invertible, we have:

$$\Pr\left[\mathbf{M}^k\in\mathcal{Q}_{n^2}\left(\mathbf{w},\frac{1}{2n^\alpha}\right)\right]=\Pr\left[\mathbf{M_U}\in\phi_k^{-1}\left(\mathcal{Q}_{n^2}\left(\mathbf{w},\frac{1}{2n^\alpha}\right)\right)\right]$$

and the same holds if we replace $\frac{1}{2n^\alpha}$ by $\frac{1}{2n^\alpha+1}$.

Now we just need to bound the above. Since $\mathbf{M_U}$ is a discrete uniform distribution, we can see that the probability mass of $\phi_k^{-1}\left(\mathcal{Q}_{n^2}(\mathbf{w},\frac{1}{2n^\alpha})\right)$ and $\phi_k^{-1}\left(\mathcal{Q}_{n^2}(\mathbf{w},\frac{1}{2n^\alpha+1})\right)$ are almost proportional to its volume. Our goal is to count and bount the discrete probability of the preimage of every hypercube. Recall that our input distribution $\mathcal{M}_{\gamma,\tau}$ is defined on the matrices $\mathbf{M_U}$ whose elements are multiples of $n^{-\tau}$.

Definition 9 (integer point)**.** $\mathbf{w}_{\text{int}} \in \mathbb{R}^{n^2}$ *is called an* integer point *if each element of* $\mathbf{w}_{\text{int}}$ *is a multiple of* $\frac{1}{n^\tau}$*.*

Definition 10 (boundary of $\mathcal{M}_{\gamma,\tau}$)**.** *Given the vector representation of a matrix* $\mathbf{w} \in \mathbb{R}^{n^2}$ *(in the space of* $\mathbf{M}^k$*). We say* $\mathbf{w}$ *is* inside the boundary of $\mathcal{M}_{\gamma,\tau}$ *in the space of* $\mathbf{M}^k$ *if for every* $i, j \in [n]$

$$\mathbf{M}^k_{\min}[i,j] \le \mathbf{w}[i,j] \le \mathbf{M}^k_{\max}[i,j]$$

Analogously, we say the $\mathbf{w}$ *is* inside the boundary of $\mathcal{M}_{\gamma,\tau}$ *in the space of* $\mathbf{M_U}$ *if for every* $i, j \in [n]$

$$\frac{1}{n} - \frac{1}{n^\gamma} \le \mathbf{w}[i,j] \le \frac{1}{n}$$

Besides, for every $r > 0$*, we say a set of points (e.g., a hypercube* $\mathcal{Q}_{n^2}(\mathbf{w}, r)$*) is* inside the boundary of $\mathcal{M}_{\gamma,\tau}$ *in the space of* $\mathbf{M}^k$ *(resp.* $\mathbf{M_U}$*) if every point in the set (e.g.,* $\mathbf{w}' \in \mathcal{Q}_{n^2}(\mathbf{w}, r)$*) is inside the boundary of* $\mathcal{M}_{\gamma,\tau}$ *in the space of* $\mathbf{M}^k$ *(resp.* $\mathbf{M_U}$*).*

With the definitions above, we can say that each instance of $\mathbf{M_U}$ is an integer point. Then, we have

Lemma 18. *For every* $\mathbf{w}$ *in* $\mathbb{R}^{n^2}$ *and* $r > n^{-\tau}$*, let* N_0 *be the number of integer points in* $\mathcal{Q}_{n^2}(\mathbf{w}, r)$*. If* $\mathcal{Q}_{n^2}(\mathbf{w}, r)$ *is inside the boundary of* $\mathcal{M}_{\gamma,\tau}$ *in the space of* $\mathbf{M_U}$*, then*

$$(2rn^\tau - 1)^{n^2} \le N_0 \le (2rn^\tau + 1)^{n^2}$$

Proof. We only need to consider one dimension. For an edge of length $2r$, it contains at most $2rn^\tau + 1$ integer points. On the other hand, it contains at least $\lfloor 2rn^\tau \rfloor > 2rn^\tau - 1$ integer points. The inequality follows by combining the n^2 dimensions. □

We have shown that the key point is to bound $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})]$ and $\Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha+1})]$ for valid centers $\mathbf{w}$. Here $\mathbf{w}$ should be regarded as an approximation of $\mathbf{M}^k$. If $\mathbf{w}$ has a preimage $\mathbf{w}'$ (should be regarded as the vector representation of $\mathbf{M_U}$), then $\forall \mathbf{M_U} \in \{\mathbf{M_U} | \phi_k(\mathbf{M_U}) \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})\}, ||\mathbf{M_U} - \mathbf{w}'|| \le \frac{1}{2n^\alpha \cdot ||\phi_k^{-1}||^{-1}_{\text{Lip}}}$. So we have $\{\mathbf{M_U} | \phi_k(\mathbf{M_U}) \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})\} \subseteq \mathcal{Q}_{n^2}(\mathbf{w}', \frac{1}{2n^\alpha \cdot ||\phi_k^{-1}||^{-1}_{\text{Lip}}})$. Similarly, $\forall \mathbf{M_U} \in \mathcal{Q}_{n^2}(\mathbf{w}', \frac{1}{||\phi_k||_{\text{Lip}}(2n^\alpha+1)}), ||\mathbf{M}^k - \mathbf{w}|| \le \frac{1}{2n^\alpha+1}$, namely $\mathcal{Q}_{n^2}(\mathbf{w}', \frac{1}{||\phi_k||_{\text{Lip}}(2n^\alpha+1)}) \subseteq \{\mathbf{M_U} | \phi_k(\mathbf{M_U}) \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha+1})\}$. So if the hypercubes centered at $\mathbf{w}'$ are inside the boundary of $\mathcal{M}_{\gamma,\tau}$ in the space of $\mathbf{M_U}$, we have:

$$|\{\mathbf{M_U} | \phi_k(\mathbf{M_U}) \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha})\}| \le \left(\frac{n^\tau}{n^\alpha \cdot ||\phi_k^{-1}||^{-1}_{\text{Lip}}} + 1\right)^{n^2}$$

and

$$|\{\mathbf{M_U} | \phi_k(\mathbf{M_U}) \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{1}{2n^\alpha+1})\}| \ge \left(\frac{2n^\tau}{||\phi_k||_{\text{Lip}}(2n^\alpha+1)} - 1\right)^{n^2}$$

By the set $\mathcal{E}$ we already let $\mathbf{M_U}$ to be not close to the boundary. We will verify in Lemma 20 that the hypercubes above are indeed inside the boundary of $\mathcal{M}_{\gamma,\tau}$.

However, before dealing with the boundary, a problem arises: if $\mathbf{w}$ does not have a preimage, then we can never apply this argument. But those $\mathbf{w}$ with preimages should be dense. Our strategy is to find a $\mathbf{w}$ that has a preimage nearby.

Lemma 19. *For every $2 \le k \le n$ and every valid $\mathbf{w} \in \mathbb{R}^{n^2}$, i.e., for every $(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\}$, $\mathbf{M}^k_{\mathcal{E}\,\min}[i,j] - \frac{1}{2n^\alpha} < \mathbf{w}[i,j] < \mathbf{M}^k_{\mathcal{E}\,\max}[i,j] + \frac{1}{2n^\alpha}$. There exists $\mathbf{w}'' \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{2k}{n^{\tau+3}})$ such that $\mathbf{w}''$ has a preimage, namely, there exists an $\mathbf{M}_{\mathbf{U}}$ such that $\phi_k(\mathbf{M}_{\mathbf{U}}) = \mathbf{w}''$.*

Proof. We prove it in a geometric way. For every integer point $\mathbf{v} = \left(\frac{k_{1,1}}{n^\tau}, \dots, \frac{k_{n,n}}{n^\tau}\right)$, where $k_{1,1}, ..., k_{n,n}$ are integers, and each element of $\mathbf{v}$ lies in the interval $\left[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}\right]$. We define $\mathcal{C}_{\mathbf{v}}$ to be the smallest hypercube such that (1) each side of the hypercube is strictly parallel to the axis (2) it contains the following point set

$$\left\{ \phi_k(\mathbf{v} + \mathbf{v}') \,\middle|\, \text{for every } i,j \in [n], \mathbf{v}'[i,j] = 0 \text{ or } \frac{1}{n^\tau} \right\}$$

Because ϕ_k is a monotone function, we know that every vector $\mathbf{v}''$ that lies between $\mathbf{v}$ and $\mathbf{v} + \frac{1}{n^\tau}\mathbf{J}$ will be mapped to a point at $\mathcal{C}_{\mathbf{v}}$, where $\mathbf{J}$ is the all-1 matrix. In this way, we give a cover to the space of $\mathbf{M}^k$, where some hypercubes overlap with others. The hypercubes are always inside the boundary of the space of $\mathbf{M}^k$ (i.e., the hypercubes do not exceed $\mathbf{M}^k_{\min}$ and $\mathbf{M}^k_{\max}$) because valid $\mathbf{w}$ is distant from the boundary. We will prove this in the next lemma.

As each hypercube $\mathcal{C}_{\mathbf{v}}$ at least contains one point that has a preimage (e.g., $\mathbf{v}$). We only need to bound the edge length of the hypercube, which is the largest possible (max-norm) distance from each valid $\mathbf{w}$ to a point $\mathbf{w}''$ that has a preimage. For each hypercube $\mathcal{C}_{\mathbf{v}}$, by Corollary 10 we have

$$\begin{aligned}
\max_{\mathbf{v}',\mathbf{v}'' \in \mathcal{C}_{\mathbf{v}}} ||\phi_k(\mathbf{v}') - \phi_k(\mathbf{v}'')|| &= \max_{\mathbf{v}',\mathbf{v}'',u,v} \left|(\phi_k(\mathbf{v}') - \phi_k(\mathbf{v}''))[u,v]\right| \\
&\le \max_{u,v\in[n]} \sum_{i,j\in[n]} \left| \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}_{\mathbf{U}}[i,j]} \cdot n^{-\tau} \right| \\
&= \max_{u,v\in[n]} \sum_{i,j\in[n]} \left| \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} \cdot n^{-\tau-3} \right| \\
&\le n^{-\tau-3} \left(k + \frac{k^2}{n^3} + \frac{k^3}{6n^6} \right) \le 2kn^{-\tau-3}
\end{aligned}$$

□

With Lemma 19 we have shown that there exists $\mathbf{w}'' \in \mathcal{Q}_{n^2}(\mathbf{w}, \frac{2k}{n^{\tau+3}})$ such that $\mathbf{w}''$ has a preimage. Then we can apply a similar argument. Moreover, we have

$$\mathcal{Q}_{n^2}(\mathbf{w}'', r - \frac{2k}{n^{\tau+3}}) \subseteq \mathcal{Q}_{n^2}(\mathbf{w}, r) \subseteq \mathcal{Q}_{n^2}(\mathbf{w}'', r + \frac{2k}{n^{\tau+3}})$$

Recall that r is either $\frac{1}{2n^\alpha}$ or $\frac{1}{2n^\alpha+1}$. So we can get: when $\mathbf{w}$ is not close to the boundary in the space of $\mathbf{M}^k$, such that all the hypercubes above are inside the boundary of $\mathcal{M}_{\gamma,\tau}$ in the space of $\mathbf{M}^k$, we have

$$\left(\frac{2n^\tau}{||\phi_k||_{\text{Lip}}} \cdot \left(r - \frac{2k}{n^{\tau+3}} \right) - 1 \right)^{n^2} \le |\{\mathbf{M}_{\mathbf{U}} | \mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}, r)\}| \le \left(\frac{2n^\tau}{||\phi_k^{-1}||^{-1}_{\text{Lip}}} \cdot \left(r + \frac{2k}{n^{\tau+3}} \right) + 1 \right)^{n^2}$$

This is exactly the bound we need. Now we justify that the hypercubes are indeed inside the boundary.

Lemma 20. *Given $2 \le \gamma < \alpha - 12.2$ and $\tau > \alpha$ constants. For every valid center $\mathbf{w}_{\text{valid}}$ (i.e., $\mathbf{w}_{\text{valid,local}}$ or $\mathbf{w}_{\text{valid,global}}$), and a point $\mathbf{w}'' \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid}}, \frac{2k}{n^{\tau+3}})$ that has a preimage, both hypercubes $\mathcal{Q}_{n^2}(\mathbf{w}'', \frac{1}{2n^{\alpha}+1} - \frac{2k}{n^{\tau+3}})$ and $\mathcal{Q}_{n^2}(\mathbf{w}'', \frac{1}{2n^{\alpha}} + \frac{2k}{n^{\tau+3}})$ are inside the boundary of $\mathcal{M}_{\gamma,\tau}$ in the space of $\mathbf{M}^k$.*

Proof. We know that for each index $(i,j) \in \{1,\dots,n\} \times \{1,\dots,n\}$, the (i,j)-th element of $\mathbf{w}_{\text{valid}}$ is inside the range $[\mathbf{M}^k_{\mathcal{E}\,\min}[i,j] - \frac{1}{2n^{\alpha}}, \mathbf{M}^k_{\mathcal{E}\,\max}[i,j] + \frac{1}{2n^{\alpha}}]$. Notice that the second hypercube contains the first one. The hypercubes are inside the boundary if

$$\left(\mathbf{M}^k_{\mathcal{E}\,\max}[i,j] + \frac{1}{2n^{\alpha}}\right) + \frac{2k}{n^{\tau+3}} + \left(\frac{1}{2n^{\alpha}} + \frac{2k}{n^{\tau+3}}\right) < \mathbf{M}^k_{\max}[i,j]$$

$$\left(\mathbf{M}^k_{\mathcal{E}\,\min}[i,j] - \frac{1}{2n^{\alpha}}\right) - \frac{2k}{n^{\tau+3}} - \left(\frac{1}{2n^{\alpha}} + \frac{2k}{n^{\tau+3}}\right) > \mathbf{M}^k_{\min}[i,j]$$

This is true because

$$||\mathbf{M}^k_{\max} - \mathbf{M}^k_{\mathcal{E}\,\max}|| > ||\phi_k^{-1}||^{-1}_{\text{Lip}} \cdot \frac{1}{n^{\gamma+3}} = \Theta\left(\frac{k}{n^{\gamma+6}}\right) > \frac{1}{n^{\alpha}} + \frac{4k}{n^{\tau+3}}$$

which is true given $2 \le \gamma < \alpha - 12.2$ and $\tau > \alpha$. And the same holds for $\mathbf{M}^k_{\mathcal{E}\,\min} - \mathbf{M}^k_{\min}$. □

We give a brief summary of what we have achieved above. To give both an upper bound and a lower bound to the probability mass of hypercubes $\mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid}}, r)$, where r is either $\frac{1}{2n^{\alpha}}$ or $\frac{1}{2n^{\alpha}+1}$. We first relocate $\mathbf{w}_{\text{valid}}$ to $\mathbf{w}'' \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid}}, \frac{2k}{n^{\tau+3}})$ such that $\mathbf{w}''$ has a preimage $\mathbf{M}_{\mathbf{U}}$ of $\phi(\mathbf{M}_{\mathbf{U}})$. The existence of such $\mathbf{w}''$ is promised by Lemma 19. It now suffices to lower bound $\mathcal{Q}_{n^2}(\mathbf{w}'', r - \frac{2k}{n^{\tau+3}})$, which is contained in $\mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid}}, r)$, and to upper bound $\mathcal{Q}_{n^2}(\mathbf{w}'', r + \frac{2k}{n^{\tau+3}})$, which contains $\mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid}}, r)$.

By the Lipschitz analysis (Lemma 15) of the function $\phi_k(\mathbf{M}_{\mathbf{U}}) = \mathbf{M}^k$ and its inverse $\phi_k^{-1}(\mathbf{M}^k) = \mathbf{M}_{\mathbf{U}}$, we know that the distance between matrices is preserved given either ϕ_k or ϕ_k^{-1}. So, every point $\mathbf{v}$ inside $\mathcal{Q}_{n^2}(\mathbf{w}'', r \pm \frac{2k}{n^{\tau+3}})$ has a preimage $\phi_k^{-1}(\mathbf{v})$ close to $\phi_k^{-1}(\mathbf{w}'')$, and every point $\mathbf{v}'$ close to $\phi_k^{-1}(\mathbf{w}'')$ should also have a image $\phi(\mathbf{v}')$ contained in $\mathcal{Q}_{n^2}(\mathbf{w}'', r \pm \frac{2k}{n^{\tau+3}})$. The distortion of the distance is quantified by the Lipschitz constants $||\phi_k||_{\text{Lip}}, ||\phi_k^{-1}||^{-1}_{\text{Lip}}$. Recall that $\mathbf{M}_{\mathbf{U}}$ is uniformly sampled from a discrete and equidistant distribution. To bound $\mathcal{Q}_{n^2}(\mathbf{w}'', r \pm \frac{2k}{n^{\tau+3}})$ we just count the number of points that within max-norm distance to $\phi_k^{-1}(\mathbf{w}'')$, which is also a hypercube.

9.3 Proof of the main theorem

Now we are ready to prove the main theorem of this section:

Theorem 6 (main result for the volume and embeddings part)**.** *For every $N = O(\text{polylog}(n))$, and every $2 \le k \le n$. Let $N_{\text{core}} = \lceil \sqrt{N} \rceil$, $N_{\overline{\text{core}}} = n^2 - N_{\text{core}}$. Let $\tau > \alpha$, $2 \le \gamma < \alpha - 12.2$, $\eta = \alpha - 7.1$ be constants. Let $\mathbf{w}_{\text{core}} \in \mathbb{R}^{N_{\text{core}}}$ to be an arbitrary assignment to arbitrary N_{core} elements of $\mathbf{M}^k$, and $\mathbf{w}_{\overline{\text{core}}} \in \mathbb{R}^{N_{\overline{\text{core}}}}$ to be an assignment to the remaining $N_{\overline{\text{core}}}$ elements. We have:*

$$\Pr_{\mathbf{M} \leftarrow \mathcal{M}_{\gamma,\tau}}\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{core}}, \frac{1}{2n^{\eta}}\right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{core}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^{\alpha}}\right)\right] \le 55 \cdot n^{(\gamma-\eta+3.1)N_{\text{core}}}$$

Proof. Recall that we have shown

$$\begin{aligned}
&\Pr\left[(\mathbf{M}^k)_{\text{core}} \in \mathcal{Q}_{N_{\text{core}}}\left(\mathbf{w}_{\text{core}}, \frac{1}{2n^\eta}\right), \mathbf{M} \in \mathcal{E} \,\middle|\, \left(\mathbf{M}^k\right)_{\overline{\text{core}}} \in \mathcal{Q}_{N_{\overline{\text{core}}}}\left(\mathbf{w}_{\overline{\text{core}}}, \frac{1}{2n^\alpha}\right)\right] \\
\leq &\frac{\sum_{\mathbf{w}_{\text{valid,local}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,local}}, \frac{1}{2n^\alpha})]}{\sum_{\mathbf{w}_{\text{valid,global}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,global}}, \frac{1}{2n^\alpha+1})]}
\end{aligned}$$

Then the number of "valid, local" points is upper bounded by: $\left(\frac{n^{-\eta}+n^{-\alpha}}{n^{-\alpha}}\right)^{N_{\text{core}}}$. As for the "valid, global" points, the number is at least $\left(\frac{||\phi_k^{-1}||_{\text{Lip}}^{-1}(n^{-\gamma}-2n^{-\gamma-3})-n^{-\alpha}}{n^{-\eta}} \cdot \frac{\frac{2}{2n^\eta+1}-n^{-\alpha}}{n^{-\alpha}}\right)^{N_{\text{core}}}$, the number of grid points times the number of length-$n^{-\alpha}$ hypercubes inside each length-$n^{-\eta}$ hypercubes.

We bound it below

$$\begin{aligned}
\frac{\sum_{\mathbf{w}_{\text{valid,local}}} 1}{\sum_{\mathbf{w}_{\text{valid,global}}} 1} &= \left(\frac{n^{-\eta}}{||\phi_k^{-1}||_{\text{Lip}}^{-1}(n^{-\gamma}-2n^{-\gamma+3})-n^{-\alpha}} \cdot \frac{n^{-\eta}+n^{-\alpha}}{\frac{2}{2n^\eta+1}-n^{-\alpha}}\right)^{N_{\text{core}}} \\
&\leq \left(\frac{2n^{3+\gamma}}{k} \cdot \frac{n^{-\eta}+n^{-\alpha}}{1-n^{-\eta}-n^{-7.1}}\right)^{N_{\text{core}}} \\
&\leq n^{(3+\gamma-\eta+\frac{4}{\log n})N_{\text{core}}}
\end{aligned}$$

Now we bound the probability.

$$\begin{aligned}
&\frac{\sum_{\mathbf{w}_{\text{valid,local}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,local}}, \frac{1}{2n^\alpha})]}{\sum_{\mathbf{w}_{\text{valid,global}}} \Pr[\mathbf{M}^k \in \mathcal{Q}_{n^2}(\mathbf{w}_{\text{valid,global}}, \frac{1}{2n^\alpha+1})]} \\
\leq &n^{(3+\gamma-\eta+\frac{4}{\log n})N_{\text{core}}} \cdot \frac{\left(\frac{2n^\tau}{||\phi_k^{-1}||_{\text{Lip}}^{-1}} \cdot \left(\frac{1}{2n^\alpha}+\frac{2k}{n^{\tau+3}}\right)+1\right)^{n^2}}{\left(\frac{2n^\tau}{||\phi_k||_{\text{Lip}}}\left(\frac{1}{2n^\alpha+1}-\frac{2k}{n^{\tau+3}}\right)-1\right)^{n^2}} \\
= &n^{(3+\gamma-\eta+\frac{4}{\log n})N_{\text{core}}} \cdot \left(\frac{||\phi_k||_{\text{Lip}}}{||\phi_k^{-1}||_{\text{Lip}}^{-1}}\right)^{n^2} \cdot \left(\frac{\frac{1}{2n^\alpha}+\frac{2k}{n^{\tau+3}}+\frac{||\phi_k^{-1}||_{\text{Lip}}^{-1}}{2n^\tau}}{\frac{1}{2n^\alpha+1}-\frac{2k}{n^{\tau+3}}-\frac{||\phi_k||_{\text{Lip}}}{2n^\tau}}\right)^{n^2} \\
\leq &n^{(3+\gamma-\eta+\frac{4}{\log n})N_{\text{core}}} \cdot \left(1+4n^{-2}\right)^{n^2} \cdot \left(\frac{\frac{1}{2n^\alpha}+\frac{2.5k}{n^{\tau+3}}}{\frac{1}{2n^\alpha+1}-\frac{2.5k}{n^{\tau+3}}}\right)^{n^2} \\
\leq &n^{(3+\gamma-\eta+\frac{4}{\log n})N_{\text{core}}} \cdot \left(1+4n^{-2}\right)^{n^2} \cdot \left(1+\frac{20}{n^{\tau+2-\alpha}}\right)^{n^2} \\
\leq &55 \cdot n^{(3.1+\gamma-\eta)N_{\text{core}}} < n^{-2N_{\text{core}}}
\end{aligned}$$

where we used the facts that $\gamma < \alpha - 12.2$, $\eta = \alpha - 7.1$ and $\alpha < \tau$. □

The proof of Theorem 4 directly follows from Theorem 5 and Theorem 6 that, if we set $N_{\text{core}} = \lceil \frac{s}{\log n} \rceil$, the restricted conditional min-entropy $H_{\min}(f)$ is bounded by $-\log(n^{-2N_{\text{core}}}) \geq 2s$. By Theorem 1, we know that the average query complexity is bounded by $\Omega(\sqrt{n})$.

Acknowledgements

We would like to thank William Hoza, Dieter van Melkebeek, and Ryan Williams for the discussions and valuable suggestions. We also wish to thank the reviewers for their constructive feedback on preliminary aspects of this work.

References

[AKL+79] Romas Aleliunas, Richard M Karp, Richard J Lipton, László Lovász, and Charles Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*, pages 218–223. IEEE Computer Society, 1979.

[AKM+20] AmirMahdi Ahmadinejad, Jonathan Kelner, Jack Murtagh, John Peebles, Aaron Sidford, and Salil Vadhan. High-precision estimation of random walks in small space. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1295–1306. IEEE, 2020.

[AMS96] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 20–29, 1996.

[Arm98] Roy Armoni. On the derandomization of space-bounded computations. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 47–59. Springer, 1998.

[AW09] Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *ACM Transactions on Computation Theory (TOCT)*, 1(1):1–54, 2009.

[BBR+96] Paul Beame, Allan Borodin, Prabhakar Raghavan, Walter L Ruzzo, and Martin Tompa. Time-space tradeoffs for undirected graph traversal by graph automata. *information and computation*, 130(2):101–129, 1996.

[BCG19] Mark Braverman, Gil Cohen, and Sumegha Garg. Pseudorandom pseudo-distributions with near-optimal error for read-once branching programs. *SIAM Journal on Computing*, 49(5):STOC18–242, 2019.

[BCP+13] Joshua E Brody, Shiteng Chen, Periklis A Papakonstantinou, Hao Song, and Xiaoming Sun. Space-bounded communication complexity. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 159–172, 2013.

[BDW02] Harry Buhrman and Ronald De Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[BGS75] Theodore Baker, John Gill, and Robert Solovay. Relativizations of the p=?np question. *SIAM Journal on computing*, 4(4):431–442, 1975.

[BH98] Richard Beigel and Tirza Hirst. One help-bit doesn't help. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 124–130, 1998.

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[CCvM06] Jin-Yi Cai, Venkatesan T Chakaravarthy, and Dieter van Melkebeek. Time-space tradeoff in derandomizing probabilistic logspace. *Theory of Computing Systems*, 39(1):189–208, 2006.

[CDR+21] Gil Cohen, Dean Doron, Oren Renard, Ori Sberlo, and Amnon Ta-Shma. Error reduction for weighted prgs against read once branching programs. *Leibniz international proceedings in informatics*, 200(22), 2021.

[CDSTS23] Gil Cohen, Dean Doron, Ori Sberlo, and Amnon Ta-Shma. Approximating iterated multiplication of stochastic matrices in small space. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 35–45, 2023.

[CJSW24] Lijie Chen, Ce Jin, Rahul Santhanam, and Ryan Williams. Constructive separations and their consequences. *TheoretiCS*, 3, 2024.

[CL20] Eshan Chattopadhyay and Jyun-Jie Liao. Optimal error pseudodistributions for read-once branching programs. In *35th Computational Complexity Conference*, 2020.

[CR80] Stephen A Cook and Charles W Rackoff. Space lower bounds for maze threadability on restricted machines. *SIAM Journal on Computing*, 9(3):636–652, 1980.

[CRT05] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. Approximating the minimum spanning tree weight in sublinear time. *SIAM Journal on computing*, 34(6):1370–1379, 2005.

[CT19] Lijie Chen and Roei Tell. Bootstrapping results for threshold circuits "just beyond" known lower bounds. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 34–41, 2019.

[CW24] Kuan Cheng and Yichuan Wang. Bpl$\subseteq$ l-ac^1. In *39th Computational Complexity Conference (CCC 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

[DPT24] Dean Doron, Edward Pyne, and Roei Tell. Opening up the distinguisher: A hardness to randomness approach for bpl= l that uses properties of bpl. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 2039–2049, 2024.

[DVM14] Holger Dell and Dieter Van Melkebeek. Satisfiability allows no nontrivial sparsification unless the polynomial-time hierarchy collapses. *Journal of the ACM (JACM)*, 61(4):1–27, 2014.

[GR14] Anat Ganor and Ran Raz. Space pseudorandom generators by communication complexity lower bounds. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[Hir22] Shuichi Hirahara. Symmetry of information from meta-complexity. In *37th Computational Complexity Conference (CCC 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[Hoz21] William M Hoza. Better pseudodistributions and derandomization for space-bounded computation. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[HZ20] William M Hoza and David Zuckerman. Simple optimal hitting sets for small-success rl. *SIAM Journal on Computing*, 49(4):811–820, 2020.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 356–364, 1994.

[KI03] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 355–364, 2003.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[LP21] Yanyi Liu and Rafael Pass. Cryptography from sublinear-time average-case hardness of time-bounded kolmogorov complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 722–735, 2021.

[M+05] Shanmugavelayutham Muthukrishnan et al. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, 1(2):117–236, 2005.

[Nis90] Noam Nisan. Pseudorandom generators for space-bounded computations. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing*, pages 204–212, 1990.

[Nis92] Noam Nisan. RL $\subseteq$ SC. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 619–623, 1992.

[NRS98] Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM Journal on Computing*, 28(3):1035–1050, 1998.

[PP23] Aaron Putterman and Edward Pyne. Near-optimal derandomization of medium-width branching programs. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 23–34, 2023.

[PRZ23] Edward Pyne, Ran Raz, and Wei Zhan. Certified hardness vs. randomness for log-space. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 989–1007. IEEE, 2023.

[PSS14] Periklis Papakonstantinou, Dominik Scheder, and Hao Song. Overlays and limited memory communication. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 298–308. IEEE, 2014.

[PV10] Mihai Pătraşcu and Emanuele Viola. Cell-probe lower bounds for succinct partial sums. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 117–122. SIAM, 2010.

[PV21] Edward Pyne and Salil Vadhan. Pseudodistributions that beat all pseudorandom generators. In *36th Computational Complexity Conference (CCC 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[Pyn24] Edward Pyne. Derandomizing logspace with a small shared hard drive. In *39th Computational Complexity Conference (CCC 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

[Rei08] Omer Reingold. Undirected connectivity in log-space. *Journal of the ACM (JACM)*, 55(4):1–24, 2008.

[RR94] Alexander A Razborov and Steven Rudich. Natural proofs. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 204–213, 1994.

[RTV06] Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks on regular digraphs and the rl vs. l problem. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 457–466, 2006.

[RY20] A. Rao and A. Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[Sak96] Michael Saks. Randomization and derandomization in space-bounded computation. In *Proceedings of Computational Complexity (Formerly Structure in Complexity Theory)*, pages 128–149. IEEE, 1996.

[Sav70] Walter J Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of computer and system sciences*, 4(2):177–192, 1970.

[Sha38] Claude E Shannon. A symbolic analysis of relay and switching circuits. *Electrical Engineering*, 57(12):713–723, 1938.

[Sha49] Claude E Shannon. The synthesis of two-terminal switching circuits. *The Bell System Technical Journal*, 28(1):59–98, 1949.

[SZ99] Michael Saks and Shiyu Zhou. $\mathrm{BP_HSPACE}(S) \subseteq \mathrm{DSPACE}(S^{3/2})$. *Journal of computer and system sciences*, 58(2):376–403, 1999.

[Vad19] Salil Vadhan. Computational entropy. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*, pages 693–726. 2019.

[Weg00] Ingo Wegener. *Branching programs and binary decision diagrams: theory and applications*. SIAM, 2000.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213, 1979.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il