Systematic Data Structure Lower Bounds via the Query-With-Sketch Model

Sumegha Garg
Rutgers University, New Brunswick, NJ, USA

Songhua He
Rutgers University, New Brunswick, NJ, USA

Yuanzhi Li
Carnegie Mellon University, Pittsburgh, PA, USA

Periklis A. Papakonstantinou
Rutgers University, New Brunswick, NJ, USA

Xin Yang
Meta, Seattle, WA, USA

Abstract

We study data structure lower bounds for the *Approximate Matrix Powering* (AMP) problem. Given a substochastic, symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ and parameters k and α, the goal is to preprocess $\mathbf{M}$ so as to answer entry queries $(u,v) \mapsto \mathbf{M}^k[u,v]$ up to additive error $1/n^\alpha$. We focus on AMP in the *succinct and systematic* regime, in which the data structure stores $\mathbf{M}$ verbatim, uses an additional r bits of redundancy, and must answer queries by probing only a small number of entries of $\mathbf{M}$.

Our main conceptual contribution is a general framework for proving probe–redundancy trade-offs for systematic data structures. We introduce the *query-with-sketch* model and develop a min-entropy-based approach that lifts conditional min-entropy bounds in the absence of redundancy to probe lower bounds in the presence of redundancy. We then establish these min-entropy bounds using problem-specific analytic and algebraic tools, for the downstream applications to AMP and its variants. As a consequence, our results provide new unconditional evidence toward a conjecture of Pătraşcu and Roditty on the space required for constant-time set-disjointness queries [31].

2012 ACM Subject Classification Theory of computation → Data structures design and analysis

Keywords and phrases systematic data structure, query-with-sketch model, approximate matrix powering

Digital Object Identifier 10.4230/LIPIcs.CCC.2026.41

Acknowledgements We thank the anonymous reviewers for their valuable suggestions.

1 Introduction

Unconditional lower bounds are among the clearest windows we have into the limits of computation. In the data structure world this promise is especially tangible: computation is free, and the only scarce resources are *information* – how many bits one stores, and how many locations one probes. A rich line of work has explored the tradeoff between these two resources across a variety of models, including the standard cell-probe model [43], as well as more restrictive frameworks such as succinct data structures [23] – under which the space usage is close to the information-theoretic minimum. While the strongest known lower bounds on the number of probes per query are logarithmic in the cell probe model [25], linear lower bounds are known for succinct data structures (e.g., [13]). In this paper, we develop a new framework for proving such tradeoffs in the succinct data structure regime, under an additional restriction: the data structure stores the input in read-only memory, together with an auxiliary data structure of at most r bits. Data structures of this form are known as *systematic* data structures[1]. Lower bounds in this model have been studied extensively in prior work [27, 13, 18, 19, 20, 6, 7].

A guiding application of our lower-bound framework is the *Approximate Matrix Powering* (AMP) problem. Given a substochastic, symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ and parameters α and k, the goal is to preprocess $\mathbf{M}$ so as to answer entry queries $(u, v) \mapsto \mathbf{M}^k[u, v]$ up to additive error $1/n^\alpha$. AMP is a fundamental computational task that arises in the analysis of the long-term behavior of Markov chains [30], in iterative methods from numerical linear algebra [17], and in PageRank and related random-walk-based algorithms [5]. AMP subsumes, as special cases, the counting set intersection problem as well as its decision version, set disjointness. These set problems are central throughout algorithms and complexity theory (e.g., [8, 31, 14, 15, 24]). In this work, we obtain (to our knowledge) the first lower bounds for set intersection and set disjointness in the systematic and static data structure model.

We model a static data structure problem as a function $f : \mathcal{X}^N \times \mathcal{Q} \to \mathcal{Y}$, where $\mathcal{X}^N$ denotes the space of possible inputs (data), $\mathcal{Q}$ denotes the set of queries, and $f(X, \phi)$ is the answer to query $\phi \in \mathcal{Q}$ on input $X \in \mathcal{X}^N$. The computation proceeds in two phases. In the preprocessing phase, a systematic data structure with unbounded computational power, given access to the input X, stores X verbatim together with an additional $r = o(N)$ bits of auxiliary information, called the *redundancy*. Subsequently, upon receiving a query $\phi \in \mathcal{Q}$, the data structure is given the redundancy bits for free and must output $f(X, \phi)$ while making only a bounded number of probes to the input X. After answering the query, the data structure is allowed to update its redundancy bits[2].

To prove general lower bounds for systematic data structures, we formalize and analyze a query-with-sketch model using a novel min-entropy lemma (Lemma 8). The key power of this lemma is that it shifts the quantifier in the analysis: instead of reasoning about the information gained from a sequence of t probes conditioned on r bits of redundancy, we reason about the entropy loss in output after the algorithm sees t arbitrary input elements (without any redundancy). This reduction is the core of our framework, as it allows us to set aside the data structure algorithm itself and focus purely on the problem's inherent properties, and it can be summarized as follows.

▶ **Theorem 1** (Lifting min-entropy to data structure lower bounds)**.** *Fix a data structure problem $f : \mathcal{X}^N \times \mathcal{Q} \to \mathcal{Y}$. Fix the redundancy parameter r. If there exists a sequence of m queries $Q \in \mathcal{Q}^m$, and an input distribution μ over $\mathcal{X}^N$ such that:*

> *For* every *partial assignment σ to at most $o(N)$ elements of the input, the min-entropy of the answers to these m queries remains high. That is, the probability of* any *single answer vector $\mathbf{y}$ is exponentially small:*
>
> $$\max_{\mathbf{y}\in\mathcal{Y}^m} \Pr_{X\leftarrow\mu}\left[(f(X, Q_1), \ldots, f(X, Q_m)) = \mathbf{y} \mid \sigma\right] \le 2^{-2r}.$$

[1] It is well known that data structure lower bounds have deep connections to other areas of theoretical computer science, including rigidity bounds [10, 33], circuit lower bounds [41, 11], cryptography [16], and fine-grained complexity [21, 9]. Importantly, [9] show that improving the state-of-the-art data structure lower bounds for the function inversion problem – under systematic and non-adaptive data structures – would also imply new circuit lower bounds in Valiant's common-bits model [38, 39].

[2] Previous works [13, 7] on succinct and systematic data structures assume probe access to the redundancy bits, and do not explicitly allow the redundancy to be updated after each query. Consider the Prefix Sum problem over a bitstring $X \in \{0, 1\}^N$, where the data structure should support querying the parity of a prefix $\bigoplus_{i=1}^k X_i$, given k. [13] proved that any succinct and systematic data structure with r bits of redundancy requires $q = \Omega(N/(r + 1))$ probes to answer parity queries. However, if we allow the data structure to update its redundancy between queries, this trade-off disappears. For the sequential query order $k = 1, \ldots, N$, a data structure can simply maintain the running parity, achieving constant costs even in the worst-case. This simple example demonstrates the separation between read-only and read-write access to redundancy.

Then, any (possibly randomized) succinct and systematic data structure (with read-write access to r bits of redundancy) that answers all queries in Q correctly with $\geq 99/100$ probability, must make $\Omega(N/m)$ probes per query on average. Here, a probe returns an element of $\mathcal{X}$.

This theorem provides a unifying template for all of our results. For a given data structure problem, we partition the query set $\mathcal{Q}$ into a sequence of blocks, each containing m queries. Let t denote the amortized probe complexity of a systematic data structure D for answering all the queries. To establish a tradeoff of the form $r \cdot t = \tilde{\Omega}(N)$, we will show that the problem satisfies the min-entropy condition with block size $m = \tilde{\Theta}(r)$. This approach yields an amortized lower bound even when D knows the entire query sequence in advance and is allowed to adaptively update its redundancy after each query. Note that the min-entropy condition must hold for *every* partial assignment: even if a benevolent adversary reveals the $o(N)$ most informative input elements, knowing the m queries, there must still remain sufficient entropy in the answers. This is a strong guarantee and is technically challenging to establish for our downstream applications to AMP. See Figure 1 for outline of our results.

Comparison with previous works

Previous lower bounds for succinct and systematic data structures have been obtained for problems such as rank and fully indexable dictionaries [27, 18, 19, 20], succinct sampling from discrete distributions [6], online matrix–vector multiplication [7], and substring search, prefix sum, polynomial evaluation, and related problems [13]. In particular, Gal and Miltersen [13] established a general lower bound of the form $(r+1)t \geq \Omega(N)$ for problems satisfying a list-decoding property: namely, that the entire list of answers can be recovered from any sufficiently large subset. While this property holds for problems with strong global algebraic structure – such as polynomial evaluation, which serves as the main example in [13] – it fails for many natural problems in which answers are more local and largely independent, including the approximate matrix powering problem studied in this work.

Our approach departs from these reconstruction-based arguments. Many previous bounds, including those of [13, 7], proceed by encoding information about the input from the behavior of a low-probe data structure on a carefully chosen sequence of queries, and then compare the resulting encoding length with an entropy lower bound for the input distribution. Such arguments naturally lead to inequalities on the total transcript length, and hence to probe–redundancy tradeoffs after the query sequence is optimized. By contrast, our query-with-sketch reduction does not attempt to reconstruct the input. Instead, it reduces lower bounds to showing that the answer vector to a block of queries retains large conditional min-entropy even after an arbitrary small set of input locations has been revealed. Thus the information-theoretic object in our framework is the residual uncertainty of the answers under partial information, rather than the entropy of the input itself.

1.1 Applications to Approximate Matrix Powering

In this section, we detail our results for the fundamental problem of Approximate Matrix Powering.

▶ **Definition 2** (Approximate Matrix Powering (AMP))**.** *Fix parameters $\alpha > 0$ (a large enough constant) and an integer k $(2 \leq k \leq n)$. The data structure must solve the following task:*

- ***Input:*** *A substochastic and symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$.*
- ***Query:*** *A pair of indices $(u, v) \in [n] \times [n]$.*
- ***Output:*** *An estimation $\hat{y} \in \mathbb{R}$ that approximates the corresponding entry of the k-th matrix power within a specified additive error:*

$$\left|\hat{y} - \mathbf{M}^k[u, v]\right| \leq \frac{1}{n^{\alpha}}$$

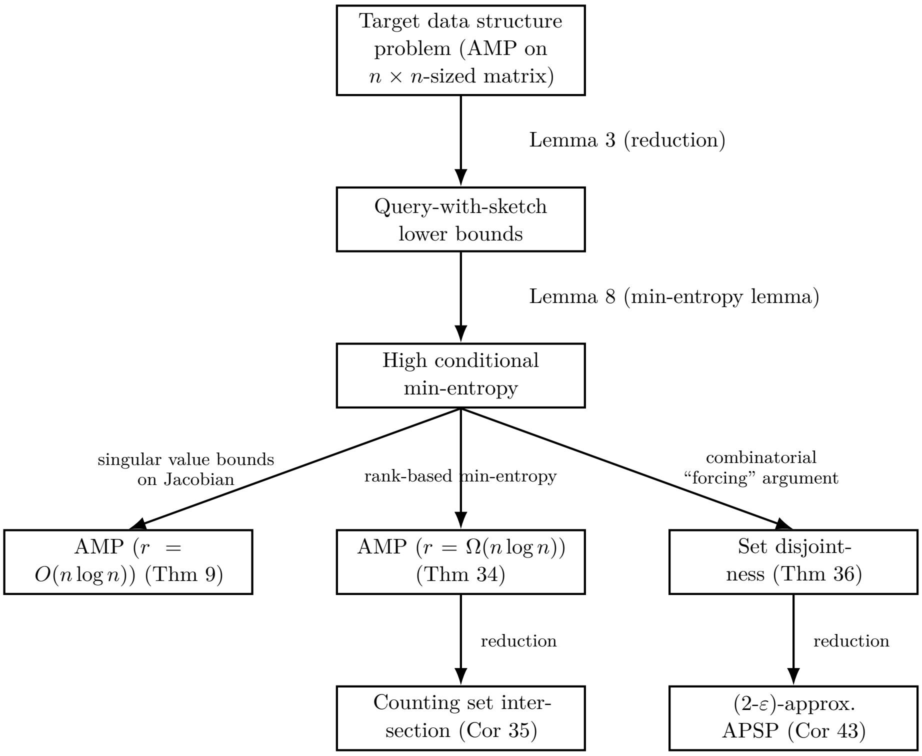

Figure 1 Our proof strategy. We reduce data structure bounds to a query-with-sketch bound (Lemma 3), which we then reduce to an information-theoretic min-entropy question (Lemma 8). We solve the latter by applying a different analytical technique for each problem.

This problem is closely related to the online matrix-vector multiplication (OMV) problem studied by [7]. That work establishes a tradeoff of $r \cdot t = \Omega(n^3)$ for any succinct or systematic data structure with r bits of redundancy and amortized probe complexity t, for answering Boolean matrix-vector multiplication queries. We emphasize that OMV and AMP are incomparable problems. On the one hand, AMP can be viewed as potentially harder than OMV, since it requires computing entries of higher powers of the matrix. On the other hand, AMP is easier in the following aspect: in OMV, the vector is part of the query and is not stored in the data structure, leading to 2^n possible queries, whereas in AMP the queries range over matrix entries and there are only n^2 possible queries.

Our data structure lower bound for the AMP problem is as follows. Since AMP involves real-valued matrices, to avoid discussion on word size, we assume that each probe returns an entry of the matrix.

Theorem (Informal statement for Theorem 9 and Theorem 34)

Fix parameters α, k. Any (possibly randomized) succinct and systematic data structure of r bits of redundancy that supports answering approximated elements of the k-th power of a substochastic and symmetric matrix $\mathbf{M}$ should make $\Omega(n^2/r)$ probes per query on average. This holds for $r \in [\Omega(n \log n), O(n^2)]$ for $k = 2$, and $r \in [\Omega(\log^2 n), O(n^2)]$ for every $k \in [3, n]$.

Since computing elements of a matrix square is a special case of matrix powering, can always be performed using $O(n)$ probes, the above lower bound holds for a full range of probe complexity. Interestingly, the cases of sublinear-in-n (Theorem 9) and superlinear-in-n

(Theorem 34) redundancy use different hard input distributions and proof techniques, where either does not seem to extend to the full range. We give a proof overview in the next subsection.

For general k, the naive algorithms give only straightforward upper bounds obtained by reading enough entries of M to compute the requested entry. We are not aware of nontrivial succinct systematic data structures for AMP that match the above probe–redundancy tradeoff across the full parameter range. Thus our result should be viewed as a lower bound establishing a new barrier for this problem, rather than as a tight characterization.

We next state our results for data structure problems on sets and graphs.

Set Intersection

First, our proof of Theorem 34 (AMP with superlinear-in-n redundancy) also yields an identical lower bound for the set intersection problem and its counting variant. Set intersection is a fundamental data structure problem, which has received significant attention in fine-grained complexity [8, 31, 14, 15, 24]. In this problem, the input consists of n subsets of the universe $[n]$, and the goal is to answer queries asking for the size of the intersection between the u-th and v-th sets. We show that, under the same parameter regime, answering each query requires $\Omega(n^2/r)$ probes on average. See Corollary 35 for the full statement.

To the best of our knowledge, data structure lower bounds for the set intersection problem were previously known only for restricted data structures that are in comparable to the succinct and systematic setting. For instance, in the pointer machine model [35], which disallows random access and forces data to be accessed by traversing pointers stored in cells, [1] showed tradeoff of $S(N)Q(N) = \Omega(N^{2-o(1)})$, where N is the total size of the input sets, $S(N)$ is the total space of the data structure, and $Q(N)$ is the probe cost in pointers followed.

Set disjointness

In 2010, Pătraşcu and Roditty [31] conjectured that any cell-probe data structure which preprocesses n sets $A_1, \ldots, A_n \subseteq [\mathrm{polylog}(n)]$ and answers set disjointness queries, that is, whether the u-th and v-th sets intersect, using constant probe complexity must use $\tilde{\Omega}(n^2)$ space. We study a related version of the set disjointness problem in which the sets $A_1, \ldots, A_n$ are subsets of $[n]$, and make the first progress toward this conjecture in the systematic data structure setting. In particular, we prove a time–space tradeoff of the form $r \cdot t = \Omega\left(\frac{n^2}{\log^3 n}\right)$, for any succinct and systematic data structure solving set disjointness.

Theorem (Informal statement for Theorem 36)

Any (possibly randomized) succinct and systematic data structure using r bits of redundancy, that answers set disjointness queries, should make $\Omega(n^2/(r\log^3 n))$ probes on average per query. Here, $r \in [\Omega(n \log n), O(n\sqrt{n}/\log^3 n)]$.

We note that the range of r is optimal for our hard input distribution, in which the total size of all sets is $O(n\sqrt{n})$. Moreover, the probe complexity is always at most $O(n)$, since a set disjointness query can be answered by probing only the two sets involved in the query.

All Pair Shortest Paths (APSP)

By a reduction from set disjointness, we also obtain a data structure lower bound for $(2-\varepsilon)$-approximate all-pairs shortest paths (APSP). Specifically, any succinct and systematic data structure that answers a sequence of distance queries within a multiplicative factor strictly less than 2, must either use r bits of redundancy or incur an average probe complexity of $\Omega(n^2/(r\log^3 n))$. This tradeoff holds for $r \in [\Omega(n\log n), O(n\sqrt{n}/\log^3 n)]$. See Corollary 43 for the full statement.

Data structures for answering approximate distance queries are commonly referred to as *distance oracles*. Thorup and Zwick [37] showed that there exists a distance oracle with $O(n^{3/2})$ space that answers distance queries within a factor of 3 in constant time. Subsequently, Pătraşcu and Roditty [31] constructed an $O(n^{5/3})$-space data structure that returns an estimate of at most $2d+1$ in constant time, where d is the true distance. On the lower-bound side, [31] established a conditional result showing that any cell-probe data structure achieving a 2-approximation in constant time must use $\tilde{\Omega}(n^2)$ space. Our result provides the first *unconditional* time-space tradeoff for approximate APSP. Moreover, the tradeoff is tight: when $r = \Omega(n^{3/2})$, the distance oracles from prior work can be stored entirely within the redundancy.

1.2 Proof Overview and Technical Contributions

In this section, we outline the proofs of our main results. We introduce the *query-with-sketch* model, which captures data structures that store a short sketch of the input and answer a fixed batch (or sequence) of queries using limited probes to the raw input. We then show a general *min-entropy lemma* that reduces lower bounds in this model to a conditional min-entropy bound. Finally, we instantiate this template for AMP using different analytic tools for different parameter regimes, obtaining strong probe–redundancy trade-offs.

1.2.1 The query-with-sketch model and the min-entropy lemma

First, we introduce the query-with-sketch model. The final data structure lower bounds consist of a reduction to this model, and a min-entropy lemma that transfers min-entropy upper bounds to query-with-sketch lower bounds. These definitions and reductions are formalized at Section 2.

The query-with-sketch model can be thought of as a succinct and systematic data structure with its queries fixed beforehand[3]. An algorithm in the query-with-sketch model aims to compute a multi-bit function $f : \{0,1\}^N \to \{0,1\}^M$ (or more generally, $f : \mathcal{X}^N \to \mathcal{Y}$) on x with the help of a short piece of information that we call "sketch", and limited number of coordinate probes to x. The sketch size should be smaller than N and M for the model to make sense. A more detailed definition and some notation are deferred to Section 2.1.

Capabilities and limits of algorithms in the query-with-sketch

The query-with-sketch model, as a natural information-theoretic model, is of independent interest. We use the Hamming weight problem to illustrate that sketch can help reduce the query (probe) complexity in a non-trivial way. In the Hamming weight problem, the input is a string from $\{0,1\}^n$ and the output is the number of 1s. In the standard query model, n probes are needed to compute, with high probability, the exact answer to this problem. But, in the query-with-sketch, given a small sketch (shorter than $\log n$ bits) of the lower-order

[3] We adopt this name from the query complexity literature, where a "query" typically refers to a low-level read from the input. However, to maintain clarity and align with data structure conventions, this paper will consistently use "probe" for a low-level read from the input X and "query" for the high-level request. Thus, what the query complexity literature calls "query complexity" corresponds directly to what we call "probe complexity" in this paper.

digits in the hamming weight, it suffices to sample n^{ε}-many bits, $\varepsilon < 1$, to compute the higher-order digits. The intuition is that the least significant digits are harder to compute than the more significant ones.

Min-entropy lemma

All of our main results are built upon the following lemma, which provides a powerful connection between probe complexity and information. This min-entropy lemma formalizes a simple yet effective intuition: an r-bit sketch is useless if the answer would have remained more than r bits uncertain, for any set of probes the algorithm could have made.

It translates the task of proving a time-space tradeoff into a more manageable, information-theoretic question: how much uncertainty (min-entropy) about the answer $f(X)$ remains, even after we are given the values of any q locations in the input X?

Lemma (Informal statement of Lemma 8)

Let $f : \mathcal{X}^N \to \mathcal{Y}$ be a function over an adversarial distribution $\mu_{\mathcal{X}^N}$. Let Π^S be a deterministic algorithm for f using an r-bit sketch S. If for some integer q, the min-entropy of $f(X)$ remains $> r$ conditioned on any partial assignment on X of length q, then Π^S must make $> \Omega(q)$ probes on average, even given sketch S.

Here is the high-level intuition of the min-entropy lemma. The proof (see Section 2.2) views any query-with-sketch algorithm as a collection of 2^r decision trees. The r-bit sketch $S(X)$ simply tells the algorithm which one of these trees to execute. A leaf at depth q in any tree corresponds to the answers from q specific probes.

Our lemma's condition says that for any such leaf, the final answer $f(X)$ is still highly ambiguous – there are more than 2^r possible correct answers consistent with that probe path. Since the r-bit sketch (which was fixed from the start) cannot possibly resolve this ambiguity, the algorithm cannot halt and must make more probes.

This lemma is powerful, but our main results require a technical strengthening. We show that the same lower bound holds even if the high min-entropy condition is only guaranteed to apply to a large, high-probability set of "good inputs". See Lemma 8 for the full, formal statement.

Connection to succinct and systematic data structures

The query-with-sketch complexity was studied for the direct sum of parity problems [29, 2], and was applied to obtain data structure lower bounds for the prefix sum problem [13]. However, the reduction in [13] was restricted to the prefix sum problem. We defer the discussion on this reduction to Section 2.3. Below, we give a reduction that is generally applicable to many data structure problems.

▶ **Lemma 3** (Reduction from data structures to query-with-sketch algorithms)**.** *Fix a data structure problem $f : \mathcal{X}^N \times \mathcal{Q} \to \mathcal{Y}$. Suppose there is a (possibly randomized) succinct and systematic data structure D with the following properties:*

- *It has a redundancy of r bits that one can update for free.*
- *For a fixed sequence of queries $Q = (\kappa_1, \ldots, \kappa_l) \in \mathcal{Q}^l$, it answers the entire sequence correctly with probability at least $99/100$.*
- *The total number of probes made to answer the entire sequence Q is at most tl in the worst case.*

Then for any integer $m \in [l]$ that divides l, there exists a starting index $i \in \{1, m+1, 2m+1, \ldots, l-m+1\}$ and a (randomized) query-with-sketch algorithm Π^S for answering the subsequence $(\kappa_i, \ldots, \kappa_{i+m-1})$ that has:

- *An r-bit sketch S.*
- *An average-case probe complexity of at most tm.*
- *A success probability of at least $99/100$.*

While our reduction applies to randomized data structures, our min-entropy lemma applies to deterministic query-with-sketch algorithms. This gap can be fixed by applying Yao's minimax principle, while we are studying the distributional complexity of data structures and query-with-sketch algorithms.

We note that our reduction applies to common variants of data structure problems in the literature, including those where the goal is to answer a single query (e.g., [13]) and those that answer, or are motivated by, a sequence of queries (e.g., [7]).

Furthermore, the lemma's requirement for success on all queries simultaneously is not a strong restriction. An algorithm that only guarantees to answer each of the l queries correctly with high probability (e.g., $\geq 1 - o(1)$) can be converted to one that satisfies our stricter requirement. This is achieved via standard probability amplification, which increases the total probe complexity by a multiplicative factor of $O(\log l)$.

We also note that, as a general lower bounding tool, our approach inherently cannot prove time-space trade-offs stronger than $r \cdot t = \Omega(N \log |\mathcal{Y}|)$. A nontrivial query-with-sketch algorithm always has $r \leq m \log |\mathcal{Y}|$, since $m \log |\mathcal{Y}|$ bits are sufficient to encode all the answers of the m consecutive queries. In addition, a nontrivial query-with-sketch algorithm never makes more than N probes to the input, which implies that $t \leq N/m$ in our reduction. The two facts together imply that there always exists query-with-sketch algorithms achieving $r \cdot t \leq N \log |\mathcal{Y}|$, for every $r \geq 1$. To the best of our knowledge, among previous works [26, 19, 18, 13, 20, 6, 7], only a recent breakthrough [7] overcame this barrier to achieve a stronger trade-off for the OuMv (online vector-matrix-vector multiplication) problem. They achieved it by showing that any efficient algorithm for OuMv would implicitly function as a compression scheme that violates Shannon's source coding theorem. While powerful, this argument is highly tailored to the algebraic properties of matrix multiplication, and the massive 2^n probe space. It does not provide a general template for breaking the $r \cdot t$ trade-off barrier.

1.2.2 Approximate matrix powering with sublinear-in-n redundancy

Warm up: a lower bound to approximating the matrix square

Let us begin our proof outline for a restricted form of the full lower bound. Our lower bound holds for symmetric matrices and for arbitrary matrix powers $k > 2$. In the proof outline we will only discuss the lower bound for squaring, $\mathbf{M}^2$, not-necessarily symmetric matrices $\mathbf{M}$. These are the only simplifying assumptions made for this high-level exposition. Specifically, in this proof overview we will show that every succinct and systematic data structure for approximating the matrix square $\mathbf{M}^2$ either requires $r = 0.5n$ bits of redundancy or $t = \Omega(n)$ number of probes for every query. After that, we show how we extend it to an $r \cdot t = \Omega(n^2)$ lower bound for general sublinear-in-n r, for computing $\mathbf{M}^k$.

While we accept every possible output within an $1/n^\alpha$ additive error, we restrict the AMP problem to a version where the output is fixed, which enables us to apply the query-with-sketch reduction. Specifically, fix a large enough constant α', we let the output to be

$[\mathbf{M}^k]_{\alpha'}$, a rounded matrix power where each element is rounded to its nearest multiple of $1/n^{\alpha'}$. We will show in Lemma 13 that the reduction to this rounded version will only incur negligible error probability, when α' is large enough compared to α.

Our two-step reduction (Lemma 3 and Lemma 8) implies that, one only needs (i) a nemesis input distribution $\mathcal{M}$; (ii) a sequence Q of n^2 queries; and (iii) to show that for every $m = r$ consecutive queries from Q, even when conditioned on knowing any $0.001n^2$ elements from $\mathbf{M}$, the answers remain an $\Omega(r)$ min-entropy.

We construct a nemesis input distribution $\mathcal{M}_{\beta,\gamma}$ of the form $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^\beta}\mathbf{M_U}$, where $\mathbf{M_U}$ is a random matrix where each element is i.i.d. (independent and identically distributed) uniformly random from the range $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For the reader's convenience, we may consider an instance where $\beta = 2$ and $\gamma = 4$. In the actual proof, $\mathbf{M_U}$ is a uniformly random symmetric matrix, and a similar proof applies. In addition, we construct the sequence of queries Q in a way that every r consecutive queries span r different rows and columns. To illustrate our technique in this warm-up, focus on an illustrative block of r diagonal queries of approximating $\mathbf{M}^2[1,1], \mathbf{M}^2[2,2], \ldots, \mathbf{M}^2[r,r]$.

Notice that $\mathbf{M}^2 = (\frac{n^2-1}{n^2})^2\mathbf{I} + \frac{2}{n^2} \cdot \frac{n^2-1}{n^2}\mathbf{M_U} + \frac{1}{n^4}\mathbf{M_U}^2$. We shall assume that the elements $\mathbf{M}[1,1], \mathbf{M}[2,2], \ldots, \mathbf{M}[r,r]$ are revealed to the algorithm for free, since this will only decrease the probe complexity. We will show that, however, the second-order term $\mathbf{M_U}^2$ remains highly uncertain.

More formally, we use σ to denote an arbitrary *partial assignment* to $0.001n^2$ elements of $\mathbf{M}$. Our reductions (the reduction from data structure lower bounds to query-with-sketch lower bounds, and the reduction to the min-entropy of the problem) help us get rid of analyzing the algorithms, and reduce it to a task of showing that the output has a high conditional min-entropy:

$$\max_{\sigma,(y_i)_{i\in[r]}} \Pr_{\mathbf{M}\leftarrow\mathcal{M}}[[\mathbf{M}^2[1,1]]_{\alpha'} = y_1, \ldots, [\mathbf{M}^2[r,r]]_{\alpha'} = y_r | \sigma] < 2^{-2r}. \tag{1}$$

Our strategy is to select $\Omega(r)$ many input elements that are not fixed by σ, so that their *Jacobian matrix* with the output elements is a diagonal matrix. That is, (if given all the other $n - \Omega(r)$ elements fixed), each output element is an increasing function of its corresponding input element, but does not depend on other elements. We call these selected elements *core* elements. We can achieve it because each element of the matrix square only depends on input elements of the same row or the same column.

More precisely, for each output element $\mathbf{M}^2[u,u]$, we try to find an input element $\mathbf{M}[u,i]$ such that $i > r$, and $\mathbf{M}[u,i]$ is not fixed by σ. That is, the element shares the same row or column with $\mathbf{M}^2[u,u]$, but does not share the same row or column with any other output element. We use $\mathbf{J}$ to denote its Jacobian matrix with the output elements, where each row corresponds to a core element $\mathbf{M}[u_i,i]$, each column corresponds to an output element $\mathbf{M}^2[u,u]$, and each $\mathbf{J}[i,u] := \frac{\partial \mathbf{M}^2[u,u]}{\partial \mathbf{M}[u_i,i]}$ quantifies how the rate $\mathbf{M}^2[u,u]$ will increase if we perturb $\mathbf{M}[u_i,i]$ by a little bit. Intuitively, the higher the rate is, the less likely the output will evaluate to a fixed value, which yields a min-entropy bound.

$$\mathbf{J} = \begin{pmatrix} \Omega(1/n^3) & 0 & \ldots & 0 \\ 0 & \Omega(1/n^3) & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & \Omega(1/n^3) \end{pmatrix}$$

We can always select $r' = \Omega(r)$ core elements since a partial assignment σ of length $0.001n^2$ can only fix a small proportion of elements at the first $r = 0.5n$ rows and the first r columns.

Given these core elements $\mathbf{M}[i_1, j_1], \ldots, \mathbf{M}[i_{r'}, j_{r'}]$, we may discard all the other elements of $\mathbf{M}$. Specifically, upper bounding the conditional probability in inequality (1) can be reduced to upper bound

$$\max_{\sigma', (y_i)_{i\in[r]}} \Pr_{\mathbf{M}\leftarrow\mathcal{M}}[[\mathbf{M}^2[1,1]]_{\alpha'} = y_1, \ldots, [\mathbf{M}^2[r,r]]_{\alpha'} = y_r | \sigma'] \tag{2}$$

where σ' is any partial assignment for all the other elements of $\mathbf{M}$ except core elements, such that σ' is consistent with σ. This is because the previous probability is a weighted sum of the latter one.[4]

The Jacobian matrix tells us that given σ', the output elements are mutually independent. For each core element $\mathbf{M}[i, j]$, its corresponding output element $\mathbf{M}^2[u, u]$, and each possible output y_u, given that $\frac{\partial \mathbf{M}^2[u,u]}{\partial \mathbf{M}[i,j]} = \Omega(1/n^3)$, we have

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}}[[\mathbf{M}^2[u,u]]_{\alpha'} = y_u | \sigma'] \leq O(1/n^{\alpha'+9})$$

which is obtained by dividing the output length $(1/n^{\alpha'})$ by the partial derivative $(\Omega(1/n^3))$ and the range length that the core element is uniformly selected from $(1/n^{\gamma+\beta} = 1/n^6)$.

Therefore, the probability in (2) is upper bounded by $O((1/n^{\alpha'+9})^{r'}) = O(1/2^{\Omega(r \log n)})$, which implies a desired $\Omega(r \log n)$ conditional min-entropy.

From $k = 2$ to $k \geq 3$: singular value bounds to the Jacobian matrix

While each element of the matrix square only depends on $2n-1$ elements of the input matrix, one cannot get a stronger probe lower bound for a smaller redundancy r. Therefore, we generalize the above arguments to $3 \leq k \leq n$, where an $r \cdot t = \Omega(n^2)$ lower bound still holds.

We use the same input distribution $\mathcal{M}$ and the sequence of queries Q as above. When $k \geq 3$, the expansion $\mathbf{M}^k = \sum_{l=0}^{k} \binom{k}{l} (\frac{n^2-1}{n^2})^{k-l} (\frac{1}{n^\beta})^l {\mathbf{M}_\mathbf{U}}^l$ is still dominated by the terms where t are small. Again, we fix and focus on a sequence of output elements $\mathbf{M}^k[1,1], \ldots, \mathbf{M}^k[r,r]$ for some $r = o(n)$.

Given a partial assignment of length $\leq 0.001n^2$, unlike the previous case, all the elements of the first r rows and columns may be fixed by σ. However, we are still going to show that the output remains a high min-entropy. Our proof aligns with the proof outline above, but with the following edits:

First, we partition the set of input elements of $\mathbf{M}$ from the last $n-r$ columns and $n-r$ rows into disjoint subsets of size r, such that each pair of elements from the same subset does not share the same row or the same column. In this way, given a partial assignment σ of length $\leq 0.001n^2$, there will be at least one subset where $r' = \Omega(r)$ elements in it are not fixed by σ. We choose these elements as our core elements.

Then, denote by $(i_1, j_1), \ldots, (i_{r'}, j_{r'})$ the core elements. We look into the Jacobian matrix between the core elements and the output elements again. Since the r output elements and r' core (input) elements span different rows and columns, their Jacobian matrix is of the following form for $K := \binom{k}{3} \frac{(n^2-1)^{k-3}}{n^{2k-6+2\beta}}$.

[4] 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 $B' \subseteq B$.

$$
\mathbf{J} = \begin{pmatrix} K\mathbf{M_U}[1,i_1]\mathbf{M_U}[j_1,1] + O(k^5/n^{2+3\beta}) & \dots & \dots & K\mathbf{M_U}[r,i_1]\mathbf{M_U}[j_1,r] + O(k^5/n^{2+3\beta}) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ K\mathbf{M_U}[1,i_{r'}]\mathbf{M_U}[j_{r'},1] + O(k^5/n^{2+3\beta}) & \dots & \dots & K\mathbf{M_U}[r,i_{r'}]\mathbf{M_U}[j_{r'},r] + O(k^5/n^{2+3\beta}) \end{pmatrix}
$$

This matrix $\mathbf{J}$ is a perturbed random matrix. The dominant term for each entry is a product of two random $\mathbf{M_U}$ elements, while the $O(k^5/n^{2+3\beta})$ term is a small, deterministic-like perturbation. Critically, because all $r + r'$ input and output indices are distinct, each $\mathbf{M_U}$ element appears at most once in the dominant terms of $\mathbf{J}$. This makes the dominant terms i.i.d. random variables.

We apply a concentration bound on its singular values [42] to show that the largest $0.9r'$ singular values of $\mathbf{J}$ are high with high probability. Intuitively, the singular values of a matrix measure the "stretch" it applies to its input vectors. A matrix with $\Omega(r')$ large singular values is robustly high-rank.

This high-rank property is the key to our min-entropy bound. We use it to bound the pre-image of a fixed output. If the map "stretches" the input space in $\Omega(r')$ dimensions, then to map to a tiny output region (our rounded answer box of side length $1/n^{\alpha'}$), the input must originate from an exponentially small-volume region. This small volume, under our uniform distribution $\mathcal{M}$, translates directly to an exponentially small probability, which is the high conditional min-entropy we need.

By a union bound over the (polynomially many) possible choices for the core element set, we can define our "good input set" $\mathcal{M}$ for which the conditional min-entropy can be bounded as the set where this high-rank property holds for all relevant Jacobians.

More specifically, given σ', we show that if two core element assignments $\mathbf{x}, \mathbf{x}'$ map to the same output $\mathbf{y}$ (within precision $1/n^{\alpha'}$), they must be geometrically close, if a constant proportion of singular values of $\mathbf{J}$ are large.

In fact, for any fixed output $\mathbf{y}$, the set of all consistent core inputs $\mathbf{x}$ is *concentrated* into an exponentially small-volume region. Under our uniform input distribution, this small volume implies an exponentially small probability, which confirms the high conditional min-entropy.

We omitted some details that will be clarified in the full proof.

1.2.3 Approximate matrix powering with superlinear-in-n redundancy

The above arguments do not easily extend to the case $r = \omega(n \log n)$, since the Jacobian matrix is no longer controlled when there are more than n core elements and more than n output elements. The singular value concentration bound [42] only works for random matrices whose elements are independent. To that end, we construct and study the following combinatorial distribution of matrices.

To better illustrate the distribution, we describe its transition graph $G = (V, E, W)$ instead, which is an undirected graph with edge weight to be its transition probability from both sides. Consider a vertex set $[n]$ that is partitioned to three parts V_1, V_2, V_3, each of size exactly $n/3$. For every pair of vertices from $V_1 \times V_2$ or $V_2 \times V_3$, we connect them by an edge of weight $1/n^3$ with independent probability $1/2$. In addition, we connect each vertex to itself by an edge of weight $1 - 1/n^2$. Note that as a substochastic matrix, we do not require each row sum of the matrix to be 1, but less than or equal to 1.

We construct our sequence of queries as the sequence of all pairs of vertices between V_1 and V_3. The sequence is constructed in a way that among every r consecutive queries, each vertex is mentioned by $\Theta(r/n)$ times.

Intuitively, the graph is a 3-layered graph with heavy lazy transitions. We will show in Observation 29 that approximating the matrix power of the transition graph reduces to counting the number of length-2 walks from each $u \in V_1$ to $v \in V_3$. Due to the heavy lazy transitions, the matrix power is dominated by the length-2 walks. In fact, we will show that even computing the *parity* of the number of walks is hard.

Although the nemesis input distribution is combinatorial, the analysis remains linear-algebraic in spirit. We will try to find a basis of the input to the output that preserves the min-entropy.

Given a partial assignment of length $\leq 0.001n^2$, we remove those vertices whose $\geq 0.01n$ neighbors are revealed. For each of the remaining vertices v from V_3, there are $O(r/n)$ queries on v. We denote by $(u_1, v), \ldots, (u_m, v) \in V_1 \times V_3$ the queries on v, for $m = O(r/n)$. We select m corresponding vertices $w_1, \ldots, w_m \in V_2$ such that none of the edges $\forall i \in [m], (w_i, v)$ are revealed by σ.

As above, we fix a partial assignment σ' that reveals all the input edges other than the selected edges above, and is consistent with σ. Lower-bounding the min-entropy conditioned on σ reduces to bounding the min-entropy conditioned on every such σ'. Given that the whole subgraph between V_1 and V_2 is revealed by σ', queries on different vertices $v \in V_3$ are independent, and the min-entropy is the sum of min-entropy on each individual v.

Fix a vertex $v \in V_3$, denote by $\hat{\mathbf{M}} \in \mathbb{F}_2^{m \times m}$ the indicator matrix of whether each $u_i \in V_1$ is connected to w_j. Denote by $\mathbf{x} \in \mathbb{F}_2^m$ the indicator vector of whether each $(w_i, v) \in E$ or not, which is a uniformly random vector conditioned on σ'. The parity of the output (the number of length-2 paths between each u_i and v) is exactly $\hat{\mathbf{M}}\mathbf{x}$. Using a classic result that a random binary matrix has a high rank with high probability ([3]), we show that with high probability every $\hat{\mathbf{M}}$ with $m \geq 10 \log n$ has a high rank, which implies a high min-entropy of the output.

1.2.4 Lower bounds to $(2 - \varepsilon)$-approximate APSP

In APSP, we study the *existence* instead of the parity of the number of length-2 paths between pairs of vertices. The graph distribution defined above is too dense, forcing the distance between vertices from V_1 and V_3 to be 2 almost surely. To that end, we work on its sparse variant instead. We still consider a graph $G = (V, E)$ of three layers V_1, V_2, V_3 with $n/3$ vertices each. However, for each pair of vertices $(u, w) \in V_1 \times V_2$ or $(w, v) \in V_2 \times V_3$, we connect them with independent probability $1/\sqrt{n} \log n$. For each pair of vertices $(u, v) \in V_1 \times V_3$, they have a length-2 path if and only if there exists a vertex $w \in V_2$ such that $(u, w), (w, v) \in E$, i.e., whether the neighbor sets of u and v intersect. Since the entropy of the graph is $\Theta(n\sqrt{n}/\log n)$, we prove an $\Omega(n^2/\log^3 n)$ time-space trade-off for every $\Omega(n \log^2 n) \leq r \leq O(n\sqrt{n}/\log^3 n)$, where the range of r is nearly tight.

A similar proof to the above can no longer apply here. While the output can still be formulated as a matrix-vector product of the form $\hat{\mathbf{M}}\mathbf{x}$, the arithmetic is over semirings (i.e., ANDs and ORs). There does not exist a natural notion of a "basis" in this setting.

Instead, we introduce a combinatorial *forcing argument*. We show that for a sparse graph, a single "disjoint" (0) answer for a query (u, v) *forces* a large number of potential edges to be non-existent (0). The high min-entropy comes from two facts: (i) with high probability a large proportion $(1 - 1/\text{polylog}(n))$ of answers are 0 (Lemma 42); and (ii) only with exponentially small probability that *many* such forcing events (from a block of queries) can be simultaneously satisfied by a random graph.

We omitted details on how to apply the forcing argument to the input graph conditioned on a partial assignment σ, and how we deal with partial assignments that reveal too much information about the answer. We deal with these problems formally in Section 5.

2 Query-with-sketch model and the min-entropy lemma

In this section we give the two abstract ingredients used throughout the paper: the query-with-sketch model and a min-entropy lemma for proving lower bounds in this model. We then show how lower bounds in the query-with-sketch model imply lower bounds for succinct and systematic data structures.

Because the query-with-sketch model will serve as the target of our lower bounds, we make its execution model explicit: the sketch is fixed from the input, subsequent probes may be chosen adaptively from the transcript, and the algorithm eventually halts with an output.

2.1 Query-with-sketch model

We define the query-with-sketch (QwS) model for a fixed function $f : \mathcal{X}^N \to \mathcal{Y}$; in contrast, a data structure is typically specified by a query problem $f : \mathcal{X}^N \times \mathcal{Q} \to \mathcal{Y}$, where the query $Q \in \mathcal{Q}$ determines which value $f(X, Q)$ must be returned.

▶ **Definition 4** (Query-with-sketch algorithms). *Let $f : \mathcal{X}^N \to \mathcal{Y}$ be a function. A deterministic query-with-sketch algorithm Π^S with sketch length r consists of two objects.*

- *A sketching map $S : \mathcal{X}^N \to \{0,1\}^r$. On input X, the algorithm is given the sketch $S(X)$.*
- *A deterministic transition rule Π. For every $l \geq 0$, given a transcript*

$$C_\ell = \big(S(X), (i_1, a_1), \ldots, (i_\ell, a_\ell)\big),$$

where $i_j \in [N]$ is a probed coordinate and $a_j \in \mathcal{X}$ is the value returned by the probe, the rule Π either outputs a new probe $i_{\ell+1} \in [N]$ or halts with an answer $y \in \mathcal{Y}$.

The computation on input X starts from $C_0 = (S(X))$. If $\Pi(C_\ell) = i_{\ell+1} \in [N]$, the algorithm probes coordinate $i_{\ell+1}$, receives $a_{\ell+1} = X_{i_{\ell+1}}$, and continues with $C_{\ell+1}$. If $\Pi(C_\ell) = y \in \mathcal{Y}$, the algorithm halts and outputs y.

We allow repeated probes, and count them with multiplicity. A transcript of length ℓ induces a partial assignment to at most ℓ distinct input coordinates. Let $\mathsf{T}_{\Pi^S}(X)$ denote the number of probes made before halting, with $\mathsf{T}_{\Pi^S}(X) = \infty$ if the computation does not halt. The worst-case and average-case probe complexities are

$$\mathrm{Cost}(\Pi^S) := \sup_{X \in \mathcal{X}^N} \mathsf{T}_{\Pi^S}(X), \qquad \text{Avg-Cost}_\mu(\Pi^S) := \mathbb{E}_{X\sim\mu}\big[\mathsf{T}_{\Pi^S}(X)\big].$$

▶ **Definition 5** (Randomized query-with-sketch algorithms). *A randomized query-with-sketch algorithm is a distribution over deterministic query-with-sketch algorithms. Equivalently, a random tape ρ is sampled before the sketch is formed, and the algorithm then uses the deterministic pair (S_ρ, Π_ρ) throughout the computation.*

The worst-case sketch length and worst-case probe complexity are the maximum over all random tapes and all inputs. The average-case probe complexity with respect to an input distribution μ is

$$\mathbb{E}_\rho \mathbb{E}_{X\sim\mu}\big[T_{\Pi_\rho^{S_\rho}}(X)\big].$$

The success probability is taken over both $X \sim \mu$ and the random tape ρ.

Query-with-sketch and related models

As a natural information-theoretic model of computation, variants of the query-with-sketch model have been introduced and studied in different scenarios. The query-with-sketch model is a direct generalization of the decision tree model with *help bits* [29, 2]. 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 $k-1$. In addition, each output bit is computed using an individual decision tree in their case, whereas in our setting, all the output bits are decided in a single decision tree. The lower bound to succinct but unrestricted data structures for the partial sums problem [32] also reduces to lower bounds to data structures with *published bits*. However, the model used in this work is substantially different from ours. The queries are made to a data structure instead of the input elements.

2.2 The min-entropy lemma

We show that lower bounds to query-with-sketch algorithms can be reduced to lower bounding the conditional joint min-entropy of the output.

▶ **Definition 6** (joint min-entropy). *Let X be a random variable and Y an event. The* joint min-entropy *of X and Y is defined as $H_{\min}(X,Y) := \min_x -\log \Pr[X = x \wedge Y]$. When $\Pr[Y] = 0$, $H_{\min}(X,Y)$ is defined as $+\infty$.*

We relate the complexity of query-with-sketch algorithms with a special form of conditional joint min-entropy of the output. To formally state our result, let us give necessary definitions.

▶ **Definition 7.** *A* partial assignment *to a function $f : \mathcal{X}^N \to \mathcal{Y}$ of length q is an event of the form $(X_{i_1} = x_{i_1}, \dots, X_{i_q} = x_{i_q})$, where $i_1, \dots, i_q \in [N]$ are distinct coordinates and $x_{i_j} \in \mathcal{X}$ for every $j \in [q]$. We use $\mathcal{E}_{\leq q}$ to denote the set of all partial assignments of length at most q.*

It is important to note that a partial assignment is an event about the function itself and is irrelevant to the algorithm.

▶ **Lemma 8.** *Let $f : \mathcal{X}^N \to \mathcal{Y}$ be an arbitrary function, and $\mu_{\mathcal{X}^N}$ an distribution over $\mathcal{X}^N$. Fix Π^S to be a deterministic query-with-sketch algorithm for f with average-case probe complexity $\leq 0.1q$.*

In addition, suppose there exists a set $\mathsf{X} \subseteq \mathcal{X}^N$ of "good" inputs such that

$$\Pr_{X \leftarrow \mu_{\mathcal{X}^N}}[X \in \mathsf{X}] \geq 99/100.$$

Furthermore, assume there is a positive integer $r \geq 2$ such that for every partial assignment $\sigma \in \mathcal{E}_{\leq q}$ of length $\leq q$, there exists a subset of inputs $\mathsf{X}_\sigma \subseteq \mathcal{X}^N$ such that

$$\Pr_{X \leftarrow \mu_{\mathcal{X}^N}}[X \in \mathsf{X} \backslash \mathsf{X}_\sigma | \sigma] \leq 2^{-2r}$$

and the output of f remains highly uncertain:

$$H_{\min}(f(X), X \in \mathsf{X} \cap \mathsf{X}_\sigma | \sigma) > 2r$$

Then, for Π^S to achieve $\geq 99/100$ success probability given a random input from $\mu_{\mathcal{X}^N}$, its sketch length must be $|S| > r$.

For our purpose of proving a query-with-sketch lower bound for randomized algorithms given a fixed sequence of queries, we could prove distributional lower bounds for deterministic query-with-sketch algorithms instead, by the standard Yao's minimax principle.

The sets $\mathsf{X}, \mathsf{X}_\sigma$ describe good inputs where the min-entropy can be well-bounded. We note that X and X_σ are optional. If we set $\mathsf{X} = \mathsf{X}_\sigma = \mathcal{X}^N$ for every σ, there is no constraint on the inputs.

Proof of Lemma 8. Assume for the sake of contradiction that the sketch length is $\le r$. Without loss of generality, we let the sketch length be r. We will show that the success probability of Π^S cannot achieve $\ge 99/100$, which is a contradiction.

We characterize Π^S as a set of 2^r decision trees. Each input X is assigned to a decision tree T_S indexed by its sketch S. $f(X)$ is computed correctly if T_S given input X produces the correct answer $f(X)$. Each leaf of T_S can be characterized by a pair (y, σ), corresponding to the output at the leaf and the probe history. Let E_S denote the event that "X is assigned with sketch S", and E_q the event that "Π^S given input X makes no more than q probes". Then $\Pr[E_q] \ge 0.9$, or the average probe complexity would exceed $0.1q$. We have

$$
\begin{aligned}
&\Pr_{X \leftarrow \mu_{\mathcal{X}^N}}[\Pi^S(X) = f(X)] \\
=&\sum_{S \in \{0,1\}^r} \sum_{(y,\sigma) \text{ is a leaf of } T_S} \Pr[E_S \text{ and } \sigma \text{ and } f(X) = y] \\
\le& 0.11 + \sum_{S \in \{0,1\}^r} \sum_{\substack{(y,\sigma) \text{ is a leaf of } T_S \\ \text{of depth } \le q}} \Pr[E_S \text{ and } \sigma \text{ and } f(X) = y \text{ and } X \in \mathsf{X} \text{ and } E_q] \\
\le& 0.11 + \sum_{S \in \{0,1\}^r} \sum_{\substack{(y,\sigma) \text{ is a leaf of } T_S \\ \text{of depth } \le q}} \Pr[\sigma] \cdot \Pr[f(X) = y \text{ and } X \in \mathsf{X} \text{ and } E_q | \sigma] \\
\le& 0.11 + \sum_{S \in \{0,1\}^r} \sum_{\substack{(y,\sigma) \text{ is a leaf of } T_S \\ \text{of depth } \le q}} \Pr[\sigma] \cdot (\Pr[f(X) = y \text{ and } X \in \mathsf{X} \cap \mathsf{X}_\sigma \text{ and } E_q | \sigma] + 2^{-2r}) \\
<& 0.11 + \sum_{S \in \{0,1\}^r} \sum_{\substack{(y,\sigma) \text{ is a leaf of } T_S \\ \text{of depth } \le q}} \Pr[\sigma] \cdot 2^{-2r+1} \\
\le& 0.11 + \sum_{S \in \{0,1\}^r} 2^{-2r+1} \\
\le& 2^{-r+1} + 0.11 < 0.99
\end{aligned}
$$

The second line follows our assumption that $X \in \mathsf{X}$ with $\ge 99/100$ probability. The fourth and the fifth line follow our assumption to the conditional joint min-entropy and $\Pr[X \in \mathsf{X}_\sigma | \sigma]$. The sixth line follows from the fact that the partial assignments σ of different leaves of the same decision tree are disjoint events of the input. ◀

2.3 Reduction from data structures

We prove Lemma 3, which reduces data structure lower bounds to query-with-sketch lower bounds.

Proof to Lemma 3. The data structure D can be characterized as a sequence of l/m query-with-sketch algorithms for computing each block of m queries. The sketch to each algorithm is the (updated) redundancy after the end of its previous algorithm. The probe made by all the query-with-sketch algorithms are exactly the probe made by D to answer Q. In addition, the error probability of each query-with-sketch algorithm is at most the error probability of the data structure D. Besides, in both algorithms, the succinct space (sketch) exclusively depends on the input $X \in \mathcal{X}^n$. Among these algorithms, at least one of them has average-case probe complexity below average $\leq tm$. ◀

Through the above reduction and min-entropy lemma, to obtain a data structure lower bound one only need to (i) construct an input distribution $\mu_{\mathcal{X}^n}$, a subset X of $\mathcal{X}^N$, and subsets X_σ for every σ; (ii) show that for every partial assignment $\sigma \in \mathcal{E}_{\leq q}$, $H_{\min}(f(X), X \in \mathsf{X} \cap \mathsf{X}_\sigma | \sigma) > 2r$; and (iii) show that with high probability a random input falls in X (resp., falls in X_σ conditioned on σ, for every σ).

Comparison to a previous reduction

The query-with-sketch complexity of the *direct sum* of boolean problems was introduced and studied in [29, 2]. They showed that any l decision trees for computing the direct sum of the parity of m bits each, with $l-1$ help bits (i.e., the sketch), at least one of the decision trees must have a depth $\geq m$. Later, [13] got a data structure lower bound for prefix sum by a reduction from the help-bits lower bound for direct-sum parity. However, their reduction is specific to the prefix sum problem. Our reduction goes in a different way and is generally applicable to succinct and systematic data structures whose min-entropy can be bounded. In addition, our reduction is applicable to every problem f whose conditional min-entropy is bounded. The remainder of this paper will be devoted to showing the min-entropy bounds on the aforementioned problems. To compare our approach with the previous work, we explain the lower bound in [13] below.

In the prefix sum problem, the input is a length-n bit string X. The data structure, given an arbitrary index $k \in [n]$, should output the parity of $\sum_{i=1}^{k} X_i$. In [13], they showed that every succinct and systematic data structure for the prefix sum with r bits of redundancy requires $\Omega(n/(r+1))$ probes to the input bits. Their proof idea is as follows. Assume for contradiction that there exists a data structure with r bits of redundancy that makes $< \frac{n}{2(r+1)}$ bit probes for every given index $k \in [n]$. Split the input string X into $r+1$ blocks of about equal length, the parity of each block can be determined using $< n/(r+1)$ probes. However, [29] shows that even with r help bits (i.e., the sketch), to compute $r+1$ instances of parity of $n/(r+1)$ bits, at least one of the $r+1$ decision trees has depth $\geq n/(r+1)$.

3 Lower bounds to approximate matrix powering with sublinear-in-n redundancy

In this section, we study the approximate matrix powering problem. For this problem, we establish an $r \cdot t = \Omega(n^2)$ trade-off for data structures with redundancy $r < 0.25 n \ln n$ and amortized probe complexity t. The subsequent section addresses the case where $r \in [10 n \log n, n^2/1000]$, proving the same trade-off with a different technique. Let us begin by defining our rounding notation and then formally state the problem.

Approximate matrix powering (AMP) problem

Fix a constant $\alpha > 0$ and a number $k \geq 2$. The input is a substochastic symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$. The succinct and systematic data structure, given a pair of indices $(u, v) \in [n] \times [n]$, must produce a value ans such that $|\mathbf{M}^k[u, v] - ans| \leq 1/n^\alpha$.

▶ **Theorem 9** (Lower bound to AMP with sublinear-in-n redundancy)**.** *Fix $3 \leq k \leq n$ and a sufficiently large constant α. For every (possibly randomized) succinct and systematic data structure with $r \in (c\log^2 n, 0.25n \ln n)$ bits of redundancy that answers the whole approximated matrix power $\mathbf{M}^k$ with additive error $1/n^\alpha$ with $\geq 9/10$ probability, its total number of probes to $\mathbf{M}$ is $\Omega(n^4/r)$ at the worst case, where c is a large enough constant.*

We note that we prove a strong statement where in average, answering each query requires $\Omega(n^2/r)$ probes.

We give a distributional lower bound to the following nemesis input distribution.

Nemesis input distribution

Fix a large enough integer $n \geq 1$, and β, γ large enough constant parameters that depends on α. Let $\mathcal{M}_{\beta,\gamma}$ be a distribution of $n \times n$ matrices $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^\beta}\mathbf{M_U}$, where $\mathbf{M_U}$ is a random symmetric matrix whose upper triangle elements $\mathbf{M}[i,j]$ are independent and identically distributed (i.i.d.) uniformly from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$, for every $i \leq j$. We write $\mathbf{M} \leftarrow \mathcal{M}_{\beta,\gamma}$ (or simply $\mathbf{M} \leftarrow \mathcal{M}$) for $\mathbf{M}$ sampled from distribution $\mathcal{M}_{\beta,\gamma}$.

To apply Lemma 3, we also need to fix a sequence of queries (i.e., a permutation to the set of possible queries to the data structure $[n] \times [n]$). We will show that to answer each of the consecutive $3m := 3r/\ln n$ queries, a query-with-sketch algorithm either requires r bits of redundancy or $\Omega(n^2)$ probes to the input matrix.

Specifically, we fix the following sequence of queries, which is a concatenation of $(n-1)$ shifted main diagonals of $\mathbf{M}^k$ (excluding the main diagonal itself):

$$Q = (\underbrace{(1,2),(2,3),\ldots,(n,1)}_{\text{Shift 1: } v=u+1}, \underbrace{(1,3),(2,4),\ldots,(n,2)}_{\text{Shift 2: } v=u+2}, \ldots, \underbrace{(1,n),(2,1),\ldots,(n,n-1)}_{\text{Shift } n-1\text{: } v=u-1})$$

Formally, the sequence Q consists of $n-1$ blocks, where for every $i \in [n-1]$, the i-th block contains n queries (u,v) defined by:

- $u \in [n]$;
- $v = ((u+i-1) \bmod n) + 1$.

All we need from the construction is the following property.

▶ **Observation 10.** *For any window of $3j$ consecutive queries from Q (where $3j < n$), one can select a subset of j queries from this window such that all the $2j$ indices (that is, both the row u and column v from each selected query) are pairwise distinct.*

Proof. First, we show that any $3j$ consecutive queries span $3j$ different rows and $3j$ different columns.

Since $3j < n$, the $3j$ consecutive queries must belong to at most 2 different shifted main diagonals. For queries belong to the same diagonal, they clearly belong to different rows and columns.

Given two queries (u,v), (u',v') that respectively belong to the i-th and the $(i+1)$-th diagonals. If they belong to a window of length $3j < n$, we have $u' \leq u - 2$. They do not share the same row. In addition,

$$(v - v') \bmod n = (u - u' - 1) \bmod n \neq 0$$

when $u' \leq u - 2$. They do not share the same column as well.

The existence of the j queries follows from a greedy construction: we greedily pick unselected queries that have distinct rows and columns, until we cannot pick any more queries. Since the $3j$ elements belong to different rows and columns, picking one query (u,v) from it will disable at most 2 other queries from the window: at most one from column u, and at most one from row v. Therefore, one can always pick j many queries. ◀

Our query-with-sketch framework is built for exact problems. To apply it to AMP, which is an approximate problem (allowing additive error), we first prove our lower bound on an exact variant we call *rounded AMP*. In this problem, the algorithm must output the true value rounded to a precision of $1/n^{\alpha'}$.

A lower bound for this rounded problem implies a lower bound for the original AMP problem because the two problems are almost identical. They only differ on a small set of "boundary" inputs–inputs where the true value is too close to a rounding point to be distinguished. We formalize this by first showing that this property holds in each dimension. Then we apply a union bound over all the dimensions.

▶ **Definition 11** (rounding notation). *Fix a rounding parameter $\alpha' > 0$. For a real number $x \in \mathbb{R}$, we denote its rounded value to its nearest multiple of $1/n^{\alpha'}$ as $[x]_{\alpha'}$. For a real matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, we use $[\mathbf{M}]_{\alpha'}$ to denote the matrix where this rounding is applied element-wise.*

Rounded AMP problem

Fix a rounding constant $\alpha' > 0$ and a number $k \geq 2$. The input is a substochastic symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$. The succinct and systematic data structure, given a pair of indices $(u, v) \in [n] \times [n]$, must output $[\mathbf{M}^k[u, v]]_{\alpha'}$.

Given Lemma 3, we only need to show that for every $3m = 3r/\ln n$ consecutive queries, every query-with-sketch algorithm with r bits of sketch that computes the $3m$ queries with probability $\geq 99/100$, the algorithm makes $\Omega(n^2)$ probes in the worst case.

By Lemma 8, proving such a lower bound reduces to lower bounding the conditional min-entropy of the m query answers by $2r$, conditioned on assignments to the input matrix $\mathbf{M}$ of length $q := 10^{-4}n^2$. For this input distribution, we will construct a nontrivial good input set $\mathsf{M} \subseteq \mathbb{R}^{n\times n}$. And we will trivially set $\mathsf{M}_\sigma = \mathbb{R}^{n\times n}$ for every σ. Formally, the remainder of this section will be devoted to proving the following technical lemma.

▶ **Lemma 12.** *Fix $k \geq 3$ and constants α', β, γ such that $\gamma \geq 2$, $\beta \geq \gamma + 2$ and $\alpha' \geq 3\beta + 2\gamma + 3.6$. For every large enough integer n, there exists a set of "good" matrices $\mathsf{M} \subseteq \mathbb{R}^{n\times n}$ such that*

$$\Pr_{\mathbf{M}\leftarrow\mu_{\mathcal{X}^n}}[\mathbf{M} \in \mathsf{M}] \geq 1 - o(1).$$

In addition, fix $r \in (c\log^2 n, n\ln n/2)$, for every $3m = 3r/\ln n$ consecutive queries $(u_1, v_1), \ldots, (u_{3m}, v_{3m})$ from Q, and every partial assignment σ of length $q = 10^{-4}n^2$ to $\mathbf{M}$, we have

$$H_{\min}(([\mathbf{M}^k[u_1, v_1]]_{\alpha'}, \ldots, [\mathbf{M}^k[u_{3m}, v_{3m}]]_{\alpha'}), \mathbf{M} \in \mathsf{M}|\sigma) > 2r$$

where c is a constant that depends only on α', β, γ.

In addition, the following lemma shows that the outputs close to the boundary has a negligible measure. Its proof is deferred to Appendix A.

▶ **Lemma 13.** *Fix $k \geq 3$ and constants $\alpha, \alpha', \beta, \gamma$ such that $\gamma \geq 2$, $\beta \geq \gamma+2$, $\alpha' \geq 3\beta+2\gamma+3.6$ and $\alpha > \alpha' + 2$. For every large enough integer n, we have*

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}}\left[\exists u, v \in [n], a \in \mathbb{N}, \left|\mathbf{M}^k[u, v] - a/2n^{\alpha'}\right| \leq 1/n^{\alpha}\right] < o(1)$$

Given Lemma 12 and Lemma 13, our lower bound to approximate matrix powering follows.

Proof to Theorem 9. Let us first show that any data structure for rounded AMP with success probability $\geq 99/100$ either requires $\geq r$ redundancy or $\geq \Omega(n^4/r)$ worst-case probe complexity to answer queries in Q.

By Lemma 3, given the constructed sequence Q of queries, we only need to show that for every $3m$ consecutive queries from Q, every randomized query-with-sketch algorithm that computes the $3m$ queries with probability $\geq 99/100$ must pay $\geq r$ bits of sketch or $\geq 0.1q = 0.3tm$ average-case probe complexity.

By Yao's minimax principle, given our input distribution $\mathcal{M}$, we instead show that for every m queries and every deterministic query-with-sketch algorithm with success probability $\geq 99/100$ given a random input from $\mathcal{M}$, either r bits of sketch or $0.1q = 0.3tm$ average-case probe complexity is required. The lower bound follows from Lemma 8 and Lemma 12.

Now, we show that the same lower bound applies to the original AMP. We achieve it by a reduction from the rounded AMP. That is, we design a data structure D' for the rounded AMP given a data structure D for the AMP, with the same redundancy, probe complexity, and only an $o(1)$ increase in fail probability.

D' simply outputs the output of D rounded to the nearest multiple of $1/n^{\alpha'}$. By the correctness of D, for every query (u, v),

$$|\mathbf{M}^k[u, v] - D(u, v)| \leq 1/n^{\alpha}.$$

$\mathbf{M}^k[u, v]$ rounds to another value only if there exists a number $a \in \mathbb{N}$ such that one of the following two cases happen:

$$D(u, v) \leq a/2n^{\alpha'} \leq \mathbf{M}^k[u, v]$$

or

$$D(u, v) \geq a/2n^{\alpha'} \geq \mathbf{M}^k[u, v],$$

which happens only if

$$|\mathbf{M}^k[u, v] - a/2n^{\alpha'}| \leq |\mathbf{M}^k[u, v] - D(u, v)| \leq 1/n^{\alpha}.$$

However, by Lemma 13, this happens with $< o(1)$ probability. Therefore, the success probability of D' is at most $o(1)$ worse than the probability of D. The above lower bound also applies to AMP. ◀

3.1 Partial derivatives between the inputs and outputs

We are going to show that, for every $3m = 3r/\ln n$ consecutive queries $((u_1, v_1), \ldots, (u_m, v_m))$ that span m different rows and columns, and every partial assignment σ of length $q = 0.001n^2$, the query answers conditioned on the partial assignment remain a high min-entropy for good inputs from M. At a high level, every partial assignment fixes a proportion of input elements. We will show that, given every partial assignment, one can always pick $\Theta(r/\log n)$ elements from the unfixed input elements that form a "basis" to the output elements. That is, each fixed and possible output only corresponds to a small subset of inputs. We capture this by studying the Jacobian matrix between the output elements and the input elements. Note that such a basis does not always exist, where we construct a large enough good input set M that guarantees the existence of the basis.

To start with, we regard each output element $\mathbf{M}^k[u,v]$ as a function of n^2 input elements $\mathbf{M}[i,j]$, and study their partial derivatives $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]}$. Recall that $\mathbf{M} = \frac{n^2-1}{n^2}\mathbf{I} + \frac{1}{n^\beta}\mathbf{M_U}$ where each element $\mathbf{M_U}[i,j] = \mathbf{M_U}[j,i]$ of $\mathbf{M_U}$ is uniformly random from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. We ignore for now the fact that $\mathbf{M}$ is a symmetric matrix. The partial derivative under the symmetric constraint is simply $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} + \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[j,i]}$ when $i \neq j$.

We will use the following observation multiple times in the proof.

▶ **Observation 14.** *For every integer $n > 1$,*

$$\left(1 - \frac{1}{n}\right)^n \leq e^{-1} \leq \left(1 - \frac{1}{n}\right)^{n-1}$$

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

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

when $n \geq 2$.

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

when $n \geq 2$. ◀

To bound the partial derivatives $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]}$, we bound $\frac{\partial \mathbf{M_U}^k[u,v]}{\partial \mathbf{M_U}[i,j]}$ first.

▶ **Lemma 15.** *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 $3 \leq k \leq n$, for every element $[i,j]$ of $\mathbf{M_U}$ and every element $[u,v]$ of ${\mathbf{M_U}}^k$ such that $u \neq i$ and $j \neq v$. When $k \geq 3$, we have:*

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

Specifically, when $k = 3$,

$$\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \mathbf{M_U}[u,i]\mathbf{M_U}[j,v]$$

Proof of Lemma 15. Note that

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

When $u \neq i$ and $j \neq v$, the partial derivative is

$$\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} = \sum_{g=1}^{k-2}\left(\sum_{u=h_0,h_1,\ldots,h_g=i,h_{g+1}=j,\ldots,h_{k-1},h_k=v}\left(\prod_{t=0,t\neq g}^{k-1}\mathbf{M_U}[h_t,h_{t+1}]\right)\right)$$

In this equation the appearance of $\mathbf{M_U}[i,j]$ is canceled at position g in the summation of the product. This directly gives the expression for $k=3$. Since for all $i,j \in [n]$, $\mathbf{M_U}[i,j] = \mathbf{M_U}[j,i] \in [\frac{1}{n}-\frac{1}{n^\gamma},\frac{1}{n}]$, 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 bounds all elements are equivalent. Since g can only be picked from $\{1,2,...,k-2\}$, there are $(k-2)n^{k-3}$ distinct sequences $h_0,\ldots,h_k$. We get

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

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

▶ **Lemma 16.** *Let $\gamma,\beta \geq 2$, 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 $3 \leq k \leq n$, every element (i,j) of $\mathbf{M}$ and every element (u,v) of $\mathbf{M}^k$ such that u,v,i,j are pairwise distinct, we have*

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \binom{k}{3}\frac{(n^2-1)^{k-3}}{n^{2k-6+2\beta}}\left(\mathbf{M_U}[u,i]\mathbf{M_U}[j,v]+\mathbf{M_U}[u,j]\mathbf{M_U}[i,v]\right)+O(k^5/n^{2+3\beta}).$$

Proof. Notice that

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

and

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

While u,v,i,j are pairwise distinct, $\frac{\partial {\mathbf{M_U}}^t[u,v]}{\partial \mathbf{M_U}[i,j]}$ is 0 when $t \leq 2$. Therefore,

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \binom{k}{3}\frac{(n^2-1)^{k-3}}{n^{2k-6+2\beta}}\mathbf{M_U}[u,i]\mathbf{M_U}[j,v]+O(k^5/n^{2+3\beta}).$$ ◀

3.2 Jacobian matrix and singular value concentration bounds

While the previous section studies the partial derivatives between individual input elements and output elements of different rows and columns. We extend it to the interplay between multiple input elements and multiple output elements, which formalizes as the Jacobian matrix. We will show that, fix a number $m \le n/4$, for every m input elements and m output elements whose column indices and row indices are pairwise distinct, the largest $0.9m$ singular values of their Jacobian matrix are controlled with high probability. We use a recent work [42] on singular values concentration bounds to achieve it.

▶ **Definition 17.** *Fix $m \le n/4$, m input indices $(i_1, j_1), \ldots, (i_m, j_m)$, and m output indices $(u_1, v_1), \ldots, (u_m, v_m)$. When $i_1 \ne j_1, \ldots, i_m \ne j_m$, their Jacobian matrix is defined as*

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \mathbf{M}^k[u_1,v_1]}{\partial \mathbf{M}[i_1,j_1]} + \frac{\partial \mathbf{M}^k[u_1,v_1]}{\partial \mathbf{M}[j_1,i_1]} & \cdots & \cdots & \frac{\partial \mathbf{M}^k[u_m,v_m]}{\partial \mathbf{M}[i_1,j_1]} + \frac{\partial \mathbf{M}^k[u_m,v_m]}{\partial \mathbf{M}[j_1,i_1]} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial \mathbf{M}^k[u_1,v_1]}{\partial \mathbf{M}[i_m,j_m]} + \frac{\partial \mathbf{M}^k[u_1,v_1]}{\partial \mathbf{M}[j_m,i_m]} & \cdots & \cdots & \frac{\partial \mathbf{M}^k[u_m,v_m]}{\partial \mathbf{M}[i_m,j_m]} + \frac{\partial \mathbf{M}^k[u_m,v_m]}{\partial \mathbf{M}[j_m,i_m]} \end{pmatrix}$$

Given an $m \times m$ Hermitian matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$, we denote by $\lambda_1(\mathbf{A}) \ge \cdots \ge \lambda_m(\mathbf{A})$ eigenvalues of $\mathbf{A}$.

The singular values $\sigma_1(\mathbf{A}), \ldots, \sigma_m(\mathbf{A})$ of a (possibly non-Hermitian) matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$, are the eigenvalues of $\sqrt{\mathbf{A}^T\mathbf{A}}$ arranged in non-increasing order.

We will use the following standard tools from matrix analysis.

▶ **Proposition 18** (Weyl's inequality for singular values [22])**.** *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times m}$ be general matrices. For every $k \in [m]$,*

$$|\sigma_k(\mathbf{A} + \mathbf{B}) - \sigma_k(\mathbf{A})| \le \|\mathbf{B}\|_{op}$$

where $\|\mathbf{B}\|_{op} = \sigma_1(\mathbf{B})$ is the operator norm.

▶ **Proposition 19** ([36])**.** *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times m}$ where $\mathrm{rank}(\mathbf{B}) = 1$. For every $k \in [m]$,*

$$\sigma_{k+1}(\mathbf{A}) \le \sigma_k(\mathbf{A} + \mathbf{B}) \le \sigma_{k-1}(\mathbf{A})$$

where we denote $\sigma_0(\mathbf{A}) = +\infty$ and $\sigma_{m+1}(\mathbf{A}) = 0$.

▶ **Proposition 20** (Geršgorin circle theorem [22])**.** *Let $\mathbf{A} \in \mathbb{C}^{m\times m}$. Let $R_i(\mathbf{A}) = \sum_{j\ne i} |\mathbf{A}[i,j]|$ denote the absolute row sums of $\mathbf{A}$ other than $\mathbf{A}[i,i]$. And consider the m Geršgorin discs*

$$\{z \in \mathbb{C} : |z - \mathbf{A}[i,i]| \le R_i(\mathbf{A})\}$$

for $i \in [m]$. The eigenvalues of $\mathbf{A}$ are in the union of Geršgorin discs

$$G(A) = \bigcup_{i=1}^{m} \{z \in \mathbb{C} : |z - \mathbf{A}[i,i]| \le R_i(\mathbf{A})|\}.$$

We will use the following concentration bound on the intermediate singular values of a random matrix [34, 42]. We adapt their result in our setting and show the following in Appendix B.

▶ **Lemma 21.** *For $i \in \{1, 2\}$, let X_i, Y_i, Z_i, W_i be i.i.d. uniform on some symmetric interval (the interval may depend on i). Define*

$$U_1 := X_1 + Y_1 + Z_1 + W_1, \qquad U_2 := X_2Y_2 + Z_2W_2.$$

Let μ_i be the distribution of U_i.

Fix μ to be either μ_1 or μ_2. Let m a large enough number, and $\mathbf{A} \in \mathbb{R}^{m\times m}$ be a matrix whose elements are i.i.d. drawn from μ. There exists constants $0 < C_1 < C_2$ and $C_3 > 0$ such that for all ℓ between 1 and m,

$$\Pr\left[\frac{C_1\ell}{\sqrt{m}} \leq \sigma_{m+1-\ell}(\mathbf{A}) \leq \frac{C_2\ell}{\sqrt{m}}\right] \geq 1 - \exp(-C_3\ell)$$

Basically, a random matrix of the above form has high singular values to a large proportion of its dimensions with high probability. Our Jacobian matrix is a combination of scaled, shifted and perturbed random matrices of the above form. Formally, we give the following concentration bound on our Jacobian matrix.

▶ **Lemma 22.** *Let $\beta, \gamma \geq 2$ be constants such that $\beta \geq \gamma + 2$. For every large enough n and every $m \leq n/4$. Fix a sequence of indices $(i_1, j_1), \ldots, (i_m, j_m)$ of $\mathbf{M}$ and a sequence of indices $(u_1, v_1), \ldots (u_m, v_m) \in [n] \times [n]$ of $\mathbf{M}^k$ such that $i_1, j_1, \ldots, i_m, j_m, u_1, v_1, \ldots, u_m, v_m$ are $4m$ pairwise distinct numbers. With probability $\geq 1 - \exp(-\Omega(m))$, we have $\sigma_{0.9m}(\mathbf{J}) = \Omega(k^3\sqrt{m}/n^{2\beta+\gamma+1})$.*

In addition, the above holds true even if each element of $\mathbf{J}$ is perturbed additively by at most $O(k^5/n^{2+3\beta})$.

Proof. Let $K := \binom{k}{3}\frac{(n^2-1)^{k-3}}{n^{2k-6+2\beta}} = \Theta(k^3/n^{2\beta})$. We define four auxiliary matrices whose singular values are bounded, and their combination is exactly $\mathbf{J}$. Notice that the expectation of each element of $\mathbf{M_U}$ is exactly $E := \frac{1}{n} - \frac{1}{2n^\gamma}$. Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B} \in \mathbb{R}^{m\times m}$ such that for every $a, b \in [m]$,

$$\mathbf{A}_1[a,b] := 6\sqrt{2}n^{2\gamma}\left(\left(\mathbf{M_U}[u_b, i_a] - E\right)\left(\mathbf{M_U}[j_a, v_b] - E\right) + \left(\mathbf{M_U}[u_b, j_a] - E\right)\left(\mathbf{M_U}[i_a, v_b] - E\right)\right),$$

$$\mathbf{A}_2[a,b] := \sqrt{3}n^\gamma\left(\mathbf{M_U}[u_b, i_a] + \mathbf{M_U}[j_a, v_b] + \mathbf{M_U}[u_b, j_a] + \mathbf{M_U}[i_a, v_b] - 4E\right),$$

$$\mathbf{B}[a,b] := \frac{\partial \mathbf{M}^k[u_b, v_b]}{\partial \mathbf{M}[i_a, j_a]} + \frac{\partial \mathbf{M}^k[u_b, v_b]}{\partial \mathbf{M}[j_a, i_a]} - K(\mathbf{M_U}[u_b, i_a]\mathbf{M_U}[j_a, v_b] + \mathbf{M_U}[u_b, j_a]\mathbf{M_U}[i_a, v_b]).$$

By Lemma 16, we can rewrite the Jacobian matrix as

$$\mathbf{J} = \frac{K}{6\sqrt{2}n^{2\gamma}}\mathbf{A}_1 + \frac{K}{\sqrt{3}n^\gamma}E\mathbf{A}_2 + 2KE^2\mathbf{1}\mathbf{1}^T + \mathbf{B}$$

where $\mathbf{1}$ denotes the all-1 vector of length m. One can verify that both $\mathbf{A}_1, \mathbf{A}_2$ are random matrices satisfying the condition in Lemma 21. Both $\mathbf{A}_1, \mathbf{A}_2$ have i.i.d. random elements because $\mathbf{M_U}$ is a uniformly random symmetric matrix, and the indices i, j, u, v do not repeat.

By Geršgorin circle theorem (Proposition 20), for every $i \in [m]$, we have

$$\sigma_i(\mathbf{B}) = \sqrt{\lambda_i(\mathbf{B}^T\mathbf{B})} \leq O(mk^5/n^{2+3\beta}). \tag{3}$$

We know from Lemma 21 that for every $\mathbf{A} = \mathbf{A}_1, \mathbf{A}_2$.

$$\Pr\left[C_1\sqrt{m} \leq \sigma_1(\mathbf{A}) \leq C_2\sqrt{m}\right] \geq 1 - \exp(-C_3 m)$$

and for every $l \in [m]$,

$$\Pr\left[C_1 l/\sqrt{m} \leq \sigma_{m+1-l}(\mathbf{A}) \leq C_2 l/\sqrt{m}\right] \geq 1 - \exp(-C_3 l).$$

By Proposition 19,

$$\sigma_{0.9m}\left(\frac{K}{\sqrt{3}n^\gamma}E\mathbf{A}_2 + 2KE^2\mathbf{1}\mathbf{1}^T\right) \geq \frac{K}{\sqrt{3}n^\gamma}E\sigma_{0.9m+1}(\mathbf{A}_2)$$

By Weyl's inequality (Proposition 18) and triangle inequality, we have

$$\sigma_{0.9m}(\mathbf{J}) \geq \frac{K}{\sqrt{3}n^{\gamma}} E \cdot \sigma_{0.9m+1}(\mathbf{A}_2) - \frac{K}{6\sqrt{2}n^{2\gamma}} \cdot \sigma_1(\mathbf{A}_1) - \sigma_1(\mathbf{B}).$$

By inequality (3) and the concentration bounds on the singular values of $\mathbf{A}_1, \mathbf{A}_2$, the following holds with high probability

$$\sigma_{0.9m}(\mathbf{J}) \geq \frac{K}{\sqrt{3}n^{\gamma}} E \cdot 0.1 C_1 \sqrt{m} - \frac{K}{6\sqrt{2}n^{2\gamma}} \cdot C_2 \sqrt{m} - O(mk^5/n^{2+3\beta})$$

which implies that $\sigma_{0.9m}(\mathbf{J}) = \Omega(K\sqrt{m}/n^{\gamma+1}) = \Omega(k^3\sqrt{m}/n^{2\beta+\gamma+1})$ when $\beta \geq \gamma + 2$.

The above holds true even if each element of $\mathbf{J}$ is perturbed additively by at most $O(k^5/n^{2+3\beta})$, since the absolute values of elements of $\mathbf{B}$ remains $O(k^5/n^{2+3\beta})$. ◀

3.3 From singular values to the min-entropy: proof to Lemma 12

We have prepared to prove our technical lemma, which is to show that there exists a good input set M such that the joint min-entropy conditioned on any partial assignment of length $\leq q = 10^{-4}n^2$ is high. Our good input set M will be defined as inputs where the Jacobian matrices have well-controlled singular values. Before that, we partition the set of input elements into groups of *core* elements of size m, so that for every partial assignment σ, we can always pick one such group of core elements that form a basis of the output elements.

▶ **Definition 23.** *For every window of $3m$ consecutive queries from Q, fix once and for all one subset of m queries whose $2m$ row and column indices are pairwise distinct. Such a subset exists by Observation 10. Let $(u_1, v_1), \ldots, (u_m, v_m)$ denote this fixed subset for the window. We define $\mathcal{C}$ to be an arbitrary collection of disjoint sequences from*

$$([n] \setminus \{u_1, \ldots, u_m, v_1, \ldots, v_m\}) \times ([n] \setminus \{u_1, \ldots, u_m, v_1, \ldots, v_m\})$$

such that, for every $((i_1, j_1), \ldots, (i_m, j_m)) \in \mathcal{C}$, the indices $i_1, j_1, \ldots, i_m, j_m$ are pairwise distinct. Moreover, we choose $\mathcal{C}$ so that

$$|\mathcal{C}| \geq \left\lfloor \frac{(n-2m)(n-2m-1)}{3m} \right\rfloor.$$

We call $\mathcal{C}$ the candidate core set *associated with this window.*

The existence of such a partition is guaranteed by the same construction as the query sequence Q, where by Observation 10, the permutation of indices we defined as Q guarantees that no m consecutive elements share the same column or the same row. Given $\mathcal{C}$, we can define our good input set M.

▶ **Definition 24.** *We define M as the set of inputs satisfying the following condition. For every window of $3m$ consecutive queries from Q, let $(u_1, v_1), \ldots, (u_m, v_m)$ be the fixed subset of queries chosen in Definition 23, and let $\mathcal{C}$ be its candidate core set. Then for every sequence $((i_1, j_1), \ldots, (i_m, j_m)) \in \mathcal{C}$, the Jacobian matrix between $\mathbf{M}[i_1, j_1], \ldots, \mathbf{M}[i_m, j_m]$ and $\mathbf{M}^k[u_1, v_1], \ldots, \mathbf{M}^k[u_m, v_m]$ satisfies*

$$\sigma_{0.9m}(J) = \Omega\left(\frac{k^3\sqrt{m}}{n^{2\beta+\gamma+1}}\right).$$

▶ **Lemma 25.** *There exists a constant c such that for every $m \in [c \log n, n/4]$, $\Pr_{\mathbf{M} \leftarrow \mathcal{M}}[\mathbf{M} \in \mathsf{M}] > 1 - o(1)$.*

Proof. Fix a window of $3m$ consecutive queries from Q, and consider the fixed subset of m queries and the candidate core set $\mathcal{C}$ from Definition 23. By Lemma 22, for every fixed sequence in $\mathcal{C}$, the corresponding Jacobian matrix has the desired singular-value bound with probability at least $1 - \exp(-\Omega(m))$.

It remains to union bound over all events that appear in the definition of $\mathcal{M}$. There are $O(n^2)$ windows of $3m$ consecutive queries in Q. For each such window, the candidate core set $\mathcal{C}$ contains at most $O(n^2/m)$ disjoint sequences. Hence the total number of Jacobian matrices over which we union bound is at most $O(n^4/m)$, which is polynomial in n. Therefore,

$$\Pr[\mathbf{M} \notin \mathcal{M}] \leq O(n^4/m) \cdot \exp(-\Omega(m)) = o(1),$$

provided $m \geq c \log n$ for a sufficiently large constant c. ◀

Given guarantees that the Jacobian matrices have high singular values, given an arbitrary partial assignment σ of length $\leq q = 10^{-4}n^2$, there always exists a sequence of input indices from $\mathcal{C}$ where a large proportion of these input elements are unfixed by σ. Those elements, due to their nice properties in their Jacobian matrix with the output, form a basis to the output. In this way, the entropy of the input preserves to the output. Before proving Lemma 12, we need the following definition and basic tools.

▶ **Definition 26.** *Given an arbitrary partial assignment σ, the* core indices *of $\mathbf{M}$ is a length-$(m' := 0.9m)$ subsequence $((i_1, j_1), \ldots, (i_{m'}, j_{m'}))$ of an arbitrary sequence from $\mathcal{C}$, such that $\mathbf{M}[i_1, j_1], \ldots, \mathbf{M}[i_{m'}, j_{m'}]$ and $\mathbf{M}[j_1, i_1], \ldots, \mathbf{M}[j_{m'}, i_{m'}]$ are not fixed by σ. We use $(\mathbf{M})_{\text{core}} := (\mathbf{M}[i_1, j_1], \ldots, \mathbf{M}[i_{m'}, j_{m'}])$ to denote the vector representation of core elements. In addition, we use $(\mathbf{M}^k)_{\text{out}} := (\mathbf{M}^k[u_1, v_1], \ldots, \mathbf{M}^k[u_m, v_m])$ to denote the vector representation of the selected output elements. Also, we always use $[\mathbf{v}]_{\alpha'}$ to denote the vector where each element is rounded to its nearest multiple of $1/n^{\alpha'}$.*

We will show that, for every fixed (and rounded) output, the large singular values of the Jacobian matrix between core elements and output elements forces $\Omega(m)$ dimensions of the input to be concentrated. The min-entropy will be reduced to measuring the volume of a hyperrectangle in the uniform input space.

▶ **Proposition 27** (Cauchy interlacing theorem for singular values [22])**.** *Let $\mathbf{J} \in \mathbb{R}^{m \times n}$ a matrix where $m \leq n$, and $\mathbf{J}'$ a matrix obtained by deleting one row from it. Then*

$$\sigma_1(\mathbf{J}) \geq \sigma_1(\mathbf{J}') \geq \sigma_2(\mathbf{J}) \geq \sigma_2(\mathbf{J}') \geq \cdots \geq \sigma_{m-1}(\mathbf{J}) \geq \sigma_{m-1}(\mathbf{J}').$$

▶ **Proposition 28** (Lagrange's mean value theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and differentiable function. For every two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, there exists $\mathbf{c} = (1 - t_0)\mathbf{a} + t_0\mathbf{b}$ for some $t_0 \in (0, 1)$ such that*

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a})$$

where $\nabla f(\mathbf{c})$ is the gradient of f evaluated at the point $\mathbf{c}$.

Proof to Lemma 12. We have shown in Lemma 25 that a random input drawn from $\mathcal{M}$ is a good input with high probability. What remains is to show that the conditional joint min-entropy of the output is high for every partial assignment σ of bounded length.

Given a partial assignment σ of length $\leq q = 10^{-4}n^2$, there exists a sequence $((i_1, j_1), \ldots, (i_m, j_m)) \in \mathcal{C}$ where at most $0.1m$ elements of $\mathbf{M}$ or their symmetric elements are revealed by σ. Otherwise there would be $> \frac{1}{2} \cdot \lfloor \frac{(n-2m)(n-2m-1)}{3m} \rfloor \cdot 0.1m > q$ elements revealed by σ. We fix an arbitrary such sequence, and the core indices $(i_1, j_1), \ldots, (i_{m'}, j_{m'})$ be its subsequence obtained by removing all the elements and the symmetric elements fixed by σ. For every possible output vector $\mathbf{y}$, we have

$$\Pr_{\mathbf{M}\leftarrow\mathcal{M}}\left[[(\mathbf{M}^k)_{\text{out}}]_{\alpha'} = \mathbf{y} \text{ and } \mathbf{M} \in \mathsf{M}\middle|\, \sigma\right] \leq \max_{\sigma'} \Pr_{\mathbf{M}\leftarrow\mathcal{M}}\left[[(\mathbf{M}^k)_{\text{out}}]_{\alpha'} = \mathbf{y} \text{ and } \mathbf{M} \in \mathsf{M}\middle|\, \sigma'\right] \quad (4)$$

where the maximum is taken over all partial assignments σ' on all the input elements other than core elements and their symmetric elements, such that σ' is consistent with σ. We denote $f : \mathbb{R}^{m'} \to \mathbb{R}^m$ as the function of $(\mathbf{M}^k)_{\text{out}}$ given core elements, and all the other input elements fixed by σ'. And we denote by $f_i(\cdot)$ the i-th element of the output vector.

Let $\|\cdot\|_\infty$ denote max norm, i.e., the maximum absolute value of all elements. Fix an arbitrary instance $\mathbf{M}'$ that is consistent σ' such that $[(\mathbf{M'}^k)_{\text{out}}]_{\alpha'} = \mathbf{y}$. Then, for matrices $\mathbf{M}$ consistent with σ' and with the same output, we have

$$\|f((\mathbf{M})_{\text{core}}) - f((\mathbf{M}')_{\text{core}})\|_\infty \leq 1/n^{\alpha'}.$$

By the mean value theorem, there exists m vectors $\mathbf{c}_1, \ldots, \mathbf{c}_m$, such that for every $\ell \in [m]$,

$$f_\ell((\mathbf{M})_{\text{core}}) - f_\ell((\mathbf{M}')_{\text{core}}) = \nabla f_\ell(\mathbf{c}_\ell) \cdot ((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})$$

where $\nabla f_\ell(\mathbf{c}_\ell)$ is exactly the ℓ-th column of the Jacobian matrix evaluated at $\mathbf{c}_\ell$. We let $\mathbf{J} \in \mathbb{R}^{m'\times m}$ be the matrix where the ℓ-th column is exactly $\nabla f_\ell(\mathbf{c}_\ell)$ for each $\ell \in [m]$. Then we have

$$1/n^{\alpha'} \geq \|f((\mathbf{M})_{\text{core}}) - f((\mathbf{M}')_{\text{core}})\|_\infty = \|\mathbf{J}^T((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})\|_\infty. \quad (5)$$

Because by Lemma 16, when given σ' fixed, the Jacobian matrix is almost fixed but the $O(k^5/n^{2+3\beta})$ term. We have

$$\|\mathbf{J} - \mathbf{J}(\mathbf{M})\|_\infty, \|\mathbf{J} - \mathbf{J}(\mathbf{M}')\|_\infty \leq O(k^5/n^{2+3\beta})$$

for $\mathbf{J}(\mathbf{M}), \mathbf{J}(\mathbf{M}')$ respectively the Jacobian matrix evaluated at $\mathbf{M}$ and $\mathbf{M}'$. The singular value bounds given in Lemma 21 also apply to $\mathbf{J}$, since, the as stated in Lemma 21, the bound applies even when $\mathbf{J}$ is a matrix obtained by perturbing every element of a Jacobian matrix by at most $O(k^5/n^{2+3\beta})$.

Let $m'' := 0.8m$, by Lemma 21 and Cauchy interlacing theorem, $\mathbf{J} \in \mathbb{R}^{m'\times m}$ between the core elements in $\mathbf{M}$ and $(\mathbf{M}^k)_{\text{out}}$ satisfies: for every $\ell \in [0.8m]$,

$$\sigma_\ell(\mathbf{J}) = \Omega(k^3\sqrt{m}/n^{2\beta+\gamma+1}). \quad (6)$$

Consider the singular value decomposition $\mathbf{J} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ for orthogonal matrices $\mathbf{U} \in \mathbb{R}^{m'\times m'}, \mathbf{V} \in \mathbb{R}^{m\times m}$ and a diagonal matrix $\mathbf{\Sigma}$ where $\mathbf{\Sigma}[\ell,\ell] = \sigma_\ell(\mathbf{J})$ for every $\ell \in [m']$. Denote by $\mathbf{u}_1, \ldots, \mathbf{u}_{m'}$ the columns of $\mathbf{U}$, which are orthonormal vectors. Let $((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}}) = \sum_{\ell=1}^{m'} \alpha_\ell \mathbf{u}_\ell$ for some $\alpha_1, \ldots, \alpha_{m'} \in \mathbb{R}$. Then we have

$$\begin{aligned}
\|\mathbf{J}^T((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})\|_\infty &\geq \frac{1}{\sqrt{m}}\|\mathbf{J}^T((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})\|_2 \\
&= \frac{1}{\sqrt{m}}\|\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T(\sum_{\ell=1}^{m'} \alpha_\ell \mathbf{u}_\ell)\|_2 \\
&= \frac{1}{\sqrt{m}}\|\mathbf{\Sigma}^T\mathbf{U}^T(\sum_{\ell=1}^{m'} \alpha_\ell \mathbf{u}_\ell)\|_2 \\
&= \sqrt{\frac{\sum_{\ell=1}^{m'} \sigma_\ell(\mathbf{J})^2\alpha_\ell^2}{m}}
\end{aligned}$$

where the first inequality is by Cauchy-Schwarz, the third line is because the L_2 norm remains the same under orthogonal transformations.

Notice that for every $\ell \in [m']$, $\alpha_\ell = \mathbf{u}_\ell \cdot ((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})$. Combining the above with inequality (5) and equation (6), for every $\ell \in [m'']$, we have

$$((\mathbf{M})_{\text{core}} - (\mathbf{M}')_{\text{core}})^T \mathbf{u}_\ell \leq \frac{\sqrt{m}}{n^{\alpha'}\sigma_\ell(\mathbf{J})} \leq O(n^{2\beta+\gamma-\alpha'+1}/k^3).$$

Therefore, given $\mathbf{y}$ and σ', the possible input set is contained in the solution set of the above m'' inequalities, whose volume is proportional to the volume of the hyperrectangle with m'' dimensions concentrated in the uniform input space.

Notice that the total volume of possible inputs $\mathbf{M}$ given σ' is $1/n^{(\beta+\gamma)m'}$. For every pair of different possible matrices consistent to σ', their maximum L_2 distance is at most $\sqrt{m'}/n^{\beta+\gamma}$. Therefore, the maximum volume of the hyperrectangle is upper-bounded by

$$(O(n^{2\beta+\gamma-\alpha'+1}/k^3))^{m''} \cdot (\sqrt{m'}/n^{\beta+\gamma})^{m'-m''}.$$

Recall that $m'' = 0.8m$, and $m' = 0.9m$. By inequality (4),

$$\begin{aligned}
&\Pr_{\mathbf{M}\leftarrow\mathcal{M}}\left[[(\mathbf{M}^k)_{\text{out}}]_{\alpha'} = \mathbf{y} \text{ and } \mathbf{M} \in \mathsf{M} \middle| \sigma\right] \\
\leq& (O(n^{2\beta+\gamma-\alpha'+1}/k^3))^{m''} \cdot (\sqrt{m}/n^{\beta+\gamma})^{m'-m''} \cdot n^{(\beta+\gamma)m'} \\
\leq& \exp(m''(2\beta+\gamma-\alpha'+1)\ln n) \cdot \exp((m'-m'')(0.5-\beta-\gamma)\ln n) \cdot \exp(m'(\beta+\gamma)\ln n) \\
\leq& \exp((2.4\beta+1.6\gamma-0.8\alpha'+0.85)m\ln n) \\
\leq& \exp(-2r)
\end{aligned}$$

while $\alpha' \geq 3\beta + 2\gamma + 3.6$ and $m = \frac{r}{\ln n} < n$.

Since the conditional min-entropy of m selected output elements is at least $2r$, the conditional min-entropy of all the $3m$ output elements is also at least $2r$. ◀

4 Lower bounds to approximate matrix powering with superlinear-in-n redundancy

The above approach does not easily extend to proofs to $r \cdot t = \Omega(n^2)$ lower bounds for $r = \omega(n \ln n)$ because the Jacobian matrix of $\omega(n)$ core elements and output elements can not be easily controlled. Our analysis above relies on the fact that the core elements, output elements, and their symmetric elements do not share rows and columns. To that end, we use a different input distribution which leads to the same trade-off for $r \in [10n \log n, n^2/1000]$. Notably, the lower bound derived using this distribution also implies a lower bound for the fundamental *counting set intersection* problem in data structures.

Nemesis input distribution

Fix $n \geq 1$ a large enough integer. Fix a partition $V_1 := \{1, 2, \ldots, n/3\}, V_2 := \{n/3 + 1, \ldots, 2n/3\}, V_3 := \{2n/3 + 1, \ldots, n\}$ to $[n]$. Let $\mathcal{M}$ be a distribution of $n \times n$ symmetric matrices $\mathbf{M}$ with the following conditions.

1. For every $(u, v) \in (V_1 \times V_2) \cup (V_2 \times V_3)$,

$$\mathbf{M}[u,v] = \mathbf{M}[v,u] = \begin{cases} 1/n^3 & \text{with independent probability } 1/2 \\ 0 & \text{with independent probability } 1/2. \end{cases}$$

2. For every $u \in [n]$, $\mathbf{M}[u, u] = 1 - 1/n^2$.
3. For every other $(u, v) \in [n] \times [n]$, $\mathbf{M}[u, v] = 0$.

In other words, its underlying graph is a three-layered undirected graph with heavy lazy transitions. For each pair of vertices of consecutive layers, we connect them with independent probability $1/2$. We have the following observation.

▶ **Observation 29.** *For every pair of vertices $(u, v) \in V_1 \times V_3$ and every $2 \le k \le n$, if there are exactly a length-2 simple paths from u to v,*

$$\mathbf{M}^k[u, v] = \binom{k}{2} \left(\frac{n^2 - 1}{n^2}\right)^{k-2} \frac{a}{n^6} + O(k^4/n^9).$$

Proof. Let $a_2, \dots, a_k$ respectively denote the number of length-$2, \dots, k$ walks from u to v without self loops. Then each $a_i = O(n^{i-1})$. Specifically, for every odd i, $a_i = 0$. Because the underlying graph of M without self-loops is a bipartite graph where u, v belong to the same part. We get

$$\mathbf{M}^k[u, v] = \binom{k}{2} \left(\frac{n^2 - 1}{n^2}\right)^{k-2} \frac{a_2}{n^6} + \sum_{i=4}^{k} \binom{k}{i} \left(\frac{n^2 - 1}{n^2}\right)^{k-i} \frac{a_i}{n^{3i}}$$

where the summation is upper bounded by the sum of a geometric series, and is dominated by the case $i = 4$. ◀

Since $k^4/n^9 = o(k^2/n^6)$, approximating the power of $\mathbf{M}$ can be used to answer the number of length-2 paths between vertices from V_1 and V_3 exactly. In this way, we get rid of the matrix, and are able to focus on a clean combinatorial task. In fact, we will show that even computing the parity of $\mathbf{M}^k[u, v]$ requires high data structure complexity.

This problem can also be rephrased in terms of counting set intersection: for every pair of vertices $(u, v) \in V_1 \times V_3$, the number of length-2 paths from u to v is exactly the size of the intersection of their neighbor sets.

The sequence of queries that we construct is similar to the above. We let it to be a concatenation of $n/3$ disjoint bijections between V_1 and V_3:

$$Q = (\underbrace{(1, 2n/3 + 1), (2, 2n/3 + 2), \dots, (n/3, n)}_{\text{identity bijection}}, \dots, \underbrace{(1, n), (2, 2n/3 + 1), \dots, (n/3, n - 1)}_{\text{shifted bijection}})$$

Formally, for every $i \in [n/3]$, the i-th bijection is exactly a sequence of pairs (u, v) where

- $u \in V_1$;
- $v = (u + i - 2) \bmod (n/3) + 2n/3 + 1$.

The following observation is immediate from our construction.

▶ **Observation 30.** *For every consecutive $m \ge n$ queries in Q, there does not exist an index $v \in [n]$ that is covered more than $\lceil \frac{m}{n/3} \rceil$ times.*

Even though the input distribution is a clean uniform distribution, the conditional min-entropy remains non-trivial to analyze. Our strategy, analogous to the previous section, is to find a "basis" from the input elements to the outputs. We show that we can always find such a basis for a large set M of "good inputs".

▶ **Definition 31.** *Let* $\hat{\mathbf{M}} \in \mathbb{F}_2^{n\times n}$ *denote the binary matrix where each element* $\hat{\mathbf{M}}[i,j] = 1$ *if and only if* $\mathbf{M}[i,j] \neq 0$.

Let $\mathsf{M} \subseteq \mathbb{R}^{n\times n}$ *be a set of "good" input matrices* $\mathbf{M}$ *such that for every* $d \in [10\log n, n/3]$, *and every* $V_1' \subseteq V_1, V_2' \subseteq V_2$, $|V_1'| = |V_2'| = d$, *the submatrix has a high rank:*

$$\mathrm{rank}(\hat{\mathbf{M}}[V_1', V_2']) \geq d/2.$$

Notice that the submatrix is a uniformly random square matrix, where there are classic results bounding its rank.

▶ **Proposition 32** (page 38 of [3]; see also page 1 of [12])**.** *Let* $\hat{\mathbf{M}}$ *be a uniformly random* $n \times n$ *matrix over* $\mathbb{F}_2$. *For every* $k \in \{0, \ldots, n\}$,

$$\Pr[\mathrm{rank}(\hat{\mathbf{M}}) = n-k] = \frac{1}{2^{k^2}} \frac{\prod_{i=1}^{n}(1-1/2^i)\prod_{i=k+1}^{n}(1-1/2^i)}{\prod_{i=1}^{n-k}(1-1/2^i)\prod_{i=1}^{k}(1-1/2^i)}.$$

Formally, the following lemma shows that the input matrix is good with high probability.

▶ **Lemma 33.** *With probability* $\geq 1 - o(1)$, *a random input matrix from* $\mathcal{M}$ *is good.*

Proof. For every $d \in [10\log n, n/3]$, there are $\binom{n/3}{d}^2 \leq (\frac{en/3}{d})^{2d}$ submatrices of dimension $d \times d$. For each submatrix $\hat{\mathbf{M}}[V_1', V_2']$, and every $k \in [d/2+1, \ldots, d]$, by Proposition 32 we have

$$\Pr[\mathrm{rank}(\hat{\mathbf{M}}[V_1', V_2']) = d-k] = \frac{1}{2^{k^2}} \frac{\prod_{i=1}^{d}(1-1/2^i)\prod_{i=k+1}^{d}(1-1/2^i)}{\prod_{i=1}^{d-k}(1-1/2^i)\prod_{i=1}^{k}(1-1/2^i)} \leq 2^{d-d^2/4}.$$

Therefore,

$$\Pr[\mathrm{rank}(\hat{\mathbf{M}}[V_1', V_2']) \leq d/2] \leq (d/2)2^{d-d^2/4}.$$

By union bound, the probability that at least one of the submatrices have a low rank is

$$\Pr[\mathbf{M} \text{ is bad}] \leq \sum_{d=10\log n}^{n/3} (\frac{en/3}{d})^{2d}(d/2)2^{d-d^2/4} \leq \sum_{d=10\log n}^{n/3} 2^{-5\log^2 n} = o(1). \qquad \blacktriangleleft$$

Equipped with the above tools, we are ready to prove the lower bound to AMP with high redundancy.

▶ **Theorem 34** (Lower bound to AMP with superlinear-in-n redundancy)**.** *Fix* $2 \leq k \leq n$ *and a constant* $\alpha > 6$. *For every (possibly randomized) succinct and systematic data structure with* $r \in [10n\log n, n^2/1000]$ *bits of redundancy that answers the whole approximated matrix power* $[\mathbf{M}^k]_\alpha$ *correctly with* $\geq 99/100$ *probability, its total number of probes is* $\Omega(n^4/r)$ *in the worst case.*

Proof. For each $(u,v) \in V_1 \times V_3$, to approximate $[\mathbf{M}^k]_\alpha$ for $\alpha > 6$, by Observation 29, the algorithm should also be able to answer exactly the number of length-2 paths from u to v.

By Yao's minimax principle, there is a randomized data structure that answer the fixed sequence of queries Q with high probability efficiently for every possible input only if there exists a deterministic data structure that with the same success probability and the same complexity given a random input drawn from $\mathcal{M}$. In below, we focus on deterministic data structures and query-with-sketch algorithms.

By Lemma 3 and Lemma 8, the lower bound reduces to showing that for every block of $m = 20r$ consecutive queries from Q, and for every partial assignment σ of length at most $q := 0.001n^2$ to $\mathbf{M}$, the joint min-entropy of the output conditioned on σ is $\geq 2r$. We will focus on the min-entropy of the unrounded elements of $\mathbf{M}^k$ in below, as it is equivalent to the min-entropy of the rounded elements of $[\mathbf{M}^k]_\alpha$, given our input distribution $\mathcal{M}$. Besides, in applying Lemma 8, we only pick a good input set M constructed above, and set $\mathsf{M}_\sigma = \mathbb{R}^{n\times n}$ for every σ.

By our construction, the input matrix $\mathbf{M}$ is fixed other than two submatrices, $\mathbf{M}[V_1 \times V_2]$ and $\mathbf{M}[V_2 \times V_3]$. Given a partial assignment of length at most q, there are at most $0.1n$ vertices from $V_1 \cup V_3$, where each of their corresponding rows (if in V_1) or columns (if in V_3) has $\geq 0.01n$ elements revealed from σ. We call these vertices "bad vertices", and other vertices "good vertices". Let $Q' = ((u_1, v_1), \ldots, (u_{m'}, v_{m'}))$ be the subsequence of the m consecutive queries containing only pairs of good vertices. By Observation 30, we have

$$m' \geq m - 0.1n \cdot (\frac{m}{n/3} + 1) \geq 0.69m$$

There exists a sequence $(w_1, \ldots, w_{m'})$ such that the sequence of triples $(u_1, v_1, w_1), \ldots, (u_{m'}, v_{m'}, w_{m'})$ are *edge-disjoint* triples where none of the edges $(u_i, w_i), (w_i, v_i)$ are revealed by σ, which means that none of pairs of indices $(u, v), (v, w)$, or (u, w) belong to multiple triples. Its existence directly follows from a construction: we decide each w_i one by one. For each (u_i, v_i), given $w_{<i}$ fixed, among $n/3$ indices from V_2, at least $n/3 - 2\lceil \frac{m}{n/3} \rceil - 0.02n \geq 0.19n$ elements of $\mathbf{M}[u_i, \cdot]$ and $\mathbf{M}[\cdot, v_i]$ are unfixed by σ and $w_{<i}$, which means there always exists an w_i that satisfies our condition.

Observe that for every partial assignment σ of length $\leq q$,

$$\begin{aligned} & H_{\min}(\mathbf{M}^k[u_1, v_1], \ldots, \mathbf{M}^k[u_{m'}, v_{m'}], \mathbf{M} \in \mathsf{M} | \sigma) \\ \geq & \min_{\sigma'} H_{\min}(\mathbf{M}^k[u_1, v_1], \ldots, \mathbf{M}^k[u_{m'}, v_{m'}], \mathbf{M} \in \mathsf{M} | \sigma') \end{aligned}$$

where σ' is selected from all possible partial assignments to $\mathbf{M}$ that is consistent to σ and reveals all the elements other than the m' elements $\mathbf{M}[w_1, v_1], \ldots, \mathbf{M}[w_{m'}, v_{m'}]$.

Fix a vertex $v \in V_3$ such that there are $\geq 10 \log n$ queries on v in Q'. Since $m' \geq 0.69m \geq 13.8r \geq 138n \log n$, at least $\geq 202m'/207$ queries from Q' are made on such vertices $v \in V_3$. Or otherwise there will be less than

$$\frac{n}{3} \cdot 10 \log n + \frac{202m'}{207} \leq m'$$

queries in Q'.

Let $d \geq 10 \log n$ be the number of queries on v in Q', and $(u'_1, v, w'_1), \ldots, (u'_d, v, w'_d)$ the triples. By our construction of w and Q, none of the vertices $u'_i \in V_1$ and $w'_i \in V_2$ repeat in the d triples. In addition, conditioned on σ', different columns of the matrix power $\mathbf{M}^k[\cdot, v]$ are independent. The outputs $\mathbf{M}^k[u'_1, v], \ldots, \mathbf{M}^k[u'_d, v]$ only depend on variables $\mathbf{M}[w'_1, v], \ldots, \mathbf{M}[w'_d, v]$. More precisely, each $\mathbf{M}^k[u'_i, v]$ is a linear combination of the variables: for every $i \in [d]$,

$$\mathbf{M}^k[u'_i, v] = \underbrace{\sum_{w: \forall j, w \neq w'_j} \mathbf{M}[u'_i, w]\mathbf{M}[w, v]}_{\text{fixed given } \sigma'} + \sum_{j=1}^{d} \mathbf{M}[u'_i, w'_j]\mathbf{M}[w'_j, v]$$

We only need to show that the set of vectors $\{(\hat{\mathbf{M}}[u'_i, w'_j])_{j \in [d]}\}_{i \in [d]}$ form a basis of rank at least $d/2$. If it is true, the parity of the output elements, which can always be presented as a linear combination of these vectors, will have $\geq d/2$ min-entropy.

Formally, let $\mathbf{y} \in \mathbb{F}_2^d$ be the vector where $\mathbf{y}(i) = \hat{\mathbf{M}}^k[u_i', v]$. Let $\mathbf{x} \in \mathbb{F}_2^d$ be the vector where $\mathbf{x}(i) = \hat{\mathbf{M}}[w_i, v]$. Let $V_1' = \{u_1', \dots, u_d'\}$ and $V_2' = \{w_1', \dots, w_d'\}$ Then for every σ', there exists a vector $\mathbf{w} \in \mathbb{F}_2^d$ such that

$$\mathbf{y} = \mathbf{w} + \hat{\mathbf{M}}[V_1', V_2']\mathbf{x}$$

where $\mathbf{x}$ is a uniformly random vector over $\mathbb{F}_2^d$, given σ'. Therefore,

$$\begin{aligned}
&H_{\min}(\mathbf{M}^k[u_1', v], \dots, \mathbf{M}^k[u_d', v], \mathbf{M} \in \mathsf{M}|\sigma') \\
\geq& \min_{\sigma'} H_{\min}(\mathbf{y}|\sigma') \\
=& \min_{\sigma'} H_{\min}(\hat{\mathbf{M}}[V_1', V_2']\mathbf{x}|\sigma') \\
=& \min_{\sigma'} \operatorname{rank}(\hat{\mathbf{M}}[V_1', V_2']) \\
\geq& d/2.
\end{aligned}$$

Here, the min-entropy of the output is greater than or equal to the min-entropy of $\mathbf{y}$ because by Observation 29, $\mathbf{M}^k[u, v]$ determines the number of length-2 walks from u to v, and hence its parity.

Therefore, each vertex v of d queries contribute $\geq d/2$ min-entropy. In addition, different columns $\mathbf{M}^k[\cdot, v]$ of the output are independent given σ'. The total min-entropy of the output is at least $\frac{1}{2} \cdot \frac{202m'}{207}$, where $\frac{202m'}{207}$ is the least number of queries from Q' that are made on these vertices v. We conclude that

$$H_{\min}(\mathbf{M}^k[u_1, v_1], \dots, \mathbf{M}^k[u_{m'}, v_{m'}], \mathbf{M} \in \mathsf{M}|\sigma) \geq \frac{101m'}{207} \geq 6.7r > 2r$$

By Lemma 3 and Lemma 8, the lower bound follows. ◀

The lower bound to the counting set intersection follows through a reduction from AMP.

▶ **Corollary 35** (Lower bound to counting set intersection). *Given a (possibly randomized) succinct and systematic data structure with $r \in [10n \log n, n^2/1000]$ bits of redundancy that given as input $2n/3$ subsets of $[n/3]$ answers their pairwise set intersection size correctly with $\geq 99/100$ probability, its total number of probes is $\Omega(n^4/r)$ at the worst case.*

We note that no super-linear-in-n amortized probe lower bound for counting set intersection can be obtained. Because the intersection size of every pair of sets can be answered in $O(n)$ probes.

Proof. Let the input sets follow i.i.d. uniform distribution from $\{0, 1\}^{n/3}$. Counting set intersection is exactly counting the number of length-2 paths given an input from $\mathcal{M}$, which we have established a lower bound above. ◀

5 Lower bounds to set disjointness and $(2 - \varepsilon)$-approximate APSP

In this section, we study the lower bound to the set disjointness, which is the decision version of counting set intersection. The same nemesis distribution and analysis also implies the same probe-redundancy trade-off for the $(2 - \varepsilon)$-approximation of APSP. We state our proof in the context of set disjointness.

Set disjointness problem

Given as input n sets $A_1, \dots, A_n \subseteq [n]$. The succinct and systematic data structure, given a pair of indices $(u, v) \in [n] \times [n]$, must answer whether $|A_u \cap A_v| = 0$. For simplicity, we denote the answer as $Ans_{u,v}$, which is a binary bit that is 1 iff the answer is NO.

For the convenience of our analysis, each A_u also denotes a binary vector of length n, where the w-th bit denotes whether $w \in A_u$.

The uniform input distribution is no longer hard for this problem: when each vector is i.i.d. uniformly from $\{0,1\}^n$, with high probability for every u, v, $|A_u \cap A_v| \neq 0$. To that end, we construct the following nemesis distribution where 1s are sparse in the input vectors.

Nemesis input distribution

Let $\mathcal{A}$ denote the input distribution that, for every $u, i \in [n]$, $A_u(i) = 1$ with independent probability $1/\sqrt{n}\log n$.

One can see that $\Pr[Ans_{u,v} = 1] = \Theta(1/\log^2 n)$ for every $u \neq v$.

We use a similar sequence Q of queries as above. Let $V_1 := \{1, \dots, n/2\}$ and $V_2 := \{n/2+1, \dots, n\}$. We let Q to be a concatenation of $n/2$ disjoint bijections between V_1 and V_2:

$$Q = (\underbrace{(1, n/2+1), (2, n/2+2), \dots, (n/2, n)}_{\text{identity bijection}}, \dots, \underbrace{(1, n), (2, n/2+1), \dots, (n/2, n-1)}_{\text{shifted bijection}})$$

Formally, for every $i \in [n/2]$, the i-th bijection is exactly a sequence of pairs (u, v) where

- $u \in V_1$;
- $v = (u + i - 2) \bmod (n/2) + n/2 + 1$.

The main result in this section is the following lower bound.

▶ **Theorem 36** (Lower bound to set disjointness). *Given a (possibly randomized) succinct and systematic data structure with $r \in [10n\log^2 n, n\sqrt{n}/1000\log^3 n]$ bits of redundancy that given as input n subsets of $[n]$ answers the sequence of queries Q correctly with $\geq 99/100$ probability, its total number of probes is $\Omega(n^4/(r\log^3 n))$ in the worst case.*

Different from the previous lower bounds to AMP, the answers cannot be formulated as (nearly-)linear combinations of selected input elements, where analytic or linear algebraic tools may apply. However, the sparsity of the sets still enables us to lower bounding the min-entropy. Formally, we define the following set of "good inputs".

▶ **Definition 37.** *Let A be the set of good inputs $\{A = (A_1, \dots, A_n)\}$ satisfying the following conditions*

(1) *For every $u \in [n]$,*

$$0.99\frac{\sqrt{n}}{\log n} \leq |A_u| \leq 1.01\frac{\sqrt{n}}{\log n}.$$

(2) *For every $k \in [\log^2 n, \sqrt{n}]$, $u \in V_1$ and $v \in V_2$,*

$$\sum_{i=0}^{k-1} Ans_{u+i,v} \leq \frac{2k}{\log n}$$

where $Ans_{u+i,v}$ denotes $Ans_{u+i-n/2,v}$ if $u+i > n/2$.

(3) *For every* $k \in [\log^2 n, \sqrt{n}]$*, every* $u \in V_1$ *and every* $w \in [n]$*,*

$$\sum_{i=0}^{k-1} A_{u+i}(w) \leq 2 \log n$$

where $A_{u+i}(w)$ *denotes* $A_{u+i-n/2}(w)$ *if* $u + i > n/2$*.*

We will show that, for most of indices $v \in V_2$, for a fixed assignment to all the $A_1, \ldots, A_{n/2}$, and for a fixed sequence of answers $Ans_{\cdot,v}$, a large proportion of bits in A_v are forced to be 0 if the input is good, which results in a high conditional min-entropy.

We use the following form of Chernoff bound to show that a random input is good with high probability.

▶ **Proposition 38** (simplified Chernoff bound)**.** *Let* $X_1, X_2, \ldots, X_n$ *be independent random variables such that* $X_i \in [0, 1]$*. Let* $X = \sum_{i=1}^{n} X_i$*. Then, for every* $\delta > 0$*,*

$$\Pr[|X - \mathbb{E}[X]| \geq \delta \cdot \mathbb{E}[X]] \leq 2 \cdot \exp\left(-\frac{\delta^2}{2+\delta} \cdot \mathbb{E}[X]\right).$$

▶ **Lemma 39.** *For every large enough* n*, a random input from* $\mathcal{A}$ *is good with probability at least* $99/100$*.*

Proof. We will show that each condition holds with low probability. The desired probability bound follows from the union bound.

For condition (1), we know from our input construction that $\mathbb{E}[|A_u|] = \frac{\sqrt{n}}{\log n}$. By Chernoff bound,

$$\Pr\left[\left|A_u - \frac{\sqrt{n}}{\log n}\right| \geq 0.01 \frac{\sqrt{n}}{\log n}\right] \leq \exp(-\Omega(\sqrt{n}/\log n)).$$

By union bound over all u, with $\geq 1 - o(1)$ probability condition (1) holds.

Then, we show that condition (2) holds with high probability conditioned on (1) is true. Fix an index $v \in V_2$ and an assignment α to A_v where $|\alpha| \in [0.99\frac{\sqrt{n}}{\log n}, 1.01\frac{\sqrt{n}}{\log n}]$. Conditioned on a fixed $A_v = \alpha$, for every u and i, $Ans_{u+i,v}$ are independent random variables where

$$\Pr[Ans_{u+i,v} = 0 | A_v = \alpha] = \left(1 - \frac{1}{\sqrt{n}\log n}\right)^{|\alpha|}.$$

By Observation 14 and the Taylor expansion $\exp(-x) = 1 - x + \frac{x^2}{2} - \ldots$, we have

$$\Pr[Ans_{u,v} = 1 | A_v = \alpha] \geq 1 - \exp\left(-\frac{|\alpha|}{\sqrt{n}\log n}\right) \geq \frac{|\alpha|}{\sqrt{n}\log n} - \frac{|\alpha|^2}{2n\log^2 n} \geq \frac{0.98}{\log^2 n}$$

and

$$\Pr[Ans_{u,v} = 1 | A_v = \alpha] \leq 1 - \exp\left(-\frac{|\alpha|}{\sqrt{n}\log n - 1}\right) \leq \frac{|\alpha|}{\sqrt{n}\log n - 1} \leq \frac{1.02}{\log^2 n}.$$

By Chernoff bound,

$$\Pr\left[\sum_{i=0}^{k-1} Ans_{u+i,v} > \frac{2k}{\log n} \middle| A_v = \alpha\right] \leq \exp(-1.9k/\log n) \leq O(n^{-2.7})$$

By union bound over u, v, and k, with $\geq 1 - o(1)$ probability condition (2) holds.

For condition (3), notice that $\mathbb{E}[\sum_{i=0}^{k-1} A_{u+i}(w)] = \frac{k}{\sqrt{n}\log n}$. By Chernoff bound, for every k, u, w,

$$\Pr\left[\sum_{i=0}^{k-1} A_{u+i}(w) > 5\log n\right] \le O(n^{-2.7}).$$

Again, by union bound, condition (3) holds with $\ge 1 - o(1)$ probability. ◀

However, the above set of good inputs still does not guarantee a high conditional min-entropy. Imagine a magical algorithm that finds all the 1s in all the strings in $O(n\sqrt{n}/\log n)$ probes. No entropy would remain in the input. To that end, we construct good input sets A_σ for every partial assignment σ of length $\le q$. Intuitively, the good input sets A_σ ensures that high uncertainty remains in the unprobed part of the input.

▶ **Definition 40.** *For every partial assignment σ of length $\le 10^{-4}n^2$, let A_σ be the set of good inputs $\{A = (A_1, \ldots, A_n)\}$ such that for every shifted bijection $(1, v_1), \ldots, (n/2, v_{n/2})$ in the sequence of queries Q, no more than $0.15n$ indices v_i in V_2 satisfy the following condition: let $W_{i,v_i} \subseteq [n]$ denote the set of indices w such that $A_i(w) = 1$ and $A_{v_i}(w)$ is revealed by σ; then $|W_{i,v_i}| > 0.5\sqrt{n}/\log n$.*

To show that a random input falls in A_σ with high probability conditioned on σ, we will use the following standard Chernoff bound.

▶ **Proposition 41** (standard Chernoff bound [28]). *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0,1\}$ and $\Pr[X_i = 1] \in (0,1)$ for $i \in [n]$. Then, for $X = \sum_{i=1}^n X_i, \mu = \mathbb{E}[X]$, and any $\delta > 0$,*

$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu.$$

▶ **Lemma 42.** *For every partial assignment σ of length $\le 10^{-4}n^2$, we have*

$$\Pr_{A \leftarrow \mathcal{A}}[A \in \mathsf{A} - \mathsf{A}_\sigma | \sigma] \le \exp(-0.0048n\sqrt{n}/\log n).$$

Proof. We show that for every fixed shifted bijection $(1, v_1), \ldots, (n/2, v_{n/2})$, the condition is satisfied with high probability. And we conclude with a union bound over all the n shifted bijections.

Given σ of length at most q, there are at most $0.01n$ vectors in which $\ge 0.01n$ bits are revealed from σ. Without loss of generality, we assume that the first $0.49n$ pairs $(1, v_1), \ldots, (0.49n, v_{0.49n})$ do not contain such vectors.

For each of these pairs (i, v_i) for $i \in [0.49n]$, there are $\ge 0.98n$ indices $w \in [n]$ such that neither of $A_i(w)$ nor $A_{v_i}(w)$ are revealed by σ. Without loss of generality, we assume these indices are $w = 1, \ldots, 0.98n$. When $A \in \mathsf{A}$ and $|W_{i,v_i}| > 0.5\sqrt{n}/\log n$, we should have $\sum_{w=1}^{0.98n} A_i(w) < 0.51\sqrt{n}/\log n$, by $|A_i| \le 1.01\sqrt{n}/\log n$. We apply the simplified Chernoff bound (Proposition 38) with $\mathbb{E}[\sum_{w=1}^{0.98n} A_i(w)|\sigma] = 0.98\sqrt{n}/\log n$, and obtain

$$\Pr[\sum_{w=1}^{0.98n} A_i(w) < 0.51\sqrt{n}/\log n|\sigma] \le 2\cdot\exp(-\frac{47}{243}\cdot 0.51\sqrt{n}/\log n) \le 2\cdot\exp(-0.098\sqrt{n}/\log n).$$

Let $X_1, \ldots, X_{0.49n} \in \{0,1\}$ be the indicators of whether $|W_{i,v_i}| > 0.5\sqrt{n}/\log n$. Then, conditioned on σ, these indicators are independent random bits, each evaluating to 1 with probability $\le 2 \cdot \exp(-0.098\sqrt{n}/\log n)$. $A \in \mathsf{A} - \mathsf{A}_\sigma$ only if $X := \sum_{i=1}^{0.49n} X_i > 0.05n$.

By applying the standard Chernoff bound (Proposition 41) with $(1+\delta)\mu = 0.05n$ and $\mu \leq 0.98n\exp(-0.098\sqrt{n}/\log n)$, we get

$$\Pr[X > 0.05n] < \exp(-0.0048n\sqrt{n}/\log n).$$

The desired inequality follows by a union bound over all $n/2$ shifted bijections. ◀

Given an input with the above conditions, we are able to prove the lower bound.

Proof to Theorem 36. Again, by Yao's minimax principle, we may focus on proving lower bounds for data structures given a random input from $\mathcal{A}$. By Lemma 3 and Lemma 8, the lower bound reduces to showing that for every $m := 120r\log^3 n$ consecutive queries $(u_1, v_1), \ldots, (u_m, v_m)$ from Q, and for every partial assignment σ of length at most $q := 10^{-4}n^2$ to $\mathbf{M}$, the joint min-entropy of the output conditioned on σ is $\geq 2r$. We may always assume that m is a multiple of $n/2$, as this will only increase m by at most $n/2 = o(r\log^3 n)$. In addition, this ensures that the consecutive m queries is a concatenation of shifted bijections. As shown in Lemma 39 and Lemma 42, the input falls in A with high probability, and falls in A_σ with high probability for every σ.

Observe that for every partial assignment σ of length $\leq q$,

$$\begin{aligned} & H_{\min}(Ans_{u_1,v_1}, \ldots, Ans_{u_m,v_m}, A \in \mathsf{A} \cap \mathsf{A}_\sigma | \sigma) \\ \geq & \min_{\sigma'} H_{\min}(Ans_{u_1,v_1}, \ldots, Ans_{u_m,v_m}, A \in \mathsf{A} \cap \mathsf{A}_\sigma | \sigma') \end{aligned}$$

where σ' is selected from all possible partial assignments to $A = (A_1, \ldots, A_n)$ that is consistent to σ and reveals all the $A_1, \ldots, A_{n/2}$. We will show that the min-entropy remains high even conditioned on σ'.

Before bounding the min-entropy, let us further pick an analyzable part from the remaining input. Given σ of length at most q, there are at most $0.01n$ vectors in which $\geq 0.01n$ bits are revealed from σ. We call these vectors "bad vectors", and other vectors "good vectors".

By our construction to Q, for every m consecutive queries, each A_u for $u \in [n]$ is involved in exactly $m/(n/2)$ queries. Therefore, at least $0.98m$ queries are among good vectors. In addition, at most $0.3m$ queries (u,v) has $|W_{u,v}| > 0.5\sqrt{n}\log n$. Denote by $m' \geq 0.68m$ the number of remaining queries, and the sequence of remaining queries Q'.

Fix an index $v \in V_2$ such that it is contained in at least $m'/3n$ queries in Q'. There are at least $\geq 0.31n$ many such vertices, otherwise the total number of queries in Q' is

$$< 0.31n \cdot \frac{2m}{n} + \left(\frac{n}{2} - 0.31n\right) \cdot \frac{m'}{3n} \leq m'$$

where $2m/n$ is the maximum number of queries that involve a vertex $v \in V_2$ in m consecutive queries.

We denote by $d \geq m'/3n > \log^2 n$ the number of queries that contain v, and $(u'_1, v), \ldots, (u'_d, v) \in Q'$ these queries. Given all the vectors in V_1 fixed by σ', these queries are independent of other queries in Q'.

Assuming a good input, by the condition (2) of Definition 37, at most $\frac{2 \cdot 2m/n}{\log n} = \frac{4m}{n\log n}$ out of d queries have answers of 1. Recall that condition (2) gives an upper bound to the sum of answers of a consecutive sequence of queries to every fixed v. Note from the construction to Q that the m consecutive queries contain exactly $2m/n \leq \sqrt{n}$ consecutive queries on v. While Q' is obtained by removing elements from the m queries, the sum of answers of queries in Q' is smaller than or equal to the consecutive sum in Q, and is also bounded by condition (2).

Without loss of generality, we assume that they are the first $d' \leq \frac{4m}{n \log n}$ queries:

$Ans_{u'_1,v} = \cdots = Ans_{u'_{d'},v} = 1.$

For each $u \in [n]$, we use A'_u to denote the vector A_u removed all the indices w where $A_v(w)$ is revealed by σ. We will show that the Hamming weight of $A'_{u'_{d'+1}} \vee \cdots \vee A'_{u'_d}$ is high. That is, given σ', for every fixed and possible answer to good inputs, a large enough proportion of bits of A'_v are forced to be 0, to avoid intersecting with any $A'_{u'_{d'+1}}, \ldots, A'_{u'_d}$. Therefore, the conditional probability this happens is always small.

Notice that by condition (1) of Definition 37 and Definition 40,

$$\sum_{i=d'+1}^{d} |A'_{u'_i}| \geq (d-d') \cdot 0.49 \frac{\sqrt{n}}{\log n} > \frac{9.7r \log^2 n}{\sqrt{n}}$$

where for each index $w \in [n]$, by condition (3) of Definition 37,

$$\sum_{i=d'+1}^{d} |A'_{u'_i}(w)| \leq 2 \log n.$$

Therefore, for the Hamming weight of the OR of these binary strings,

$$\left| A'_{u_{d'+1}} \vee \cdots \vee A'_{u_d} \right| \geq \frac{9.7r \log^2 n}{\sqrt{n}} \cdot \frac{1}{2 \log n} = \frac{4.8r \log n}{\sqrt{n}}.$$

Since given σ, A'_v is a random vector where each bit is 1 with independent probability $1/\sqrt{n} \log n$, we have

$$\Pr[A'_v \wedge (A'_{u_{d'+1}} \vee \cdots \vee A'_{u_d}) = 0 \text{ and } A \in \mathsf{A} \cap \mathsf{A}_\sigma | \sigma'] \leq (1 - 1/\sqrt{n} \log n)^{\frac{4.8r \log n}{\sqrt{n}}} \leq \exp(-4.8r/n).$$

That means that the min-entropy of queries on v is at least

$4.8 \log(e) r/n \geq 6.92r/n.$

Since the min-entropy on different $v \in V_2$ conditioned on σ' are independent, the total min-entropy is lower bounded by their sum:

$$H_{\min}(Ans_{u_1,v_1}, \ldots, Ans_{u_l,v_l}, A \in \mathsf{A} \cap \mathsf{A}_\sigma | \sigma) \geq 0.31n \cdot \frac{6.92r}{n} > 2r$$

Therefore, the lower bound holds for set disjointness. ◀

The above lower bound also applies for the $(2-\varepsilon)$-approximate APSP, for every $\varepsilon > 0$. We obtain it via a reduction to set disjointness under the nemesis distribution $\mathcal{A}$.

$(2-\varepsilon)$-approximate all-pairs shortest paths (APSP)

Fix a parameter $\varepsilon > 0$, Given as input an undirected graph $G = (V, E)$. The succinct and systematic data structure, given a sequence of all pairs of vertices $(u, v) \in \binom{V}{2}$, must answer a value $d \in [d_G(u,v), (2-\varepsilon) d_G(u,v)]$ for each (u, v), where $d_G(u,v)$ denotes the shortest distance between u and v in G.

▶ **Theorem 43** (Lower bound to $(2-\varepsilon)$-approximate APSP). *Fix $\varepsilon > 0$. Given a (possibly randomized) succinct and systematic data structure with $r \in [5n \log^2 n, n\sqrt{n}/4000 \log^3 n]$ bits of redundancy that given as input a graph G on n vertices $(2-\varepsilon)$-approximates all pairs shortest paths distances correctly with $\geq 99/100$ probability, its total number of probes is $\Omega(n^4/(r \log^3 n))$ at the worst case.*

Proof. We prove the lower bound through a reduction from set disjointness. For every instance of set disjointness, we construct an input to APSP, such that data structures to APSP, through the reduction, can be revised into data structures for set disjointness with the same success probability and complexity.

Suppose for the sake of contradiction that there exists a data structure D of bounded complexity for $(2-\varepsilon)$-APSP. Given an input from $\mathcal{A}$. We construct a graph $G = (V, E)$ of $2n$ vertices, where $V = V_1 \cup W \cup V_2$ and $2|V_1| = |W| = 2|V_2| = n$. For every $u \in V_1$ and $w \in [n]$, u connects to the w-th vertex in W if and only if $A_u(w) = 1$. Analogously, for every $v \in V_2$ and $w \in [n]$, v connects to the w-th vertex in W if and only if $A_v(w) = 1$.

For each query (u, v) from Q, we ask D about the shortest path length between $u \in V_1$ and $v \in V_2$ in G. If the answer $d \in [2, 4)$, we answer that $|A_u \cap A_v| \geq 1$; otherwise, we answer $|A_u \cap A_v| = 0$.

To see the correctness, by our construction, $d_G(u, v) = 2$ if and only if $|A_u \cap A_v| \geq 1$; otherwise $d_G(u, v) \geq 4$. The constructed graph is a bipartite graph on two parts $V_1 \cup V_2$ and W, and there cannot be paths of length 3 between u and v. We get a data structure for set disjointness with the same redundancy and the same probe complexity. Therefore, the same lower bound holds for $(2-\varepsilon)$-approximate APSP. ◀

References

1 Peyman Afshani and Jesper Sindahl Nielsen. Data structure lower bounds for document indexing problems. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, Rome, Italy, July 11-15, 2016*, volume 55 of *LIPIcs*, pages 93:1–93:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. `doi:10.4230/LIPIcs.ICALP.2016.93`.

2 Richard Beigel and Tirza Hirst. One help bit doesn't help. In Jeffrey Scott Vitter, editor, *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 124–130. ACM, 1998. `doi:10.1145/276698.276720`.

3 Eric David Belsley. *Rates of convergence of Markov chains related to association schemes*. Harvard University, 1993.

4 Vladimir I Bogachev. Measures on topological spaces. In *Measure theory*, pages 476–583. Springer, 2007.

5 Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998. `doi:10.1016/S0169-7552(98)00110-X`.

6 Karl Bringmann and Kasper Green Larsen. Succinct sampling from discrete distributions. In *Proceedings of the forty-fifth annual ACM symposium on Theory of Computing*, pages 775–782, 2013. `doi:10.1145/2488608.2488707`.

7 Diptarka Chakraborty, Lior Kamma, and Kasper Green Larsen. Tight cell probe bounds for succinct boolean matrix-vector multiplication. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1297–1306, 2018. `doi:10.1145/3188745.3188830`.

8 Hagai Cohen and Ely Porat. Fast set intersection and two-patterns matching. *Theoretical Computer Science*, 411(40-42):3795–3800, 2010. `doi:10.1016/j.tcs.2010.06.002`.

9 Henry Corrigan-Gibbs and Dmitry Kogan. The function-inversion problem: Barriers and opportunities. In *Theory of Cryptography Conference*, pages 393–421, 2019. `doi:10.1007/978-3-030-36030-6_16`.

10 Zeev Dvir, Alexander Golovnev, and Omri Weinstein. Static data structure lower bounds imply rigidity. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 967–978, 2019. `doi:10.1145/3313276.3316348`.

11 Pavel Dvorák, Michal Koucký, Karel Král, and Veronika Slívová. Data structures lower bounds and popular conjectures. In Petra Mutzel, Rasmus Pagh, and Grzegorz Herman, editors, *29th Annual European Symposium on Algorithms, ESA 2021, Lisbon, Portugal (Virtual Conference), September 6-8, 2021*, volume 204 of *LIPIcs*, pages 39:1–39:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ESA.2021.39.

12 Jason Fulman and Larry Goldstein. Stein's method and the rank distribution of random matrices over finite fields. *The Annals of Probability*, 43(3):1274–1314, 2015.

13 Anna Gál and Peter Bro Miltersen. The cell probe complexity of succinct data structures. *Theoretical computer science*, 379(3):405–417, 2007. doi:10.1016/j.tcs.2007.02.047.

14 Isaac Goldstein, Tsvi Kopelowitz, Moshe Lewenstein, and Ely Porat. Conditional lower bounds for space/time tradeoffs. In *Workshop on Algorithms and Data Structures*, pages 421–436, 2017. doi:10.1007/978-3-319-62127-2_36.

15 Isaac Goldstein, Moshe Lewenstein, and Ely Porat. On the hardness of set disjointness and set intersection with bounded universe. In Pinyan Lu and Guochuan Zhang, editors, *30th International Symposium on Algorithms and Computation, ISAAC 2019, Shanghai University of Finance and Economics, Shanghai, China, December 8-11, 2019*, volume 149 of *LIPIcs*, pages 7:1–7:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ISAAC.2019.7.

16 Alexander Golovnev, Siyao Guo, Thibaut Horel, Sunoo Park, and Vinod Vaikuntanathan. Data structures meet cryptography: 3sum with preprocessing. In *Proceedings of the 52nd annual ACM SIGACT symposium on theory of computing*, pages 294–307, 2020. doi:10.1145/3357713.3384342.

17 Gene H Golub and Charles F Van Loan. *Matrix computations*. JHU press, 2013.

18 Alexander Golynski. Optimal lower bounds for rank and select indexes. *Theoretical Computer Science*, 387(3):348–359, 2007. doi:10.1016/j.tcs.2007.07.041.

19 Alexander Golynski, Roberto Grossi, Ankur Gupta, Rajeev Raman, and S. Srinivasa Rao. On the size of succinct indices. In Lars Arge, Michael Hoffmann, and Emo Welzl, editors, *Algorithms - ESA 2007, 15th Annual European Symposium, Eilat, Israel, October 8-10, 2007, Proceedings*, volume 4698 of *Lecture Notes in Computer Science*, pages 371–382. Springer, 2007. doi:10.1007/978-3-540-75520-3_34.

20 Alexander Golynski, Rajeev Raman, and S Srinivasa Rao. On the redundancy of succinct data structures. In *Scandinavian Workshop on Algorithm Theory*, pages 148–159, 2008. doi:10.1007/978-3-540-69903-3_15.

21 Monika Henzinger, Sebastian Krinninger, Danupon Nanongkai, and Thatchaphol Saranurak. Unifying and strengthening hardness for dynamic problems via the online matrix-vector multiplication conjecture. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 21–30, 2015. doi:10.1145/2746539.2746609.

22 Roger A. Horn and Charles R. Johnson. *Matrix Analysis, 2nd Ed.* Cambridge University Press, 2012. doi:10.1017/CBO9781139020411.

23 Guy Joseph Jacobson. *Succinct static data structures*. Carnegie Mellon University, 1988.

24 Tsvi Kopelowitz and Virginia Vassilevska Williams. Towards optimal set-disjointness and set-intersection data structures. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, Saarbrücken, Germany (Virtual Conference), July 8-11, 2020*, volume 168 of *LIPIcs*, pages 74:1–74:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ICALP.2020.74.

25 Kasper Green Larsen. Higher cell probe lower bounds for evaluating polynomials. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 293–301. IEEE, 2012. doi:10.1109/FOCS.2012.21.

26 Peter Bro Miltersen. Lower bounds on the size of selection and rank indexes. In *SODA*, volume 5, pages 11–12, 2005. URL: http://dl.acm.org/citation.cfm?id=1070432.1070435.

27 Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. In *Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 103–111, 1995. doi:10.1145/225058.225093.

28 Rajeev Motwani and Prabhakar Raghavan. Randomized algorithms. *ACM Computing Surveys (CSUR)*, 28(1):33–37, 1996. doi:10.1145/234313.234327.

29 Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM Journal on Computing*, 28(3):1035–1050, 1998. doi:10.1137/S0097539795282444.

30 James R. Norris. *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1998.

31 Mihai Patrascu and Liam Roditty. Distance oracles beyond the thorup-zwick bound. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 815–823. IEEE, 2010. doi:10.1109/FOCS.2010.83.

32 Mihai Patrascu and Emanuele Viola. Cell-probe lower bounds for succinct partial sums. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 117–122. SIAM, SIAM, 2010. doi:10.1137/1.9781611973075.11.

33 Sivaramakrishnan Natarajan Ramamoorthy and Cyrus Rashtchian. Equivalence of systematic linear data structures and matrix rigidity. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference, ITCS 2020, Seattle, Washington, USA, January 12-14, 2020*, volume 151 of *LIPIcs*, pages 35:1–35:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.ITCS.2020.35.

34 Stanislaw J Szarek. Spaces with large distance to ℓ_∞^n and random matrices. *American Journal of Mathematics*, 112(6):899–942, 1990.

35 Robert Endre Tarjan. A class of algorithms which require nonlinear time to maintain disjoint sets. *Journal of computer and system sciences*, 18(2):110–127, 1979. doi:10.1016/0022-0000(79)90042-4.

36 Robert C Thompson. The behavior of eigenvalues and singular values under perturbations of restricted rank. *Linear Algebra and its Applications*, 13(1-2):69–78, 1976.

37 Mikkel Thorup and Uri Zwick. Approximate distance oracles. *Journal of the ACM (JACM)*, 52(1):1–24, 2005. doi:10.1145/1044731.1044732.

38 Leslie G Valiant. Graph-theoretic arguments in low-level complexity. In *International Symposium on Mathematical Foundations of Computer Science*, pages 162–176, 1977. doi:10.1007/3-540-08353-7_135.

39 Leslie G Valiant. Why is boolean complexity theory difficult. *Boolean Function Complexity*, 169(84-94):4, 1992.

40 Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

41 Emanuele Viola. Lower bounds for data structures with space close to maximum imply circuit lower bounds. *Theory of Computing*, 15(1):1–9, 2019. doi:10.4086/toc.2019.v015a018.

42 Feng Wei. Upper bound for intermediate singular values of random matrices. *Journal of Mathematical Analysis and Applications*, 445(2):1530–1547, 2017.

43 Andrew Chi-Chih Yao. Should tables be sorted? *Journal of the ACM (JACM)*, 28(3):615–628, 1981. doi:10.1145/322261.322274.

A Missing proof to Lemma 13

The probability we are going to bound is the probability that any of the dimension of the $\mathbf{M}^k[u,v]$ is not close to $a/2n^{\alpha'}$, for integers a. It is non-trivial to bound since $\mathbf{M}^k$ does not follow a uniform distribution.

Specifically, we will show that each dimension of $\mathbf{M}^k$ is close to such a value with probability $\leq 1/\text{poly}(n)$. The desired probability bound follows by a union bound. For each dimension $[u, v]$ of $\mathbf{M}^k$, we are going to show that, even when all the other elements of $\mathbf{M_U}$ are fixed by an arbitrarily partial assignment σ, the function $\phi_{k,\sigma} : \mathbb{R} \to \mathbb{R}$ that maps from $\mathbf{M_U}[u, v]$ to $\mathbf{M}^k[u, v]$ evaluates to a value that is close to $a/2n^{\alpha'}$ with low probability. This is because $\phi_{k,\sigma}$ is an increasing function with its derivative well-controlled everywhere.

First, we need the following partial derivative bounds that generalizes Lemma 15 and Lemma 16, and works for every pair of indices $[u, v], [i, j]$. Recall that we assume that each ${\mathbf{M_U}}^k[u, v]$ is a function on n^2 variables $\mathbf{M_U}[i, j]$. For simplicity, we study for now their partial derivatives without the constraint of being a symmetric matrix. Given the symmetric constraint, their partial derivative is exactly $\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} + \frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[j,i]}$ when $i \neq j$, and is $\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]}$ otherwise.

▶ **Lemma 44.** *Let $\gamma > 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every integer $2 \leq k \leq n$, for every element $[i, j]$ of $\mathbf{M_U}$ and every element $[u, v]$ of ${\mathbf{M_U}}^k$, we have the following bounds of $\frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]}$:*

$$\frac{\partial {\mathbf{M_U}}^k[u, v]}{\partial \mathbf{M_U}[i, j]} = \begin{cases} \Theta(\frac{n}{n^2}), & u = i \text{ or } v = j \\ \Theta(\frac{k-2}{n^2}), & \text{otherwise} \end{cases}$$

If $k = 1$, we have:

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

Proof. We generalize Lemma 15 to every pair of vertices $[u, v]$ and $[i, j]$. Recall that

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

Again, $\prod_{t=0,t\neq g}^{k-1} \mathbf{M_U}[h_t, h_{t+1}]$ 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 bounds all elements are equivalent. When g is picked from $\{1, 2, ..., k-2\}$, there are $(k-2)n^{k-3}$ distinct sequences $h_0, \ldots, h_k$ for each summation. g can be 0 (resp., $k-1$) only if u (resp., v) coincides with i (resp., j). More precisely, the number of extra and distinct sequences for both cases $u = i$ and $j = v$ are exactly n^{k-2}.

Therefore, when $u \neq i$ and $v \neq j$,

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

and

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

by Observation 14 and $\gamma > 2$.

Similarly, for $u = i$ or $v = j$ but not both,

$$(1 - o(1))\frac{n+k-2}{n^2} \leq \frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} \leq \frac{n+k-2}{n^2}$$

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

$$(1 - o(1))\frac{2n+k-2}{n^2} \leq \frac{\partial {\mathbf{M_U}}^k[u,v]}{\partial \mathbf{M_U}[i,j]} \leq \frac{2n+k-2}{n^2}$$ ◀

The following lemma generalizes Lemma 16.

▶ **Lemma 45.** *Let $\gamma > 2$ and $\beta \geq \gamma + 2$, and the values of elements of $\mathbf{M_U}$ range from $[\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]$. For every integer $3 \leq k \leq n$, for every element $[i,j]$ of $\mathbf{M}$ and every element $[u,v]$ of $\mathbf{M}^k$, we have the following bound of $\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]}$:*

$$\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} = \begin{cases} \Theta(k), & u = i \text{ and } v = j \\ \Theta(\frac{k^2}{n^{1+\beta}}), & u = i \text{ or } v = j, \text{ but not both} \\ \Theta(\frac{k^3}{n^{1+2\beta}}), & \text{otherwise} \end{cases}.$$

Proof. We generalize Lemma 16 to every pair of indices $[u,v]$ and $[i,j]$. Recall that

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

which is dominated by $t = 1$ when $u = i$, $v = j$; dominated by $t = 2$ when $u = i$ or $v = j$, but not both; and dominated by the term of $t = 3$ otherwise.

Combined with Lemma 44, we have the following bounds. When $u = i, v = j$,

$$\Omega(k) \leq \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} \leq k$$

When $u = i$ or $v = j$, but not both,

$$\Omega\left(\frac{k^2}{n^{1+\beta}}\right) \leq \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} \leq \frac{k^2(n+k-2)}{n^{2+\beta}}$$

When $u \neq i$ and $v \neq j$,

$$\Omega\left(\frac{k^3}{n^{1+2\beta}}\right) \leq \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[i,j]} \leq \frac{k^3(2n+k-2)}{n^{2+2\beta}}$$ ◀

Now, we are ready to prove Lemma 13.

Proof to Lemma 13. Recall that we are going to show that the following happens with high probability: for every $a \in \mathbb{N}$ and every $u, v \in [n]$, $|\mathbf{M}^k[u,v] - a/2n^{\alpha'}| > 1/n^\alpha$. We bound its fail probability for every fixed u, v. The desired bound follows by a union bound.

Let σ denote a partial assignment to all the other elements of $\mathbf{M_U}$ except $\mathbf{M_U}[u,v], \mathbf{M_U}[v,u]$. And we let $\phi_{k,\sigma} : \mathbb{R} \to \mathbb{R}$ a function that maps $\mathbf{M_U}[u,v] = \mathbf{M_U}[v,u]$ to $\mathbf{M}^k[u,v]$. We have

$$\Pr_{\mathbf{M} \leftarrow \mathcal{M}}[\exists a \in \mathbb{N}, |\mathbf{M}^k[u,v] - a/2n^{\alpha'}| \leq 1/n^\alpha] \leq \max_\sigma \Pr_{x \in [\frac{1}{n} - \frac{1}{n^\gamma}, \frac{1}{n}]}[\exists a \in \mathbb{N}, |\phi_{k,\sigma}(x) - a/2n^{\alpha'}| \leq 1/n^\alpha].$$

There are two cases. When $u = v$,

$$\frac{\partial \phi_{k,\sigma}}{\partial x} = \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M_U}[u,v]} = \frac{1}{n^\beta} \cdot \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[u,v]} = \Theta(k/n^\beta).$$

When $u \neq v$,

$$\frac{\partial \phi_{k,\sigma}}{\partial x} = \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M_U}[u,v]} + \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M_U}[v,u]} = \frac{1}{n^\beta} \cdot \left(\frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[u,v]} + \frac{\partial \mathbf{M}^k[u,v]}{\partial \mathbf{M}[v,u]} \right) = \Theta(k/n^\beta).$$

Let us call $D(a/2n^{\alpha'}, 1/n^\alpha)$ a disc, which is the set of real numbers that is within $1/n^\alpha$ distance from $a/2n^{\alpha'}$. We also need to lower bound the range length of $\mathbf{M}^k[u,v]$, which gives the lower bound of how many discs centered at $a/2n^{\alpha'}$ that f will meet. Since the range length of $\mathbf{M_U}[u,v]$ is $1/n^\gamma$, the range length of $\mathbf{M}^k[u,v]$ is $\Theta(k/n^{\gamma+\beta})$. Therefore, there will be $\Theta(kn^{\alpha'-\gamma-\beta})$ discs.

For each disc, the range length of $\mathbf{M_U}[u,v]$ such that $\phi_{k,\sigma}$ evaluates to this disc is at most $\Theta(n^\beta/kn^\alpha)$. Therefore, the total volume of $\mathbf{M_U}[u,v]$ such that $\phi_{k,\sigma}$ falls in at least one discs is at most $\Theta(n^{\alpha'-\gamma-\alpha})$. Therefore, by union bound,

$$\begin{aligned}
&\Pr_{\mathbf{M}\leftarrow\mathcal{M}}[\exists u,v \in [n], \exists a \in \mathbb{N}, |\mathbf{M}^k[u,v] - a/2n^{\alpha'}| \le 1/n^\alpha] \\
\le& n^2 \max_\sigma \Pr_{x\in[\frac{1}{n}-\frac{1}{n^\gamma},\frac{1}{n}]}[\exists a \in \mathbb{N}, |\phi_{k,\sigma}(x) - a/2n^{\alpha'}| \le 1/n^\alpha] \\
\le& n^2 \cdot \Theta(n^{\alpha'-\gamma-\alpha})/(1/n^\gamma) \\
=& \Theta(n^{\alpha'-\alpha+2})
\end{aligned}$$

which is $o(1)$ when $\alpha > 2 + \alpha'$. ◀

B Missing proof to Lemma 21

We prove Lemma 21 in this section, which is a direct application of the concentration bound on intermediate singular values given in [34, 42]. Below are definitions used in presenting the concentration bound.

▶ **Definition 46.** *Let Z be a random variable. Then the ψ_2-norm (subgaussian norm) of Z is defined as*

$$\|Z\|_{\psi_2} := \inf\left\{\lambda > 0 : \mathbb{E}\left[\exp\left(\frac{|Z|^2}{\lambda^2}\right)\right] \le 2\right\}$$

▶ **Lemma 47** (Corollary 1.11. of [42])**.** *Let $\mathbf{A}$ an $m \times m$ random matrix with i.i.d. entries that have mean 0, variance 1 and ψ_2-norm K. Assume also that there exist a constant $c(K) > 0$ that depends only on K, and positive numbers $p > 0$, $s \le c(K)\min(1/p, 1)$ such that for every $i, j \in [m]$,*

$$\sup_{u\in\mathbb{C}} \Pr[|\mathbf{A}[i,j] - u| \le s] \le ps. \tag{7}$$

Then there exist $0 < C_1 < C_2$ and $C_3 > 0$ that depend only on K and p such that for all $\ell \in [m]$,

$$\Pr\left[\frac{C_1\ell}{\sqrt{m}} \le \sigma_{m+1-\ell}(\mathbf{A}) \le \frac{C_2\ell}{\sqrt{m}}\right] \ge 1 - \exp(-C_3\ell).$$

Lemma 21 follows by showing that both p and K are constants in our distribution. We also use the following inequality.

▶ **Proposition 48** (Young's convolution inequality [4]). *Suppose f is in the Lebesgue space $L^p(\mathbb{R})$ and g is in $L^q(\mathbb{R})$ and $1/p + 1/q = 1/r + 1$ with $1 \le p, q, r \le \infty$. Then*

$$\|f * g\|_r \le \|f\|_p \|g\|_q$$

where $(f * g)(t) := \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$ is the convolution of f and g.

Proof to Lemma 21. Recall that μ is either μ_1 or μ_2, the distributions of $X_1+Y_1+Z_1+W_1$ or $X_2Y_2 + Z_2W_2$, where X_i, Y_i, Z_i, W_i are i.i.d. uniform on some symmetric intervals chosen so that the resulting distribution has variance 1. Concretely,

$$X_1, Y_1, Z_1, W_1 \leftarrow \text{Unif}\left[-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right] \qquad X_2, Y_2, Z_2, W_2 \leftarrow \text{Unif}\left[-\left(\frac{9}{2}\right)^{1/4}, \left(\frac{9}{2}\right)^{1/4}\right]$$

We will apply Lemma 47 to $\mathbf{A}$ and show that C_1, C_2, C_3 are constants, i.e., in both cases $\mu = \mu_1$ or $\mu = \mu_2$, p, K are constants.

To show that the ψ_2-norm K of $\mathbf{A}[i,j]$ is a constant, note that μ is supported on a constant-bounded interval in both cases. $K = \Theta(1)$ follows directly from the fact that bounded random variables are subgaussian [40].

What remains is to show that there exists p such that for a possibly small constant $c(K)$, inequality (7) holds. To that end, we give upper bounds to the probability density functions (PDF) f_{μ_1} and f_{μ_2}. Specifically, $\|f_\mu\|_\infty \le M$ for an absolute constant M implies that

$$\sup_{u\in\mathbb{R}} \Pr[|\mathbf{A}[i,j] - u| \le s] = \sup_{u\in\mathbb{R}} \int_{u-s}^{u+s} f_\mu(t)dt \le 2s\|f_\mu\|_\infty \le (2M)s \tag{8}$$

where it suffices to take $u \in \mathbb{R}$ since $\mathbf{A}[i,j]$ is real-valued. Thus (7) holds with $p := 2M$.

Case $\mu = \mu_1$. Let $a := \sqrt{3}/2$. The uniform density on $[-a, a]$ is $f_0(t) = \frac{1}{2a}\mathbf{1}_{\{|t|\le a\}}$, hence $\|f_0\|_\infty = \frac{1}{2a} = \frac{1}{\sqrt{3}}$. The density of $X_1 + Y_1 + Z_1 + W_1$ is the 4-fold convolution $f_{\mu_1} = f_0 * f_0 * f_0 * f_0$. Note that $\|f_0\|_1 = 1$ since f_0 is a PDF. By repeatedly applying Young's convolution inequality $\|f * g\|_\infty \le \|f\|_\infty \cdot \|g\|_1$,

$$\|f_{\mu_1}\|_\infty = \|(f_0 * f_0 * f_0) * f_0\|_\infty \le \|f_0 * f_0 * f_0\|_\infty \cdot \|f_0\|_1 \le \|f_0\|_\infty \cdot 1 = \frac{1}{\sqrt{3}}.$$

Case $\mu = \mu_2$. Let $b := \left(\frac{9}{2}\right)^{1/4}$ and define $U := X_2Y_2$. Let $f_{X,Y}(x,y) = \frac{1}{4b^2}$ be the joint PDF of X_2, Y_2, for $|x|, |y| \le b$.

To obtain a clean formula of the PDF $f_U(u)$, we use a change of variables. Let $(U, V) = (X_2Y_2, X_2)$. Then $x = v, y = u/v$. Notice the following Jacobian

$$\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1/v & -u/v^2 \end{vmatrix} = \frac{1}{|v|},$$

which implies that $f_{U,V}(u,v) = f_{X,Y}(x,y) \cdot \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = f_{X,Y}(v, \frac{u}{v}) \cdot \frac{1}{|v|}$. Hence,

$$\begin{aligned} f_U(u) &= \int_{|v|\le b,\ |u/v|\le b} f_{U,V}(u,v)dv = \int_{|v|\le b,\ |u/v|\le b} f_{X,Y}\left(v, \frac{u}{v}\right)\frac{1}{|v|}dv \\ &= \int_{|v|\in[|u|/b, b]} \frac{1}{4b^2}\frac{1}{|v|}dv = \frac{1}{2b^2}\ln\frac{b^2}{|u|} \end{aligned}$$

for $0 < |u| \leq b^2$, and

$$\|f_U\|_2^2 = \int_{-b^2}^{b^2} |f_U(u)|^2 du = \frac{1}{2b^4} \int_0^{b^2} \ln^2\left(\frac{b^2}{u}\right) du = \frac{1}{2b^2} \int_0^{\infty} t^2 e^{-t} dt = \frac{1}{b^2}$$

where we substitute $u = b^2 e^{-t}$ and use the fact that $\Gamma(3) = \int_0^{\infty} t^2 e^{-t} dt = 2$.

Let $U' := Z_2 W_2$. Note that U and U' are i.i.d. By applying Young's convolution inequality, we have

$$\|f_{\mu_2}\|_\infty = \|f_U * f_{U'}\|_\infty \leq \|f_U\|_2 \cdot \|f_{U'}\|_2 = \frac{1}{b^2} = \frac{\sqrt{2}}{3}$$

In both cases, $M := \frac{\sqrt{3}}{3} \geq \|f_\mu\|_\infty$. By (8), we have

$$\sup_{u \in \mathbb{R}} \Pr[|\mathbf{A}[i,j] - u| \leq s] \leq ps$$

for $p = \frac{2\sqrt{3}}{3}$ and $s \leq c(K)/p$. By Lemma 47, C_1, C_2, C_3 are constants. ◀