Approximate $\mathbb{F}_2$-Sketching of Valuation Functions

Grigory Yaroslavtsev
Indiana University, Bloomington, IN, USA
The Alan Turing Institute, London, UK
gyarosla@iu.edu

Samson Zhou
Indiana University, Bloomington, IN, USA
samsonzhou@gmail.com

─── **Abstract** ───

We study the problem of constructing a linear sketch of minimum dimension that allows approximation of a given real-valued function $f \colon \mathbb{F}_2^n \to \mathbb{R}$ with small expected squared error. We develop a general theory of linear sketching for such functions through which we analyze their dimension for most commonly studied types of valuation functions: additive, budget-additive, coverage, α-Lipschitz submodular and matroid rank functions. This gives a characterization of how many bits of information have to be stored about the input x so that one can compute f under additive updates to its coordinates.

Our results are tight in most cases and we also give extensions to the distributional version of the problem where the input $x \in \mathbb{F}_2^n$ is generated uniformly at random. Using known connections with dynamic streaming algorithms, both upper and lower bounds on dimension obtained in our work extend to the space complexity of algorithms evaluating $f(x)$ under long sequences of additive updates to the input x presented as a stream. Similar results hold for simultaneous communication in a distributed setting.

1 Introduction

Linear sketching is a fundamental tool in efficient algorithm design that has enabled many of the recent breakthroughs in fast graph algorithms and computational linear algebra. It has a wide range of applications, including randomized numerical linear algebra (see survey [56]), graph sparsification (see survey [44]), frequency estimation [3], dimensionality reduction [34], various forms of sampling, signal processing, and communication complexity. In fact, linear sketching has been shown to be the optimal algorithmic technique [41, 2] for dynamic data streams, where elements can be both inserted and deleted. Linear sketching is also a frequently used tool in distributed computing – summaries communicated between the processors in massively parallel computational models are often linear sketches.

In this paper we introduce a study of approximate linear sketching over $\mathbb{F}_2$ (approximate $\mathbb{F}_2$-sketching). This is a previously unstudied but natural generalization of the work of [35], which studies exact $\mathbb{F}_2$-sketching. For a set $S \subseteq [n]$ let $\chi_S \colon \mathbb{F}_2^n \to \mathbb{F}_2$ be a parity function defined as $\chi_S(x) = \sum_{i \in S} x_i$. Given a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$, we are looking for a distribution

over k subsets $\mathbf{S}_1, \ldots, \mathbf{S}_k \subseteq [n]$ such that for any input x, it should be possible to compute $f(x)$ with expected squared error at most ϵ from the parities $\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)$ computed over these sets. While looking only at linear functions over $\mathbb{F}_2$ as candidate sketches for evaluating f might seem restrictive, this view turns out to be optimal in a number of settings. In the light of recent results of [35, 32], the complexity of $\mathbb{F}_2$-sketching also characterizes the space complexity of streaming algorithms in the XOR-update model as well as communication complexity of one-way multiplayer broadcasting protocols for XOR-functions.

In matrix form, $\mathbb{F}_2$-sketching corresponds to multiplication over $\mathbb{F}_2$ of the row vector $x \in \mathbb{F}_2^n$ by a random $n \times k$ matrix whose i-th column is the characteristic vector of $\chi_{\mathbf{S}_i}$:

$$
\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ \chi_{\mathbf{S}_1} & \chi_{\mathbf{S}_2} & \cdots & \chi_{\mathbf{S}_k} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} \chi_{\mathbf{S}_1}(x) & \chi_{\mathbf{S}_2}(x) & \ldots & \chi_{\mathbf{S}_k}(x) \end{pmatrix}
$$

The goal is to minimize k, ensuring that the sketch alone is sufficient for computing f with expected squared error at most ϵ for any fixed input x. For a fixed distribution D of x, the definition of error is modified to include an expectation over D in the error guarantee. We give formal definitions below.

▶ **Definition 1** (Exact $\mathbb{F}_2$-sketching, [35]). *The* exact randomized $\mathbb{F}_2$-sketch complexity *with error δ of a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$ (denoted as $R_\delta^{lin}(f)$) is the smallest integer k such that there exists a distribution $\chi_{\mathbf{S}_1}, \chi_{\mathbf{S}_2}, \ldots, \chi_{\mathbf{S}_k}$ over k linear functions over $\mathbb{F}_2^n$ and a post-processing function $g : \mathbb{F}_2^k \to \mathbb{R}$ that satisfies:*

$$
\forall x \in \mathbb{F}_2^n : \Pr_{\mathbf{S}_1, \ldots, \mathbf{S}_k} [g(\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)) = f(x)] \geq 1 - \delta.
$$

The number of parities k in the definition above is referred to as the *dimension* of the $\mathbb{F}_2$-sketch.

▶ **Definition 2** (Approximate $\mathbb{F}_2$-sketching). *The ϵ-approximate randomized $\mathbb{F}_2$-sketch complexity of a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$ (denoted as $\bar{R}_\epsilon^{lin}(f)$) is the smallest integer k such that there exists a distribution $\chi_{\mathbf{S}_1}, \chi_{\mathbf{S}_2}, \ldots, \chi_{\mathbf{S}_k}$ over k linear functions over $\mathbb{F}_2^n$ and a post-processing function $g : \mathbb{F}_2^k \to \mathbb{R}$ that satisfies:*

$$
\forall x \in \mathbb{F}_2^n : \mathbb{E}_{\mathbf{S}_1, \ldots, \mathbf{S}_k} \left[(g(\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)) - f(x))^2 \right] \leq \epsilon
$$

If g is an unbiased estimator of f, then this corresponds to an upper bound on the variance of the estimator. For example, functions with small spectral norm (e.g. coverage functions, [57]) admit such approximate $\mathbb{F}_2$-sketches. Moreover, observe that Definition 2 is not quite comparable with an epsilon-delta guarantee, which only promises that $|g(\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)) - f(x)| \leq \epsilon$ with probability $1 - \delta$, but guarantees nothing for δ fraction of the inputs.

In addition to this worst-case guarantee, we also consider the same problem for x from a certain distribution. In this case, a weaker guarantee is required, i.e. the bound on expected squared error should hold only over some fixed known distribution D. An important case is $\mathsf{D} = U(\mathbb{F}_2^n)$, the uniform distribution over all inputs.

▶ **Definition 3** (Approximate distributional $\mathbb{F}_2$-sketching). *For a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$, we define its ϵ-approximate randomized distributional $\mathbb{F}_2$-sketch complexity with respect to a distribution D over $\mathbb{F}_2^n$ (denoted as $\bar{\mathcal{D}}_\epsilon^{lin,D}(f)$) as the smallest integer k such that there exists a distribution $\chi_{\mathbf{S}_1}, \chi_{\mathbf{S}_2}, \ldots, \chi_{\mathbf{S}_k}$ over k linear functions over $\mathbb{F}_2$ and a post-processing function $g : \mathbb{F}_2^k \to \mathbb{F}_2$ that satisfies:*

$$\mathop{\mathbb{E}}_{x \sim D} \mathop{\mathbb{E}}_{\mathbf{S}_1,\ldots,\mathbf{S}_k} \left[(g(\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)) - f(x))^2 \right] \leq \epsilon.$$

1.1 Applications to Streaming and Distributed Computing

One of the key applications of our results is to the dynamic streaming model. In this model, the input x is generated via a sequence of additive updates to its coordinates, starting with $x = 0^n$. If $x \in \mathbb{R}^n$, then updates are of the form (i, Δ_i) (turnstile model), where $i \in [n]$, and $\Delta_i \in \mathbb{R}$, which adds Δ_i to the i-th coordinate of x. For $x \in \mathbb{F}_2^n$, only the coordinate i is specified and the corresponding bit is flipped, which is known as the XOR-update model [52][1]. Dynamic streaming algorithms aim to minimize space complexity of computing a given function f for an input generated through a sequence of such updates while also ensuring fast update and function evaluation times.

Note that linear sketching over the reals and $\mathbb{F}_2$-sketching can be used directly in the respective streaming update models. Most interestingly, these techniques turn out to achieve almost optimal space complexity. It is known that linear sketching over the reals gives (almost) optimal space complexity for processing dynamic data streams in the turnstile model for *any* function f [41, 2]. However, the results of [41, 2] require adversarial streams of length triply exponential in n. In the XOR-update model, space optimality of $\mathbb{F}_2$-sketching has been shown recently in [32]. This optimality result holds even for adversarial streams of much shorter length $\Omega(n^2)$. Hence, lower bounds on $\mathbb{F}_2$-sketch complexity obtained in our work extend to space complexity of dynamic streaming algorithms for streams of quadratic length.

A major open question in this area is the conjecture of [35] that the same holds even for streams of length only $2n$. We thus complement our lower bounds on dimension of $\mathbb{F}_2$-sketches with one-way two-player communication complexity lower bounds for the corresponding XOR functions $f^+(x, y) = f(x + y)$. Such lower bounds translate to dynamic streaming lower bounds for streams of length $2n$. Furthermore, whenever our communication lower bounds hold for the uniform distribution, the corresponding streaming lower bound applies to streaming algorithms under uniformly random input updates.

Finally, our upper bounds can be used for distributed algorithms computing $f(x_1 + \cdots + x_M)$ over a collection of distributed inputs $x_1, \ldots, x_M \in \mathbb{F}_2^n$ as $\mathbb{F}_2$-sketches can be used for distributed inputs. On the other hand, our communication lower bounds also apply to the simultaneous message passing (SMP) communication model, since it is strictly harder than one-way communication.

---

[1] By slightly changing the function to $f'(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$, it is easy to see that there are functions for which knowledge of the sign of the update (i.e. whether it is +1 or -1) is not a stronger model than the XOR-update model. For some further motivation of the XOR-update model, consider dynamic graph streaming algorithms, i.e the setting when x represents the adjacency matrix of a graph and updates correspond to adding and removing the edges. Almost all known dynamic graph streaming algorithms (except spectral graph sparsification of [36]) are based on the ℓ_0-sampling primitive [26]. As shown recently, ℓ_0-sampling can be implemented optimally using $\mathbb{F}_2$-sketches [37] and hence almost all known dynamic graph streaming algorithms can handle XOR-updates, i.e. knowing whether an edge was inserted or deleted does not help.

1.2 Valuation Functions and Sketching

Submodular valuation functions, originally introduced in the context of algorithmic game theory and optimization, have received a lot of interest recently in the context of learning theory [10, 9, 18, 29, 47, 23, 22, 24, 25][2], approximation [27, 7] and property testing [16, 49, 12]. As we show in this work, valuation functions also represent an interesting study case for linear sketching and streaming algorithms. While a variety of papers exists on streaming algorithms for optimizing various submodular objectives, e.g. [48, 20, 8, 17, 15, 21, 30, 4, 11], to the best of our knowledge no prior work considers the problem of evaluating such functions under a changing input.

A systematic study of $\mathbb{F}_2$-sketching has been initiated for Boolean functions in [35]. This paper can be seen as a next step, as we introduce approximation into the study of $\mathbb{F}_2$-sketching. One of the consequences of our work is that the Fourier ℓ_1-sampling technique, originally introduced by Bruck and Smolensky [14] (see also [28, 45]), turns out to be optimal in its dependence on both spectral norm and the error parameter. For Boolean functions, a corresponding result is not known as Boolean functions with small spectral norm and necessary properties are hard to construct. Another technical consequence of our work is that the study of learning and sketching algorithms turn out to be related on a technical level despite pursuing different objectives (in learning the specific function is unknown, while in sketching it is). In particular, our hardness result for Lipschitz submodular functions uses a construction of a large family of matroids from [10] (even though in a very different parameter regime), who designed such a family to fool learning algorithms.

1.3 Our Results

A function $f\colon 2^{[n]} \to \mathbb{R}$ is α-Lipschitz if for any $S \subseteq [n]$ and $i \in [n]$, it holds that $|f(S \cup \{i\}) - f(S)| \leq \alpha$ for some constant $\alpha > 0$. A function $f\colon 2^{[n]} \to \mathbb{R}$ is submodular if:

$$f(A \cup \{i\}) - f(A) \geq f(B \cup \{i\}) - f(B) \qquad \forall A \subseteq B \subseteq [n] \text{ and } i \notin B.$$

We consider the following classes of valuation functions of the form $f\colon \mathbb{F}_2^n \to \mathbb{R}$ (all of them submodular) sometimes treating them as $f\colon 2^{[n]} \to \mathbb{R}$ and vice versa. These classes mostly cover all of existing literature on submodular functions[3]. See Table 1 for a summary of the results.

- **Additive (linear).** $f(x) = \sum_{i=1}^{n} w_i x_i$, where $w_i \in \mathbb{R}$.
  Our results: For additive functions, it is easy to show that dimension of $\mathbb{F}_2$-sketches is $O(\min(\|w\|_1^2/\epsilon, n))$ and give a matching communication lower bound for all $\epsilon \geq \|w\|_2^2$ [57].
- **Budget-additive.** $f(x) = \min(b, \sum_{i=1}^{n} w_i x_i)$ where $b, w_i \in \mathbb{R}$. An example of such functions is the "hockey stick" function $hs_\alpha(x) = \min(\alpha, \frac{2\alpha}{n} \sum_{i=1}^{n} x_i)$.
  Our results: For budget-additive functions, it is easy to show that dimension of $\mathbb{F}_2$-sketches is $O(\min(\|w\|_1^2/\epsilon, n))$ [57]. We give a matching communication bound for the "hockey stick" function for constant ϵ [57], which holds even under the uniform distribution of the input.

---

[2] We remark that in this literature the term "sketching" is used to refer to the space complexity of representing the function f itself under the assumption that it is unknown but belongs to a certain class. This question is orthogonal to our work as we assume f is known and fixed while the input x is changing.

[3] We do not discuss some other subclasses of subadditive functions because they are either superclasses of classes for which we already have an $\Omega(n)$ lower bound (e.g. submodular, subadditive, etc.) or because such a lower bound follows trivially (e.g. for OXS/XOS since for XS-functions a lower bound of $\Omega(n)$ is easy to show, see [57]).

Table 1 Linear sketching complexity of classes of valuation functions. We defer the proofs of several results to the full version [57].

Class	Error	Distribution	Complexity	Result
Additive/Budget additive $\min(b, \sum_{i=1}^n w_i x_i)$	ϵ	any	$\Theta\left(\frac{\|w\|_1^2}{\epsilon}\right)$	[57]
$\min(c\sqrt{n}, \frac{2c}{\sqrt{n}}\sum_{i=1}^n x_i)$	constant	uniform	$\Omega(n)$	[57]
Coverage	ϵ	any	$O\left(\frac{1}{\epsilon}\right)$	[57]
Matroid Rank 2	exact	any	$\Theta(1)$	Theorem 12
Graphic Matroids Rank r	exact	any	$O(r^2 \log r)$	Theorem 16
Matroid Rank r	exact	any	$\Omega(r)$	Corollary 35
Matroid Rank r	exact	uniform	$O((r \log r + c)^{r+1})$	[57]
Matroid Rank	$1/\sqrt{n}$	uniform	$\Theta(1)$	[57]
$\frac{c}{n}$-Lipschitz Submodular	constant	any	$\Theta(n)$	Theorem 28

- **Coverage.** A function f is a *coverage function* on some universe U of size m if there exists a collection $A_1, \ldots, A_n$ of subsets of U and a vector of non-negative weights $(w_1, \ldots, w_m)$ such that:

$$f(S) = \sum_{i \in \cup_{j \in S} A_j} w_i.$$

  Our results: We show a simple upper bound of $O(1/\epsilon)$ for such functions [57].
- **Matroid rank.** A pair $M = ([n], \mathcal{I})$ is called a matroid if $\mathcal{I} \subseteq 2^{[n]}$ is a non-empty set family such that the following two properties are satisfied:
  - If $I \in \mathcal{I}$ and $J \subseteq I$, then $J \in \mathcal{I}$
  - If $I, J \in \mathcal{I}$ and $|J| < |I|$, then there exists an $i \in I \setminus J$ such that $J \cup \{i\} \in \mathcal{I}$.
  The sets in $\mathcal{I}$ are called *independent*. A maximal independent set is called a *base* of M. All bases have the same size, which is called the *rank* of the matroid and is denoted as $rk(M)$. The *rank function* of the matroid is the function $rank_M \colon 2^{[n]} \to \mathbb{N}_+$ defined as:

$$rank_M(S) := \max\{|I| \colon I \subseteq S, I \in \mathcal{I}\}.$$

  It follows from the definition that $rank_M$ is always a submodular 1-Lipschitz function.
  Our results: In order to have consistent notation with the rest of the manuscript we always assume that matroid rank functions are scaled so that their values are in $[0, 1]$. Some of our results are exact, i.e. the corresponding matroid rank function is computed exactly (and in this case rescaling does not matter) while others allow approximation of the function value. In the latter case, the approximation guarantees are multiplicative with respect to the rescaled function.
  Our main theorem regarding sketching of matroid rank functions is as follows:
  ▶ **Theorem 4** (Sketching matroid rank functions). *For (scaled) matroid rank functions:*
  - *There exists an exact $\mathbb{F}_2$-sketch of size $O(1)$ for matroids of rank 2 (Theorem 12) and graphic matroids (Theorem 16).*
  - *There exists $c = \Omega(1)$ and a matroid of rank r such that a c-approximation of its matroid rank function has randomized linear sketch complexity $\Omega(r)$. Furthermore, this lower bound also holds for the corresponding one-way communication problem (Theorem 34, Corollary 35).*

This can be contrasted with the results under the uniform distribution for which matroids of rank r have an exact $\mathbb{F}_2$-sketch of size $O\left(\left(r\log r + \log\frac{1}{\epsilon}\right)^{r+1}\right)$, where ϵ is the probability of failure ([57], follows from the junta approximation of [13]). Furthermore, matroids of high rank $\Omega(n)$ can be trivially approximately sketched under product distributions, due to their concentration around their expectation (see [57] for details).

- **Lipschitz submodular.** A function $f\colon 2^{[n]} \to \mathbb{R}$ is α-Lipschitz submodular if it is both submodular and α-Lipschitz.

  Our results: We show an $\Omega(n)$ communication lower bound (and hence a lower bound on $\mathbb{F}_2$-sketch complexity) for constant error for monotone non-negative $O(1/n)$-Lipschitz submodular functions (Theorem 28). We note that this hardness result crucially uses a non-product distribution over the input variables since Lipschitz submodular functions are tightly concentrated around their expectation under product distributions (see e.g. [55, 10]) and hence can be approximated using their expectation without any sketching at all.

1.4   Overview and Techniques

1.4.1   Basic Tools: XOR Functions, Spectral Norm, Approximate Fourier Dimension

In Section 2, we introduce the basics of approximate $\mathbb{F}_2$-sketching. Most definitions and results in this section can be seen as appropriate generalizations regarding Boolean functions (such as in [35]) to the case of real-valued functions where we replace Hamming distance with expected squared distance. We then define the randomized one-way communication complexity of the two-player XOR-function $f^+(x, y) = f(x + y)$ corresponding to f. This communication problem plays an important role in our arguments as it gives a lower bound on the sketching complexity of f. We then introduce the notion of approximate Fourier dimension developed in [35]. The key structural results of [35], which characterize both the sketching complexity of f and the one-way communication complexity of f^+ under the uniform distribution using the approximate Fourier dimension, can be extended to the real-valued case as shown in Proposition 10 and Theorem 11. This characterization is our main tool for showing lower bounds under the uniform distribution of x.

Another useful basic tool is a bound on the linear sketching complexity based on the spectral norm of f, which we develop in [57]. In particular, as we show in [57], analogously to the Boolean case, we can leverage properties of the Fourier coefficients of a function f to show that the ϵ-approximate randomized sketching complexity of f is at most $O(\|\hat{f}\|_1^2/\epsilon)$. Thus, we can determine the dimension of $\mathbb{F}_2$-sketches for classes of functions whose spectral norms are well-bounded as well as functions which can be computed as Lipschitz compositions of a small number of functions with bounded spectral norm [57]. Examples of such classes include additive (linear), budget-additive and coverage functions. Finally, we argue that the dependence on the parameters in the spectral norm bound cannot be substantially improved in the real-valued case by presenting a subclass of linear functions which require sketches of size $\Omega(\|\hat{f}\|_1^2/\epsilon)$ [57]. This is in contrast with the case of Boolean functions studied in [35] for which such tightness result is not known.

1.4.2   Matroid Rank Functions, LTF, LTF∘OR

In Section 3, we present our results on sketching matroid rank and Lipschitz submodular functions. In Section 3.1 we show that matroid rank functions of matroids of rank 2 and graphic matroids have constant randomized sketching complexity. This is done by first

observing that rank functions of such matroids can be expressed as a threshold function over a number of disjunctions. Therefore, it remains to determine the sketching complexity of the threshold function on a collection of disjunctions. Unfortunately, known upper bounds for the sketching complexity of even the simpler class of linear threshold functions have a dependence on n and hence one cannot get a constant upper bound directly.

Hence we show how to remove this dependence in Section 3.1.1, also resolving an open question of Montanaro and Osborne [45]. Recall that a linear threshold function (LTF) can be represented as $f(x) = \text{sgn} \left(\sum_{i=1}^{n} w_i x_i - \theta \right)$ for some weights w_i and threshold θ, where we slightly alter the traditional definition of the sign function sgn to output 0 if the input is negative and 1 otherwise. An important parameter of an LTF is its *margin m*, which corresponds to the difference between the threshold and the value of the linear combination closest to it. We first observe that the terms with insignificant coefficients, i.e. weights that are small in absolute value, do not contribute to the final output and thus, we can ignore them. Similarly, the remaining weights can be rounded, without altering the output of the function, to a collection of weights whose size is bounded, independent of n. Furthermore, $f(x) = 0$ only if $x_i = 1$ for at most $\frac{\theta}{2m}$ of these "significant" indices i of x. Thus, we hash the significant indices to a large, but independent of n, number of buckets. As a result, either there are a small number of significant indices i with $x_1 = 1$ and there are no collisions, or there is a large number of significant indices i with $x_i = 1$. Since we can differentiate between these two cases, the sketch can output whether $f(x) = 0$ or $f(x) = 1$ with constant probability. With a more careful choice of hash functions this idea can be extended to linear thresholds of disjunctions. We show in Section 3.1.2 that a threshold function over a number of disjunctions (LTF∘OR) also has linear sketch complexity independent of n.

In Section 3.2.1, we show that there exists an $\Omega(n)$-Lipschitz submodular function f that requires a randomized linear sketch of size $\Omega(n)$. We construct such a function probabilistically by using a large family of matroid rank functions constructed by [10] with an appropriately chosen set of parameters. We show any fixed deterministic sketch fails on a matroid chosen uniformly at random from this parametric family with very high probability. In fact, even if we take a union bound over all possible sketches of bounded dimension, the failure of probability is still negligibly close to 1. By Yao's principle, the randomized linear sketch complexity follows. We then extend this result to a communication lower bound for f^+ in Section 3.2.2. In the one-way communication complexity setting, we show that there exists an $\Omega(n)$-Lipschitz submodular function f whose f^+ requires communication $\Omega(n)$.

1.4.3   Uniform Distribution

In [57], we show lower bounds for a budget additive "hockey stick" function under the uniform distribution. The lower bounds follow from a characterization of communication complexity using approximate Fourier dimension, and to complete the analysis, we lower bound the Fourier spectrum of the hockey stick function in [57]. Although our approach for matroids of rank 2 does not seem to immediately generalize to matroids of higher rank under arbitrary distributions, we show in [57] that under the uniform distribution, we can use ϵ-approximations of disjunctive normal forms (DNFs) by juntas to obtain a randomized linear sketch whose size is independent of n. Furthermore, rank functions of matroids of very high rank admit trivial *approximate* sketches under the uniform distribution as follows from standard concentration results [55] (see [57]).

2   Basics of Approximate $\mathbb{F}_2$-Sketching

2.1   Communication Complexity of XOR functions

In order to analyze the optimal dimension of $\mathbb{F}_2$-sketches, we need to introduce a closely related communication complexity problem. For $f \colon \mathbb{F}_2^n \to \mathbb{R}$ define the XOR-function $f^+ \colon \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{R}$ as $f^+(x, y) = f(x + y)$ where $x, y \in \mathbb{F}_2^n$. Consider a communication game between two players Alice and Bob holding inputs x and y respectively. Given access to a shared source of random bits Alice has to send a single message to Bob so that he can compute $f^+(x, y)$. This is known as the one-way communication complexity problem for XOR-functions (see [50, 58, 45, 39, 40, 51, 42, 53, 43, 31, 35] for related communication complexity results).

▶ **Definition 5** (Randomized one-way communication complexity of XOR function). *For a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$, the* randomized one-way communication complexity *with error δ (denoted as $R_\delta^{\to}(f^+)$) of its XOR-function is defined as the smallest size[4] (in bits) of the (randomized using public randomness) message $M(x)$ from Alice to Bob, which allows Bob to evaluate $f^+(x, y)$ for any $x, y \in \mathbb{F}_2^n$ with error probability at most δ.*

It is easy to see that $R_\delta^{\to}(f^+) \le R_\delta^{lin}(f)$ as using shared randomness Alice can just send k bits $\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)$ to Bob, who can for each $i \in [k]$ compute $\chi_{\mathbf{S}_i}(x + y) = \chi_{\mathbf{S}_i}(x) + \chi_{\mathbf{S}_i}(y)$, which is an $\mathbb{F}_2$-sketch of f on $x + y$ and hence suffices for computing $f^+(x, y)$ with probability $1 - \delta$.

Replacing the guarantee of exactness of the output in the above definition with an upper bound on expected squared error, we obtain the following definition.

▶ **Definition 6** (Randomized one-way communication complexity of approximating an XOR function). *For a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$, the* randomized one-way communication complexity *(denoted as $\bar{R}_\epsilon^{\to}(f^+)$) of approximating its XOR-function with error ϵ is defined as the smallest size(in bits) of the (randomized using public randomness) message $M(x)$ from Alice to Bob, which allows Bob to evaluate $f^+(x, y)$ for any $x, y \in \mathbb{F}_2^n$ with expected squared error at most ϵ.*

Distributional communication complexity is defined analogously for the corresponding XOR function and is denoted as $\mathcal{D}_\epsilon$.

Finally, in the simultaneous model of computation [6, 5], also called simultaneous message passing (SMP) model, there exist two players and a coordinator, who are all aware of a function f. The two players maintain x and y respectively, and must send messages of minimal size to the coordinator so that the coordinator can compute $f(x \oplus y)$.

▶ **Definition 7** (Simultaneous communication complexity of XOR function). *For a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$, the* simultaneous one-way communication complexity *with error δ (denoted as $R_\delta^{sim}(f^+)$) of its XOR-function is defined as the smallest sum of the sizes (in bits) of the (randomized using public randomness) messages $M(x)$ and $M(y)$ from Alice and Bob, respectively, to a coordinator, which allows the coordinator to evaluate $f^+(x, y)$ for any $x, y \in \mathbb{F}_2^n$ with error probability at most δ.*

Observe that a protocol for randomized one-way communication complexity of XOR function translates to a protocol for the simultaneous model of computation.

---

[4] Formally the minimum here is taken over all possible protocols where for each protocol the size of the message $M(x)$ refers to the largest size (in bits) of such message taken over all inputs $x \in \mathbb{F}_2^n$. See [38] for a formal definition.

2.2 Distributional Approximate $\mathbb{F}_2$-Sketch Complexity

Fourier analysis plays an important role in the analysis of distributional $\mathbb{F}_2$-sketch complexity over the uniform distribution. In our discussion below, we make use of some standard facts from Fourier analysis of functions over $\mathbb{F}_2^n$. For definitions and basics of Fourier analysis of functions of such functions we refer the reader to the standard text [46] and [57]. In particular, Fourier concentration on a low-dimensional subspace implies existence of a small sketch which satisfies this guarantee:

▶ **Definition 8** (Fourier concentration). *A function $f \colon \mathbb{F}_2^n \to \mathbb{R}$ is γ-concentrated on a linear subspace A_d of dimension d if for this subspace it satisfies:*

$$\sum_{S \in A_d} \hat{f}(S)^2 \geq \gamma.$$

We also use the following definition of approximate Fourier dimension from [35], adapted for the case of real-valued functions.

▶ **Definition 9** (Approximate Fourier dimension). *Let $\mathcal{A}_k$ be the set of all linear subspaces of $\mathbb{F}_2^n$ of dimension k. For $f \colon \mathbb{F}_2^n \to \mathbb{R}$ and $\epsilon \in (0, \|f\|_2^2]$ the ϵ-approximate Fourier dimension $\dim_\epsilon(f)$ is defined as:*

$$\dim_\epsilon(f) = \min_k \left\{ \exists A \in \mathcal{A}_k \colon \sum_{\alpha \in A} \hat{f}^2(\alpha) \geq \epsilon \right\}.$$

▶ **Proposition 10.** *For any $f \colon \mathbb{F}_2^n \to \mathbb{R}$, it holds that:*

$$\bar{\mathcal{D}}_\epsilon^{lin,U}(f) \leq \dim_{\|f\|_2^2 - \epsilon}(f).$$

Furthermore, approximate Fourier dimension can be used as a lower bound on the one-way communication complexity of the corresponding XOR-function. We defer the proof of the following result to [57] as it is follows closely an analogous result for Boolean functions from [35].

▶ **Theorem 11.** *For any $f \colon \mathbb{F}_2^n \to \mathbb{R}$, $\delta \in [0, 1/2]$ and $\xi = \|f\|_2^2 - \epsilon(1 + 2\delta)$ it holds that:*

$$\bar{\mathcal{D}}_\epsilon^{\to,U}(f^+) \geq \frac{\delta}{2} \cdot \dim_\xi(f).$$

3 Sketching Matroid Rank Functions

In this section we analyze sketching complexity of matroid rank functions. We start by considering the most fundamental possible matroids (of rank 2) in Section 3.1 and showing that exactly sketching the matroid rank function requires $O(1)$ complexity. Similarly, we show that exactly sketching the rank of graphic matroids only uses $O(1)$ complexity. On the other hand, we show a lower bound in Section 3.2.1 that even approximating the rank r of general matroids up to certain constant factors requires $\Omega(r)$ complexity.

To sketch matroids of rank 2, we leverage a result by Acketa [1] which characterizes the collection of independent sets of such matroids. This allows us to represent matroid rank functions for matroids of rank 2 as a linear threshold of disjunctions. Thus, we first show the randomized linear sketch complexity of (θ, m)-linear threshold functions, resolving an open question by Montanaro and Osborne [45].

3.1 Matroids of Rank 2 and Graphic Matroids

In this section, we show that there exists a constant-size sketch that can be used to compute exact values of matroid rank functions for matroids of rank 2.

▶ **Theorem 12.** *For every matroid M of rank 2 it holds that $R^{lin}_{\frac{1}{3}}(rank_M) = O(1)$.*

It is well-known that matroids of rank 2 admit the following characterization (see e.g. [1]).

▶ **Fact 13.** *The collection of size 2 independent sets of a rank 2 matroid can be represented as the edges in a complete graph that has edges of some number of disjoint cliques removed.*

We define the following function as a threshold on the Hamming weight of a binary vector x

$$\mathsf{HAM}_{\leq d}(x) = \begin{cases} 0, & \text{if } \sum_{i=1}^n x_i \leq d + \frac{1}{2} \\ 1, & \text{otherwise.} \end{cases}$$

We use a series of technical lemmas in the following section to prove the following result, which says that linear threshold functions can be succinctly summarized:

▶ **Theorem 14.** *The function $\mathsf{HAM}_{\leq d}\left(\bigvee_{i \in S_1} x_i, \bigvee_{i \in S_2} x_i, \dots\right)$ has a randomized linear sketch of size $O(d^2 \log d)$.*

The following fact that upper bounds the sketch complexity for functions with small support:

▶ **Fact 15** (Folklore, see e.g. [45, 35]). *For any function $f : \{0,1\}^n \to \{0,1\}$ with $\min_{z \in \{0,1\}} \mathbf{Pr}_{x \in \{0,1\}^n}(f(x) = z) \leq \epsilon$ it holds that $R^{lin}_{\delta}(f) \leq \log \frac{2^{n+1}\epsilon}{\delta}$.*

Using Fact 13, Theorem 14, and Fact 15, we prove Theorem 12 by writing the matroid rank function for M as a linear threshold function of disjunctions.

Proof of Theorem 12. We first claim $\mathbb{F}_2$-sketching complexity of the rank function of any rank 2 matroid M is essentially the same as the complexity of the corresponding Boolean function that takes value 1 if $rank_M(x) = 2$ and takes value 0 otherwise. Indeed, let the function above be denoted as f_M. Without loss of generality, we can assume that all singletons are independent sets in M as otherwise the rank function of M does not depend on the corresponding input. Hence $rank_M(x) = 0$ if and only if $x = 0^n$. Thus $R^{lin}_{\delta}(rank_M) = R^{lin}_{\delta}(f_M) + O(\log 1/\delta)$ as by Fact 15 we can use $O(\log 1/\delta)$-bit sketch to check whether $x = 0^n$ first and then evaluate $rank_M$ using f_M. Recall from Fact 13 that matroids of rank 2 can be represented as edges in a complete graph with edges corresponding to some disjoint union of cliques removed.

Let $S_1, \dots, S_t$ be the collection of vertex sets of disjoint cliques defining a rank 2 matroid M in Fact 13. Without loss of generality, we can assume that $|\cup_{i=1}^t S_i| = n$ by adding singletons. Then:

$$f_M(x) = \mathsf{HAM}_{\geq 2}\left(\bigvee_{j \in S_1} x_j, \bigvee_{j \in S_2} x_j, \dots, \bigvee_{j \in S_t} x_j\right),$$

where $\mathsf{HAM}_{\geq 2}(z_1, \dots, z_t) = 1$ if and only if $\sum_{i=1}^t z_i \geq 2$ is the threshold Hamming weight function. By Theorem 14, the sketch complexity of $f_M(x)$ is $O(1)$, since the Hamming weight threshold is $d = 2$. ◀

Since the independent bases of a graphic matroid $M(G)$ are the spanning forests of G, the matroid rank function of a graphic matroid of rank r can be expressed as

$$f_M(x) = \mathsf{HAM}_{\geq r}\left(\bigvee_{j \in S_1} x_j, \bigvee_{j \in S_2} x_j, \ldots, \bigvee_{j \in S_t} x_j\right),$$

where each S_i is a separate spanning forest. Therefore, Theorem 14 yields a $O(r^2 \log r)$ space linear sketch for graphic matroids of rank r.

▶ **Theorem 16.** *For every graphic matroid M of rank r, it holds that $R_{\frac{1}{3}}^{lin}(rank_M) = O(r^2 \log r)$.*

We use the remainder of the Section 3.1 to prove Theorem 14, while resolving an open question by Montanaro and Osborne [45].

3.1.1 Linear Threshold Functions

We first define linear threshold functions (LTFs) and (θ, m)-LTFs.

▶ **Definition 17.** *A function $f : \{0,1\}^n \to \{0,1\}$ is a* linear threshold function *(LTF) if there exist constants $\theta, w_1, w_2, \ldots, w_n$ such that $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i x_i\right)$, where $\mathsf{sgn}(y) = 0$ for $y < 0$ and $\mathsf{sgn}(y) = 1$ for $y \geq 0$ is the Heaviside step function.*

▶ **Definition 18.** *A monotone linear threshold function $f : \{0,1\}^n \to \{0,1\}$ is a (θ, m)-LTF if $m \leq \min_{x \in \{0,1\}^n} |-\theta + \sum_{i=1}^n w_i x_i|$, where θ is referred to as the* threshold *and m as the* margin *of the LTF.*

Although (θ, m)-LTFs have previously been shown to have randomized linear sketch complexity $O\left(\frac{\theta}{m} \log n\right)$ [42], Montanaro and Osborne asked whether any (θ, m)-LTF can be represented in the simultaneous model with $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$ communication.

▶ **Question 19** ([45]). *Let $g(x, y) = f(x \oplus y)$, where f is a (θ, m)-LTF. Does there exist a protocol for g in the simultaneous model with communication complexity $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$?*

Note that the difference between $\log n$ and $\log \frac{\theta}{m}$ is crucial for obtaining constant randomized linear sketch complexity for functions for matroid rank 2. We answer Question 19 in the affirmitive and show the stronger result that (θ, m)-LTFs admit a randomized linear sketch of size $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$. We first show that we can completely ignore all variables whose weights are significantly smaller than $2m$ in evaluating a (θ, m)-LTF.

▶ **Lemma 20.** *Let $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i x_i\right)$ be a (θ, m)-LTF. For $1 \leq i \leq n$, let $w_i' = w_i$ if $w_i \geq 2m$ and $w_i' = 0$ otherwise. Then $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i' x_i\right)$.*

As noted, Lemma 20 implies that we can ignore not only variables with zero weights, but all variables whose weights are less than $2m$. We now bound the support of the set $\{x \mid f(x) = 0\}$, where f is a (θ, m)-LTF, and apply Fact 15.

▶ **Lemma 21.** *For any (θ, m)-LTF, there exists a randomized linear sketch of size $O\left(\frac{\theta}{m} \log n\right)$.*

Proof. Let $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i x_i\right)$ be a (θ, m)-LTF. By Lemma 20, the output of f remains the same even if we only consider the variables S with weight at least $2m$. On the other hand, if $f(x) = 0$, then at most $\frac{\theta}{2m}$ variables in S can have value 1. Equivalently, at most $\frac{\theta}{2m}$ indices i can have $x_i = 1$ if $f(x) = 0$. Thus, the number of $x \in \{0,1\}^n$ with $f(x) = 0$ is at most $\sum_{0 \leq i \leq \theta/2m} \binom{n}{i} \leq (n+1)^{\lceil \theta/2m \rceil}$. Applying Fact 15, there exists a randomized linear sketch for f, of size $O\left(\frac{\theta}{m} \log n\right)$. ◀

In order to fully prove Question 19 and obtain a dependence on $\log \frac{\theta}{m}$ rather than $\log n$, we use the following two observations. First, we show in Lemma 22 that the weights of a (θ, m)-LTF can be rounded to a set that contains $O\left(\frac{\theta}{m}\right)$ elements. Second, we show in Theorem 25 that we can then use hashing to reduce the number of variables down to $\mathsf{poly}\left(\frac{\theta}{m}\right)$ before applying Lemma 21.

▶ **Lemma 22.** *Let $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i x_i\right)$ be a (θ, m)-LTF. Then there exists a set W with $|W| = O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$, and a margin $m' = \Theta(m)$ such that $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w'_i x_i\right)$, where each $w'_i \in W$ and f is a (θ, m')-LTF.*

The following result is also useful for our construction of a sketch for a (θ, m)-LTF.

▶ **Lemma 23** ([33]). *There is a randomized linear sketch with size $O(1)$ for the function*

$$
\mathsf{HAM}_{n,d|2d}(x) = \begin{cases} 1, & \text{if } ||x||_0 \le d \\ 0, & \text{if } ||x||_0 \ge 2d \end{cases}
$$

on instances $\{x | x \in \{0,1\}^n \text{ and } ||x||_0 \le d \text{ or } ||x||_0 \ge 2d\}$.

▶ **Fact 24.** *If $h : [n] \to [M]$ is a random hash function and $S \subseteq [n]$, then the probability that there exist $x, y \in S$ with $h(x) = h(y)$ is at most $\frac{|S|^2}{M}$.*

▶ **Theorem 25.** *Any (θ, m)-LTF admits a randomized linear sketch of size $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$.*

Proof. Let $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i x_i\right)$ be a (θ, m)-LTF. By Lemma 22, we can assume that $w_i \in W = \{2m(1 + \epsilon)^i\}_{i=0}^t$ so that the new margin $m' = \frac{4}{5} m$ and $t = \left\lceil \log_{1+\epsilon} \frac{\theta}{m} \right\rceil$ for $\epsilon = \frac{\theta}{10m}$. Recall from Lemma 21, $f(x) = 0$ only if $x_i = 1$ for at most $\frac{\theta}{2m}$ indices i of x. From Lemma 23, we can detect the instances where at least $\frac{\theta}{2m}$ indices i of x satisfy $x_i = 1$.

On the other hand, if less than $\frac{\theta}{2m}$ indices i of x satisfy $x_i = 1$, we can identify these indices and corresponding weights via hashing. Let $h : [n] \to [M]$, where $M = 5\left(\frac{\theta}{m}\right)^2$, and S be a set of indices of x, of size at most $\frac{\theta}{m}$. Then by Fact 24, the probability of a collision in h under elements of S is at most $\frac{1}{5}$. We partition $[n]$ into sets $S_{w,j}$ where $w \in W$ and $j \in [M]$ so that $S_{w,j} = \{i | h(i) = j \wedge w_i = w\}$. Therefore with probability at least $\frac{4}{5}$, there are no collisions in h under elements of S and $|S_{w,j}| \le 1$ for all $w \in W$ and $j \in [M]$.

Let $y_{w,j} = \sum_{i \in S_{w,j}} x_i$ and note that if there are no collisions in h under elements of S, then

$$
\sum_{i=1}^n w_i x_i = \sum_{(j,w) \in [M] \times W} w \left(\sum_{i \in S_{w,j}} x_i \right) = \sum_{(j,w) \in [M] \times W} w \cdot y_{w,j}.
$$

Thus, $f(x)$ is equivalent to the function $g(y) = \mathsf{sgn}\left(-\theta + \sum_{w,j} w \cdot y_{w,j}\right)$. Since $|W| = O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$, $M = 5\left(\frac{\theta}{m}\right)^2$ and $m' = \frac{4}{5} m$ is the margin for $g(y)$, then $g(y)$ depends on $O\left(\left(\frac{\theta}{m}\right)^3 \log \frac{\theta}{m}\right)$ variables $y_{w,j}$. By Lemma 21, there exists a randomized sketch for $g(y)$ of size $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$. ◀

We can also show that Theorem 25 is tight by recalling the function

$$
\mathsf{HAM}_{\le d}(x) = \begin{cases} 0, & \text{if } \sum_{i=1}^n x_i \le d + \frac{1}{2} \\ 1, & \text{otherwise.} \end{cases}
$$

Since this function is a $\left(d + \frac{1}{2}, \frac{1}{2}\right)$-LTF, it can be represented by a randomized linear sketch of size $O(d \log d)$. On the other hand, Dasgupta, Kumar and Sivakumar [19] notes that the one-way complexity of small set disjointness for two vectors x and y of weight d, which reduces to the function $\mathsf{HAM}_{\leq d}(x \oplus y)$, is $\Omega(d \log d)$. Thus, $\mathsf{HAM}_{\leq d}(x \oplus y)$ also requires a sketch of size $\Omega(d \log d)$.

3.1.2 Linear Threshold of Disjunctions

In this section, we describe a randomized linear sketch for functions that can be represented as 2-depth circuits where the top gate is a monotone linear threshold function with threshold θ and margin m, and the bottom gates are OR functions. Formally, if $g_S(x) = \bigvee_{i \in S} x_i$, q is a linear threshold function, and $w_S \geq 0$, then $f(x) = q(\ldots, g_S(x), \ldots) = \mathsf{sgn}\left(-\theta + \sum_{S \in 2^{[n]}} w_S \cdot g_S(x)\right)$.

▶ **Lemma 26.** *Let $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^{n} w_i x_i\right)$ be a (θ, m)-LTF where $w_i \in W$ for some set W. Let $h : [n] \to [M]$ be a random hash function where $M = \frac{50 \theta^2}{m^2}$ and*

$$f_h(x) = \mathsf{sgn}\left(-\theta + \sum_{(j,w) \in [M] \times W} w \left(\bigvee_{\substack{i:h(i)=j \\ w_i=w}} x_i\right)\right).$$

Then for all x, $\mathbf{Pr}\left[f_h(x) \neq f(x)\right] \leq \frac{1}{50}$.

▶ **Theorem 27.** *Let $g_S(x) = \bigvee_{i \in S} x_i$ with $w_S \geq 0$, q be a (θ, m)-LTF, and*

$$f(x) = q(\ldots, g_S(x), \ldots) = \mathsf{sgn}\left(-\theta + \sum_{S \in 2^{[n]}} w_S \cdot g_S(x)\right).$$

Then there is a randomized linear sketch for f of size $O\left(\left(\frac{\theta}{m}\right)^4 \log^2 \frac{\theta}{m}\right)$, where m is the margin of q.

Proof. We first apply Lemma 20 and Lemma 22 to q so that weights w_i can be rounded to elements of a set W with $|W| = O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$. For each $w_i \in W$, it again suffices to detect whether $\Theta(\frac{\theta}{m})$ disjunctions are nonzero. Hence to hash $O\left(\left(\frac{\theta}{m}\right)^2 \log \frac{\theta}{m}\right)$ disjunctions, it suffices to use a hash function with $M = O\left(\left(\frac{\theta}{m}\right)^4 \log^2 \frac{\theta}{m}\right)$ buckets. By Lemma 26, our resulting randomized linear sketch has size $O\left(\left(\frac{\theta}{m}\right)^4 \log^2 \frac{\theta}{m}\right)$. ◀

Proof of Theorem 14. Recall that $\mathsf{HAM}_{\leq d}(x)$ is a $\left(d + \frac{1}{2}, \frac{1}{2}\right)$-LTF. Furthermore, the set of weights W for $\mathsf{HAM}_{\leq d}(x)$ consists of a single element $\{1\}$, since the coefficient of each disjunction is one. Since $M = O(d^2 \log d)$, we can construct a randomized linear sketch with size $O(d^2 \log d)$ by Lemma 26. ◀

We note that our approach can be easily generalized to the case where the disjunction include the negations of some variables as well.

3.2   Communication Complexity of Lipschitz Submodular Functions

We discuss the communication complexity of Lipschitz submodular functions in this section. We first show in Section 3.2.1 that there exists an $\Omega(n)$-Lipschitz submodular function f that requires a randomized linear sketch of size $\Omega(n)$. We then show in Section 3.2.2 that in the one-way communication complexity model for XOR functions, there exists an $\Omega(n)$-Lipschitz submodular function f that has communication complexity $\Omega(n)$.

3.2.1   Approximate $\mathbb{F}_2$-Sketching of Lipschitz Submodular Functions

▶ **Theorem 28.** *There exist constants $c_1, c_2, \epsilon \geq 0$ and a monotone non-negative $(\frac{c_1}{n})$-Lipschitz submodular function f (a scaling of a matroid rank function) such that $\bar{R}_\epsilon^{lin}(f) \geq c_2 n$.*

Proof. Our proof uses a construction of a large family of matroid rank functions given in [10], Theorem 8. The construction uses the following notion of lossless bipartite expanders:

▶ **Definition 29** (Lossless bipartite expander). *Let $G = (U \cup V, E)$ be a bipartite graph. For $J \subseteq U$ let $\Gamma(J) = \{v | \exists u \in U : \{u, v\} \in E\}$. Graph G is a (D, L, ϵ)-lossless expander if:*

$$|\Gamma(\{u\})| = D \quad \forall u \in U$$
$$|\Gamma(J)| \geq (1 - \epsilon)D|J| \quad \forall J \subseteq U, |J| \leq L.$$

Here we need different parameters than in [10] so we restate their theorem as follows:

▶ **Theorem 30** ([10]). *Let $(U \cup V, E)$ be a (D, L, ϵ)-lossless expander with $|U| = k$ and $|V| = n$ and let $b = 8 \log k$. If $D \geq b$, $L = 4D/b - 2$ and $\epsilon = \frac{b}{4D}$ then there exists a family of sets $\mathcal{A} \subseteq 2^{[n]}$ and a family of matroids $\{M_\mathcal{B} : \mathcal{B} \subseteq \mathcal{A}\}$ with the following properties:*
- *$|\mathcal{A}| = k$ and for every $A \in \mathcal{A}$ it holds that $|A| = D$.*
- *For every $\mathcal{B} \subseteq \mathcal{A}$ and every $A \in \mathcal{A}$, we have:*

$$rank_{M_\mathcal{B}}(A) = \begin{cases} b & \text{if } A \in \mathcal{B} \\ D & \text{if } A \in \mathcal{A} \setminus \mathcal{B} \end{cases}$$

We use the following construction of lossless expanders from [54], see also [10].

▶ **Theorem 31** ([54]). *Let $k \geq 2$ and $\epsilon \geq 0$. For any $L \leq k$, let $D \geq 2 \log k / \epsilon$ and $n \geq 6DL/\epsilon$. Then a (D, L, ϵ)-lossless expander exists.*

In the above theorem we can set parameters as follows:

$$D = \frac{n}{3 \cdot 2^7}, \quad L = 2^3, \quad \epsilon = 2^{-3}, \quad k = 2^{n/3 \cdot 2^{11}}, \quad b = \frac{n}{3 \cdot 2^8}.$$

Note that under this choice of parameters we have $6DL/\epsilon = n$ and $\frac{2 \log k}{\epsilon} = D$ and hence a (D, L, ϵ)-lossless expander with parameters set above exists.

Now consider the family of matroids $\mathcal{M}$ given by Theorem 30 using the expander construction above. The rest of the proof uses the probabilistic method. We will show non-constructively that there exists a matroid in this family whose rank function does not admit a sketch of dimension $d = o(n)$. Let $\mathcal{D} = U(\mathcal{A})$ be the uniform distribution over $\mathcal{A}$. By Yao's principle it suffices to show that there exists a matroid rank function for which any deterministic sketch fails with a constant probability over this distribution. In the proof below we first show that any fixed deterministic sketch succeeds on a randomly chosen matroid from $\mathcal{M}$ with only a very tiny probability, probability $2^{2^{-\Omega(n)}}$, and then take a union bound over all 2^{dn} sketches of dimension at most d.

Indeed, fix any deterministic sketch $\mathcal{S}$ of dimension $d = n/2^{11}$. Let $\{b_1, \ldots, b_{2^d}\}$ be the set of all possible binary vectors of length d corresponding to the possible values of the sketch, so that each $b_i \in \{0,1\}^d$.

Let $S_{b_i} = \{A \in \mathcal{A} : \mathcal{S}(A) = b_i\}$. Let $t = \frac{1}{4}2^{n/2^{11}}$ and $G = \{b_i \in \{0,1\}^d | |S_{b_i}| \geq t\}$. The following proposition follows by a simple calculation.

▶ **Proposition 32.** *If $t = \frac{1}{4}2^{n/2^{11}}$ then $\frac{1}{k}\sum_{b_i \in G}|S_{b_i}| \geq \frac{3}{4}$.*

Proof. We have:

$$\frac{1}{k}\sum_{b_i \in G}|S_{b_i}| \geq 1 - \frac{1}{k}\sum_{b_i \,:\, |S_{b_i}| < \frac{k}{4 \cdot 2^d}}|S_{b_i}| \geq 1 - \frac{1}{k}\cdot\frac{k}{4 \cdot 2^d}\cdot 2^d \geq \frac{3}{4}. \qquad \blacktriangleleft$$

Let $S^1_{b_i} = \{A \in S_{b_i} : rank_{M_{\mathcal{B}}}(A) = b\}$ and $S^2_{b_i} = \{A \in S_{b_i} : rank_{M_{\mathcal{B}}}(A) = D\}$. We require the following lemma.

▶ **Lemma 33.** *Let $t = \frac{1}{4}2^{n/2^{11}}$ and $d = n/2^{11}$. There exists a matroid $M_{\mathcal{B}} \in \mathcal{M}$ such that for all deterministic sketches $\mathcal{S}$ of dimension d and all $b_i \in G$, $\min(|S^1_{b_i}|, |S^2_{b_i}|) \geq \frac{1}{4}|S_{b_i}|$.*

Fix the set $\mathcal{B}$ constructed in Lemma 33 and consider the function $rank_{M_{\mathcal{B}}}$. Consider distribution $\mathcal{D}$ over the inputs. The probability that any deterministic sketch over this distribution makes error at least $D - b$ is at least:

$$\frac{1}{k}\sum_{b_i \in \{0,1\}^n}\min(|S^1_{b_i}|, |S^2_{b_i}|) \geq \frac{1}{k}\sum_{b_i \in G}\min(|S^1_{b_i}|, |S^2_{b_i}|) \geq \frac{1}{k}\sum_{b_i \in G}\frac{1}{4}|S_{b_i}|,$$

where the last inequality holds by Lemma 33. Thus by Proposition 32, the probability is at least $\frac{3}{4} \times \frac{1}{4} \geq \frac{1}{6}$.

Finally, the construction of [10] ensures that the function $rank_{M_{\mathcal{B}}}$ takes integer values between 0 and D. Using this and the fact that matroid rank functions are 1-Lipschitz, we can normalize it by dividing all values by D and ensure that the resulting function is $O(1/n)$-Lipschitz and takes values in $[0,1]$, while the sketch makes error at least $(D-b)/D = \frac{1}{2}$. ◀

3.2.2 One-Way Communication of Lipschitz Submodular Functions

In this section, we strengthen the lower bound shown above, extending it to the corresponding one-way communication problem. We use the same notation as in the previous section.

▶ **Theorem 34.** *There exists a constant $c_1 > 0$ and a $\frac{c_1}{n}$-Lipschitz submodular function such that $R^{\rightarrow}_{1/3} = \Omega(n)$.*

By restricting the n-dimensional elements to r coordinates and observing that the construction outputs matroids of rank b or D that are separated by a constant gap, we obtain the following result using the same proof:

▶ **Corollary 35.** *There exists $c = \Omega(1)$ such that a c-approximation of matroid rank functions has randomized one-way communication complexity $R^{\rightarrow}_{1/3} = \Omega(r)$ where r is the rank of the underlying matroid.*

─── **References** ───

1 Dragan M Acketa. On the enumeration of matroids of rank-2. *Zbornik radova Prirodnomatematickog fakulteta–Univerzitet u Novom Sadu*, 8:83–90, 1978.

2 Yuqing Ai, Wei Hu, Yi Li, and David P. Woodruff. New Characterizations in Turnstile Streams with Applications. In *31st Conference on Computational Complexity, CCC*, pages 20:1–20:22, 2016.

3 Noga Alon, Yossi Matias, and Mario Szegedy. The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

4 Sepehr Assadi, Sanjeev Khanna, and Yang Li. Tight bounds for single-pass streaming complexity of the set cover problem. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 698–711, 2016.

5 László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication Complexity of Simultaneous Messages. *SIAM J. Comput.*, 33(1):137–166, 2003.

6 László Babai and Peter G. Kimmel. Randomized Simultaneous Messages: Solution of a Problem of Yao in Communication Complexity. In *Proceedings of the Twelfth Annual IEEE Conference on Computational Complexity*, pages 239–246, 1997.

7 Ashwinkumar Badanidiyuru, Shahar Dobzinski, Hu Fu, Robert Kleinberg, Noam Nisan, and Tim Roughgarden. Sketching valuation functions. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1025–1035, 2012.

8 Ashwinkumar Badanidiyuru, Baharan Mirzasoleiman, Amin Karbasi, and Andreas Krause. Streaming submodular maximization: massive data summarization on the fly. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, pages 671–680, 2014.

9 Maria-Florina Balcan, Florin Constantin, Satoru Iwata, and Lei Wang. Learning Valuation Functions. In *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, pages 4.1–4.24, 2012.

10 Maria-Florina Balcan and Nicholas J. A. Harvey. Learning submodular functions. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC*, pages 793–802, 2011.

11 MohammadHossein Bateni, Hossein Esfandiari, and Vahab S. Mirrokni. Almost Optimal Streaming Algorithms for Coverage Problems. In *Proceedings of the 29th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA*, pages 13–23, 2017.

12 Eric Blais and Abhinav Bommireddi. Testing Submodularity and Other Properties of Valuation Functions. In *8th Innovations in Theoretical Computer Science Conference, ITCS*, pages 33:1–33:17, 2017.

13 Eric Blais, Krzysztof Onak, Rocco Servedio, and Grigory Yaroslavtsev. Concise representations of discrete submodular functions, 2013.

14 Jehoshua Bruck and Roman Smolensky. Polynomial Threshold Functions, AC^0 Functions, and Spectral Norms. *SIAM J. Comput.*, 21(1):33–42, 1992.

15 Amit Chakrabarti and Anthony Wirth. Incidence Geometries and the Pass Complexity of Semi-Streaming Set Cover. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016*, pages 1365–1373, 2016.

16 Deeparnab Chakrabarty and Zhiyi Huang. Testing Coverage Functions. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP, Proceedings, Part I*, pages 170–181, 2012.

17 Chandra Chekuri, Shalmoli Gupta, and Kent Quanrud. Streaming Algorithms for Submodular Function Maximization. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP, Proceedings, Part I*, pages 318–330, 2015.

18 Mahdi Cheraghchi, Adam R. Klivans, Pravesh Kothari, and Homin K. Lee. Submodular functions are noise stable. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1586–1592, 2012.

19   Anirban Dasgupta, Ravi Kumar, and D. Sivakumar. Sparse and Lopsided Set Disjointness via Information Theory. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX, and 16th International Workshop, RANDOM. Proceedings*, pages 517–528, 2012.

20   Erik D. Demaine, Piotr Indyk, Sepideh Mahabadi, and Ali Vakilian. On Streaming and Communication Complexity of the Set Cover Problem. In *Distributed Computing - 28th International Symposium, DISC. Proceedings*, pages 484–498, 2014.

21   Yuval Emek and Adi Rosén. Semi-Streaming Set Cover. *ACM Trans. Algorithms*, 13(1):6:1–6:22, 2016.

22   Vitaly Feldman and Pravesh Kothari. Learning Coverage Functions and Private Release of Marginals. In *Proceedings of The 27th Conference on Learning Theory, COLT*, pages 679–702, 2014.

23   Vitaly Feldman, Pravesh Kothari, and Jan Vondrák. Representation, Approximation and Learning of Submodular Functions Using Low-rank Decision Trees. In *COLT 2013 - The 26th Annual Conference on Learning Theory*, pages 711–740, 2013.

24   Vitaly Feldman and Jan Vondrák. Tight Bounds on Low-Degree Spectral Concentration of Submodular and XOS Functions. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*, pages 923–942, 2015.

25   Vitaly Feldman and Jan Vondrák. Optimal Bounds on Approximation of Submodular and XOS Functions by Juntas. *SIAM J. Comput.*, 45(3):1129–1170, 2016.

26   Gereon Frahling, Piotr Indyk, and Christian Sohler. Sampling in Dynamic Data Streams and Applications. *Int. J. Comput. Geometry Appl.*, 18(1/2):3–28, 2008.

27   Michel X. Goemans, Nicholas J. A. Harvey, Satoru Iwata, and Vahab S. Mirrokni. Approximating submodular functions everywhere. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 535–544, 2009.

28   Vince Grolmusz. On the Power of Circuits with Gates of Low L_1 Norms. *Theor. Comput. Sci.*, 188(1-2):117–128, 1997.

29   Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan Ullman. Privately Releasing Conjunctions and the Statistical Query Barrier. *SIAM J. Comput.*, 42(4):1494–1520, 2013.

30   Sariel Har-Peled, Piotr Indyk, Sepideh Mahabadi, and Ali Vakilian. Towards Tight Bounds for the Streaming Set Cover Problem. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS*, pages 371–383, 2016.

31   Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of Protocols for XOR Functions. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*, pages 282–288, 2016.

32   Kaave Hosseini, Shachar Lovett, and Grigory Yaroslavtsev. Optimality of Linear Sketching under Modular Updates. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:169, 2018. URL: https://eccc.weizmann.ac.il/report/2018/169.

33   Wei Huang, Yaoyun Shi, Shengyu Zhang, and Yufan Zhu. The communication complexity of the Hamming distance problem. *Inf. Process. Lett.*, 99(4):149–153, 2006.

34   William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability*, pages 189–206, 1984.

35   Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear Sketching over F_2. In *33rd Computational Complexity Conference, CCC*, pages 8:1–8:37, 2018.

36   Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. Single Pass Spectral Sparsification in Dynamic Streams. *SIAM J. Comput.*, 46(1):456–477, 2017.

37   Michael Kapralov, Jelani Nelson, Jakub Pachocki, Zhengyu Wang, David P. Woodruff, and Mobin Yahyazadeh. Optimal Lower Bounds for Universal Relation, and for Samplers and Finding Duplicates in Streams. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 475–486, 2017.

38   Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, 1997.

39   Troy Lee and Shengyu Zhang. Composition Theorems in Communication Complexity. In *Automata, Languages and Programming, 37th International Colloquium, ICALP, Proceedings, Part I*, pages 475–489, 2010.

40   Ming Lam Leung, Yang Li, and Shengyu Zhang. Tight bounds on the randomized communication complexity of symmetric XOR functions in one-way and SMP models. *CoRR*, abs/1101.4555, 2011. `arXiv:1101.4555`.

41   Yi Li, Huy L. Nguyen, and David P. Woodruff. Turnstile streaming algorithms might as well be linear sketches. In *Symposium on Theory of Computing, STOC*, pages 174–183, 2014.

42   Yang Liu and Shengyu Zhang. Quantum and randomized communication complexity of XOR functions in the SMP model. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:10, 2013.

43   Shachar Lovett. Recent Advances on the Log-Rank Conjecture in Communication Complexity. *Bulletin of the EATCS*, 112, 2014.

44   Andrew McGregor. Graph stream algorithms: a survey. *SIGMOD Record*, 43(1):9–20, 2014.

45   Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *CoRR*, abs/0909.3392, 2009. `arXiv:0909.3392`.

46   Ryan O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.

47   Sofya Raskhodnikova and Grigory Yaroslavtsev. Learning pseudo-Boolean k-DNF and submodular functions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1356–1368, 2013.

48   Barna Saha and Lise Getoor. On Maximum Coverage in the Streaming Model & Application to Multi-topic Blog-Watch. In *Proceedings of the SIAM International Conference on Data Mining, SDM*, pages 697–708, 2009.

49   C. Seshadhri and Jan Vondrák. Is Submodularity Testable? *Algorithmica*, 69(1):1–25, 2014.

50   Yaoyun Shi and Zhiqiang Zhang. Communication complexities of symmetric XOR functions. *Quantum Inf. Comput*, pages 0808–1762, 2008.

51   Xiaoming Sun and Chengu Wang. Randomized Communication Complexity for Linear Algebra Problems over Finite Fields. In *29th International Symposium on Theoretical Aspects of Computer Science, STACS*, pages 477–488, 2012.

52   Justin Thaler. Semi-Streaming Algorithms for Annotated Graph Streams. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 59:1–59:14, 2016.

53   Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier Sparsity, Spectral Norm, and the Log-Rank Conjecture. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 658–667, 2013.

54   Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

55   Jan Vondrák. A note on concentration of submodular functions. *CoRR*, abs/1005.2791, 2010. `arXiv:1005.2791`.

56   David P. Woodruff. Sketching as a Tool for Numerical Linear Algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014.

57   Grigory Yaroslavtsev and Samson Zhou. Approximate F_2-Sketching of Valuation Functions. *CoRR*, abs/1907.00524, 2019. `arXiv:1907.00524`.

58   Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of Boolean functions. *Theor. Comput. Sci.*, 411(26-28):2612–2618, 2010.

A   Missing Proofs

Proof of Proposition 10. Indeed, let A_d be a d-dimensional subspace such that $\sum_{S \in A_d} \hat{f}^2(S) \geq \|f\|_2^2 - \epsilon$ and consider the function $g(x) = \sum_{S \in A_d} \hat{f}(S) \chi_S(x)$. Note that in order to compute all values $\chi_S(x)$ for $S \in A_d$ it suffices to evaluate d parities corresponding to sets $S_1, \ldots, S_d$ forming a basis in A_d. Values of all other parities can be computed as linear combinations. Let $\Delta(x) = f(x) - g(x)$. Then the desired guarantee follows from the following calculation:

$$\mathop{\mathbb{E}}_{x \sim U(\{0,1\}^n)} [\Delta(x)^2] = \mathop{\mathbb{E}}_{S \sim U(\{0,1\}^n)} [\hat{\Delta}(S)^2] = \sum_{S \in \{0,1\}^n} (\hat{f}(S) - \hat{g}(S))^2 = \sum_{S \notin A_d} \hat{f}(S)^2 \leq \epsilon,$$

where the first equality holds from Parseval's identity. ◀

Proof of Lemma 20. We show the stronger result that for any j such that $w_j < 2m$, then $f(x) = f(x \oplus e_j)$, where e_j is the elementary unit vector with one in the jþ position, and zeros elsewhere. This implies the lemma since it shows that any variable whose weight is less than $2m$ does not affect the output of the function or the margin of the function and thus might as well have weight zero.

Suppose, by way of contradiction, that $f(x) \neq f(x \oplus e_j)$ and without loss of generality, $f(x) = 0$ with $x_j = 0$. Since f is a linear threshold function and $f(x) = 0$, then $-\theta + \sum_{i=1}^n w_i x_i < 0$. Moreover, f is a (θ, m)-LTF, so $-\theta + \sum_{i=1}^n w_i x_i < -m$. Because $w_j < 2m$, $-\theta + w_j + \sum_{i=1}^n w_i x_i < -\theta + 2m + \sum_{i=1}^n w_i x_i < m$. But because m is the margin of the function, if $-\theta + w_j + \sum_{i=1}^n w_i x_i < m$, then it must hold that $-\theta + w_j + \sum_{i=1}^n w_i x_i < -m$. Therefore, $f(x \oplus e_j) = 0$, so x_j does not affect the output of the function or the margin of the function. ◀

Proof of Lemma 22. Observe that for any $w_i \geq 2\theta$, if $x_i = 1$, then $f(x) = 1$. Thus, if $f(x) = 1$, it suffices to consider $2m \leq w_i \leq 2\theta$.

Let $W = \{2m(1 + \epsilon)^i\}_{i=0}^t$ for $t = \lceil \log_{1+\epsilon} \left(\frac{\theta}{m}\right) \rceil$, where ϵ is some fixed constant that we set at a later time. For each i, let w_i' be the largest element in W that does not exceed w_i. Thus, $w_i' \leq w_i < (1 + \epsilon)w_i'$. Observe that since $w_i' \leq w_i$ and f is a (θ, m)-LTF, then $f(x) = 0$ implies $-m > -\theta + \sum_{i=1}^n w_i x_i \geq -\theta + \sum_{i=1}^n w_i' x_i$, so that $\mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i' x_i\right) = 0 = f(x)$ and a margin of m remains.

On the other hand, if $f(x) = 1$, then $\sum_{i=1}^n w_i x_i > \theta + m$ as f is a (θ, m)-LTF. Since $w_i' \leq w_i < (1 + \epsilon)w_i'$, then $\sum_{i=1}^n w_i' x_i > \frac{\theta + m}{1+\epsilon} > (1 - \epsilon)(\theta + m)$. Observe that $\theta \geq m$ and hence, $\sum_{i=1}^n w_i' x_i > \theta - \epsilon\theta + m - \epsilon m \geq \theta + m - 2\epsilon\theta$. Setting $\epsilon = \frac{\theta}{10m}$ shows that $\mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i' x_i\right) = 1 = f(x)$ and a margin of $m' = \frac{4}{5}m$ remains. ◀

Proof of Lemma 26. As by Lemma 22, we can assume without loss of generality that $w_i \geq 2m$ and $w \geq 2m$. Let $S = \{i | x_i = 1\}$ so that if there are no collisions under h in S, then

$$\sum_{(j,w) \in [M] \times W} w \left(\bigvee_{\substack{i:h(i)=j \\ w_i=w}} x_i \right) = \sum_i w_i x_i.$$

If $f(x) = 0$, then $|S| \leq \frac{\theta}{2m}$ so that the probability there are collisions under h in S is at most $\frac{1}{200}$ by Fact 24. Thus if $f(x) = 0$, then $f_h(x) = 0$ with probability at least $1 - \frac{1}{200}$.

If $f(x) = 1$, then either $|S| < \frac{\theta}{m}$ or $|S| \geq \frac{\theta}{m}$. If $|S| < \frac{\theta}{m}$, then the probability there are collisions under h in S is at most $\frac{1}{50}$ by Fact 24, so then $f_h(x) = 1$ with probability at least $1 - \frac{1}{50}$. If $|S| \geq \frac{\theta}{m}$, with probability at least $1 - \frac{1}{50}$, there exist $\frac{\theta}{m}$ values j such that there exists $x_i = 1$ and $h(i) = j$. Therefore, we set $f_h(x) = 1$ whenever at least $\frac{\theta}{m}$ buckets of h are non-empty. In all cases, $f_h(x) = f(x)$ with probability at least $1 - \frac{1}{50}$. ◀

Proof of Lemma 33. The proof uses the probabilistic method to show the existence of $\mathcal{B}$ with desired properties. Consider drawing a random matroid from the family $\mathcal{M}$, i.e. pick $\mathcal{B}$ to be a uniformly random subset of $\mathcal{A}$ and consider $M_{\mathcal{B}}$. Fix any deterministic sketch $\mathcal{S}$ and any $b_i \in G$. Since $|S_{b_i}| \geq t$, by the Chernoff bound, it holds that:

$$\Pr_{\mathcal{B} \subseteq \mathcal{A}} \left[|S_{b_i}^1| > \left(\frac{1}{2} + \delta \right) |S_{b_i}| \right] \leq e^{-c\delta^2 |S_{b_i}|} \leq e^{-c\delta^2 t}.$$

Setting $\delta = 1/4$, we have that the above probability is at most e^{-Ct} for some constant $C > 0$. Applying the argument above to both $S_{b_i}^1$ and $S_{b_i}^2$, we have that:

$$\Pr_{\mathcal{B} \subseteq \mathcal{A}} \left[\min(|S_{b_i}^1|, |S_{b_i}^2|) < \frac{1}{4} |S_{b_i}| \right] \leq 2e^{-Ct}.$$

Let $\mathcal{E}$ denote the event that $\min(|S_{b_i}^1|, |S_{b_i}^2|) \geq \frac{1}{4} |S_{b_i}|$.

Note that the total number of deterministic sketches of dimension d is at most 2^{dn}, since each sketch is specified by a collection of d linear functions over $\mathbb{F}_2^n$. Also note that for each sketch $|G| \leq 2^d$. Taking a union bound over all sketches and all sets G by the choice of t and d event $\mathcal{E}$ holds for all $\mathcal{S}$ and $b_i \in G$ with probability at least $1 - 2^{(n+1)d+1} e^{-Ct} \geq 1 - 2^{(n+1)d+1} 2^{-\frac{C}{4} 2^{n/2^{11}}} = 1 - o(1)$. Thus, there exists some set $\mathcal{B}$ for which the statement of the lemma holds. ◀

Proof of Theorem 34. Let $\alpha = \frac{1}{3 \cdot 2^{11}}$ and $|\mathcal{A}| = k = 2^{\alpha n}$. Suppose Alice holds $x \in \mathcal{A} \subseteq \{0,1\}^n$ and Bob holds $y \in \{0,1\}^n$. Recall that in the one-way communication model for XOR functions, Alice must pass a message of minimal length to Bob, who must then compute $f(x \oplus y)$ with some probability, say $\frac{2}{3}$. Here, we let specifically let f be a scaling of a matroid rank function, which is some monotone non-negative $\left(\frac{c_1}{n} \right)$-Lipschitz submodular function. By Yao's principle, it suffices to show that every deterministic one-way communication protocol using at most $\frac{\alpha}{4} n$ bits fails with probability greater than $\frac{1}{3}$ over $\mathcal{A}$. Suppose by way of contradiction, that Alice and Bob succeed through a deterministic one-way communication protocol, using at most $\frac{\alpha}{4} n$ bits. For the purpose of analysis, we furthermore suppose that Bob's input is fixed.

We now claim that if Alice passes a message to Bob using at most $\frac{\alpha}{4} n$ bits, then there are at least $2^{\alpha n} - 4 \cdot 2^{\alpha n/4}$ points in $\mathcal{A}$ that are represented by the same message as at least five other points. Note that Alice can partition the input space $\mathcal{A}$ into at most $2^{\alpha n/4}$ parts, each part with its own distinct representative message. The number of points *not* in parts containing at least five other points is at most $4 \cdot 2^{\alpha n/4}$. The remaining points, at least $2^{\alpha n} - 4 \cdot 2^{\alpha n/4}$ in quantity, are represented by the same message as at least five other points.

Let S be the set of points in $\mathcal{A}$ represented by a given message from Alice. Hence, Alice assigns the same message to each of these points and passes the state of the protocol to Bob. Because Bob cannot distinguish between these points and must perform a deterministic protocol, then Bob must output the same result for each of these points. Recall that we consider Bob's input $y \in \{0,1\}^n$ as fixed. Consider the family of functions

$$\mathcal{F} = \{ f : f(x \oplus y) = b \text{ or } f(x \oplus y) = D \text{ for all } x \in \mathcal{A} \}.$$

Thus, if S contains at least five points, there exists $f \in \mathcal{F}$ such that Bob errs on at least $\frac{2}{5}$ fraction of the points in S by setting $f(x \oplus y) = b$ to at least $\left\lfloor \frac{|S|-1}{2} \right\rfloor$ of the points $x \in S$ and similarly for $f(x \oplus y) = D$. Moreover, since Alice partitions the points in $\mathcal{A}$, then there exists an $f \in \mathcal{F}$ such that Bob errs on at least $\frac{2}{5}$ fraction on *all* points that are represented by the same message as at least five other points. Hence, the total number of inputs that Bob errs is at least $\frac{2}{5} \left(2^{\alpha n} - 6 \cdot 2^{\alpha n/4} \right) > \frac{1}{3} \cdot 2^{\alpha n}$ for sufficiently large values of n. This contradicts the assumption that the communication protocol, using at most $\frac{\alpha}{4} n$ bits, succeeds with probability $\frac{2}{3}$. $\blacktriangleleft$