Contents

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

Grigory Yaroslavtsev[*]   Samson Zhou[†]

September 6, 2018

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 [Woo14]), graph sparsification (see survey [McG14]), frequency estimation [AMS99], dimensionality reduction [JL84], various forms of sampling, signal processing, and communication complexity. In fact, linear sketching has been shown to be the optimal algorithmic technique [LNW14, AHLW16] 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 $\mathbb{F}_2$ ($\mathbb{F}_2$-sketching). This is a previously unstudied but natural generalization of the work of [KMSY18], 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 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 computed over these sets, denoted as $\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)$. While looking only at linear functions over $\mathbb{F}_2$ as candidate sketches might seem restrictive, we are unaware of any problem for which sketching over reals gives any advantage over $\mathbb{F}_2$ if the domain of the function is $\mathbb{F}_2^n$. Furthermore, as shown recently in [KNP+17], almost all

---

[*]Indiana University, Bloomington, IN., USA. `gyarosla@iu.edu`
[†]Indiana University, Bloomington, IN., USA. `samsonzhou@gmail.com`

dynamic graph streaming algorithms[1] can be based on versions of the ℓ_0-sampling primitive without losing optimality in space which in turn can be implemented optimally as $\mathbb{F}_2$-sketches. Examples of such algorithms include dynamic streaming algorithms for graph connectivity, k-connectivity, bipartiteness, minimum spanning tree [AGM12a], subgraph counting, min-cut, cut sparsifiers, spanners, spectral sparsifiers [AGM12b], maximal matching [CCHM15], maximum matching [Kon15, AKLY16, AKL17, CCE+16], densest subgraph [BHNT15, MTVV15, EHW16], vertex and hyperedge connectivity [GMT15] and graph degeneracy [FT16].

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 1.1 (Exact $\mathbb{F}_2$-sketching, [KMSY18]). *The* exact randomized $\mathbb{F}_2$-sketch complexity *with error δ of a function $f: \mathbb{F}_2^n \to \mathbb{R}$ (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 1.2 (Approximate $\mathbb{F}_2$-sketching). *The ϵ-approximate randomized $\mathbb{F}_2$-sketch complexity of a function $f: \mathbb{F}_2^n \to \mathbb{R}$ (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, we show that functions with small spectral norm (e.g. coverage functions, Corollary A.4) admit such approximate $\mathbb{F}_2$-sketches.

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 1.3 (Approximate distributional $\mathbb{F}_2$-sketching). *For a function $f: \mathbb{F}_2^n \to \mathbb{R}$, 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 $\bar{\mathcal{D}}_\epsilon^{lin,\mathsf{D}}(f)$) 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:*

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

---

[1]With the only exception being the work of [KLM+17] on spectral graph sparsification.

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 [Tha16][2]. 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.

It is known that linear sketching over the reals ([LNW14, AHLW16]) as well as $\mathbb{F}_2$-sketching ([KMSY18]) give (almost) optimal space complexity for processing dynamic data streams in the respective update models for *any* function f. Hence, both upper and lower bounds on $\mathbb{F}_2$-sketch complexity obtained in our work extend to space complexity of dynamic streaming algorithms. However, all existing general reductions between sketching and streaming require adversarial streams of length triply exponential in n. We thus complement our lower bounds on dimension of $\mathbb{F}_2$-sketches with one-way communication complexity lower bounds for the corresponding XOR functions. 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 results can be used for distributed algorithms computing $f(x_1 + \ldots x_M)$ over a collection of distributed inputs $x_1, \ldots, x_M$ as $\mathbb{F}_2$-sketches can be used for distributed inputs. On the other hand, our communication lower bounds also hold in the simultaneous message passing (SMP) communication model, which is strictly harder than one-way communication.

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 [BH10, BCIW12, CKKL12, GHRU13, RY13, FKV13, FK14, FV15, FV16][3], approximation [GHIM09, BDF+12] and property testing [CH12, SV14, BB17]. 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. [SG09, DIMV14, BMKK14, CGQ15, CW16, ER16, HIMV16, AKL16, BEM17], 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 [KMSY18]. 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 [BS92] (see also [Gro97, MO09]), turns out to be

---

[2]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 knowledge of the sign of the update (i.e. whether it is +1 or -1) does not help in general. 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 [KLM+17]) are based on the ℓ_0-sampling primitive [FIS08]. As shown recently, ℓ_0-sampling can be implemented optimally using $\mathbb{F}_2$-sketches [KNP+17] 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.

[3]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 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.

Class	Error	Distribution	Complexity	Result
Additive/Budget additive $\min(b, \sum_{i=1}^n w_i x_i)$	ϵ	any	$\Theta\left(\frac{\|w\|_1^2}{\epsilon}\right)$	Theorem A.7, 4.1 Corollary A.3, A.6
$\min(c\sqrt{n}, \frac{2c}{\sqrt{n}} \sum_{i=1}^n x_i)$	constant	uniform	$\Omega(n)$	Theorem 4.1
Coverage	ϵ	any	$O\left(\frac{1}{\epsilon}\right)$	Corollary A.4
Matroid Rank 2 Graphic Matroids	exact	any	$\Theta(1)$	Theorem 3.1 Theorem 3.5
Matroid Rank r	exact	any	$\Omega(r)$	Corollary 3.24
Matroid Rank r	exact	uniform	$O((r \log r + c)^{r+1})$	Corollary D.6
Matroid Rank	$1/\sqrt{n}$	uniform	$\Theta(1)$	Corollary D.8
$\frac{c}{n}$-Lipschitz Submodular	constant	any	$\Theta(n)$	Theorem 3.17

Table 1: Linear sketching complexity of classes of valuation functions

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 particular, our hardness result for Lipschitz submodular functions uses a construction of a large family of matroids from [BH10] (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[4].

- **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 $O(\min(\|w\|_1^2/\epsilon, n))$ (Corollary A.3) and give a matching communication lower bound (Theorem A.7) for all $\epsilon \geq \|w\|_2^2$.

- **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 $O(\min(\|w\|_1^2/\epsilon, n))$ (Corollary A.6). We give a matching communication bound

---

[4]We don't discuss some other subclasses of submodular 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 Appendix C).

for the "hockey stick" function for constant ϵ (Theorem 4.1) which holds even under the uniform distribution of the input.

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

  Our results: We show a simple upper bound of $O(1/\epsilon)$ (Corollary A.4) for such functions.

- **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 1.4 (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 3.1) and graphic matroids (Theorem 3.5).*
  - *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 3.23, Corollary 3.24).*

  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 (Corollary D.6, follows from the junta approximation of [BOSY13]). Furthermore, matroids of high rank $\Omega(n)$ can be trivially approximately sketched under product distributions, due to their concentration around their expectation (see Appendix D 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 3.17). 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. [Von10, BH10]) and hence can be approximated using their expectation without any sketching at all.

5

1.4 Overview and Techniques

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 of the results of [KMSY18] regarding Boolean functions 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 [KMSY18]. The key structural results of [KMSY18] 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 2.6 and Theorem 2.7. 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 Appendix A.1. In particular, as we show in Appendix A.1.1, analogously to the Boolean case, the ϵ-approximate randomized Fourier sketching complexity of a function 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 (Proposition A.5). 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 can't be substantially improved in the real-valued case by presenting a subclass of linear functions which require sketches of size $\Omega(\|\hat{f}\|_1^2/\epsilon)$ (Theorem A.7). This is in contrast with the case of Boolean functions studied in [KMSY18] for which such tightness result is not known.

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 [MO09]. Recall that a linear threshold function (LTF) can be represented as $f(x) = \mathsf{sgn}\left(\sum_{i=1}^{n} w_i x_i - \theta\right)$ for some weights w_i and threshold θ, where sgn is the sign function. An important parameter of an LTF is its *margin* which corresponds to the difference between the threshold and the value of the LTF 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 to a collection of weights without altering the output of the function and moreover, the collection is bounded in size, 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 [BH10] 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 $\Omega(n)$.

Uniform Distribution   In Section 4, 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 Section 4.1. 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 Section D 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 [Von10] (see Section D).

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 [SZ08, ZS10, MO09, LZ10, LLZ11, SW12, LZ13, TWXZ13, Lov14, HHL16, KMSY18] for related communication complexity results).

Definition 2.1 (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 $R_\delta^\rightarrow(f^+)$) of its XOR-function is defined as the smallest size[5] (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 $R_\delta^\rightarrow(f^+) \le R_\delta^{lin}(f)$ 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 2.2 (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*

---

[5]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 [KN97] for a formal definition.

(denoted as $\bar{R}_\epsilon^\rightarrow(f^+)$) 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 [BK97, BGKL03], 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 2.3 (Simultaneous communication complexity of XOR function). *For a function $f \colon \mathbb{F}_2^n \to \mathbb{R}$ the simulatenous 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.

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 [O'D14] and Appendix A.1. In particular, Fourier concentration on a low-dimensional subspace implies existence of a small sketch which satisfies this guarantee:

Definition 2.4 (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 [KMSY18], adapted for the case of real-valued functions.

Definition 2.5 (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 2.6. *For any $f \colon \mathbb{F}_2^n \to \mathbb{R}$ it holds that:*

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

Proof. 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 and the following calculation:

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

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 Appendix B as it is follows closely an analogous result for Boolean functions from [KMSY18].

Theorem 2.7. *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 [Ack78] 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 threshold Hamming weight function 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 [MO09].

3.1 Matroids of Rank 2

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

Theorem 3.1. *For every matroid M of rank 2 it holds that $R_{\frac{1}{3}}^{lin}(rank_M) = O(1)$.*

It is well-known that matroids of rank 2 admit the following characterization (see e.g. [Ack78]).

Fact 3.2. *The collection of independent sets of size 2 of a rank 2 matroid can be represented as edges in a complete graph with edges corresponding to some disjoint union of cliques removed.*

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

Using Fact 3.2, Theorem 3.3, and the following fact that upper bounds the sketch complexity for functions with small support, we prove Theorem 3.1 by writing the matroid rank function for M as a linear threshold function of disjunctions.

Fact 3.4 (Folklore, see e.g. [MO09, KMSY18])**.** *For any function $f : \{0, 1\}^n \to \{0, 1\}$ with $\min_{z \in \{0,1\}} \mathbf{Pr}_x(f(x) = z) \leq \epsilon$ it holds that $R_\delta^{lin}(f) \leq \log \frac{2^{n+1}\epsilon}{\delta}$.*

Proof of Theorem 3.1: $\mathbb{F}_2$-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 doesn't depend on the corresponding input. Hence $rank_M(x) = 0$ if and only if $x = 0^n$. Thus $R_\delta^{lin}(rank_M) = R_\delta^{lin}(f_M) + O(\log 1/\delta)$ as by Fact 3.4

9

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 3.2 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 $S_1, \ldots, S_t$ be the collection of vertex sets of disjoint cliques defining a rank 2 matroid M in Fact 3.2. Without loss of generality, we can assume that $|\cup_{i=1}^t S_i| = n$ by adding singletons. Then:

$$f_M(x) = Ham_{\geq 2} \left(\bigvee_{j \in S_1} x_j, \bigvee_{j \in S_2} x_j, \ldots, \bigvee_{j \in S_t} x_j \right),$$

where $Ham_{\geq 2}(z_1, \ldots, z_t) = 1$ if and only if $\sum_{i=1}^t z_i \geq 2$ is the threshold Hamming weight function. By Theorem 3.3, the sketch complexity of $f_M(x)$ is $O(1)$, since the Hamming weight threshold is $d = 2$.

$\square$

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

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

where each S_i is a separate spanning forest. Therefore, Theorem 3.3 yields a $O(r^2 \log r)$ space linear sketch for graphic matroids of rank r.

Theorem 3.5. *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 3.3, while resolving an open question by Montanaro and Osborne [MO09].

3.1.1 Linear Threshold Functions

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

Definition 3.6. *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 $\mathsf{sgn}(x) = 0$ for $x < 0$ and $\mathsf{sgn}(x) = 1$ for $x \geq 0$ is the Heaviside step function.*

Definition 3.7. *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 $O\left(\frac{\theta}{m} \log n\right)$ [LZ13], 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 3.8 ([MO09]). *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 3.8 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 the margin $2m$ in evaluating a (θ, m)-LTF.

Lemma 3.9. *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' if $w_i \geq 2m$ and $w_i' = 0$ otherwise. Then $f(x) = \mathsf{sgn}\left(-\theta + \sum_{i=1}^n w_i' x_i\right)$.*

10

Proof. 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 jth 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, $f(x) = -1$ with $x_j = -1$. Since f is a linear threshold function and $f(x) = -1$, then $-\theta + \sum_{i=1}^{n} w_i x_i < 0$. Moreover, f is a $(\theta, 2m)$-LTF, so $-\theta + \sum_{i=1}^{n} w_i x_i < -m$. Because $w_j < m$, $-\theta + w_j + \sum_{i=1}^{n} w_i x_i < -\theta + 2m + \sum_{i=1}^{n} w_i x_i < m$. 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. ∎

As noted, Lemma 3.9 implies that we can ignore not only variables with zero weights, but all variables whose weights are less than $2m$. We now apply Fact 3.4 by showing that the support of the set $\{x \mid f(x) = -1\}$, where f is a (θ, m)-LTF.

Lemma 3.10. *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 3.9, 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 $f(x) = -1$, 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 $f(x) = -1$. Thus, the number of $x \in \{0,1\}^n$ with $f(x) = -1$ is at most $\sum_{0 \leq i \leq \theta/2m} \binom{n}{i} \leq (n+1)^{\lceil \theta/2m \rceil}$. Applying Fact 3.4, there exists a randomized linear sketch for f, of size $O\left(\frac{\theta}{m} \log n\right)$. ∎

In order to fully prove Question 3.8 and obtain a dependence on $\log \frac{\theta}{m}$ rather than $\log n$, we use the following two observations. First, we show in Lemma 3.11 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 3.14 that we can then use hashing to reduce the number of variables down to $\mathsf{poly}\left(\frac{\theta}{m}\right)$ before applying Lemma 3.10.

Lemma 3.11. *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 $m' = \Omega(m)$ 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.*

Proof. 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 $t = \left\lceil \log_{1+\epsilon}\left(\frac{\theta}{m}\right)\right\rceil$, 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 $\frac{4}{5}m$ remains. ∎

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

Lemma 3.12. *[HSZZ06] There is a randomized linear sketch with size $O(1)$ for the function*

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

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

Fact 3.13. *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 $\frac{|S|^2}{2M}$.*

Theorem 3.14. *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 3.11, 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 3.10, $f(x) = 0$ only if $x_i = 1$ for at most $\frac{\theta}{2m}$ indices i of x. From Lemma 3.12, 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 3.13, 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 $|S_{w,j}| \leq 1$ 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

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

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 3.10, 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 3.14 is tight by considering the following function. Let

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

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 [DKS12] notes that the one-way complexity of small set disjointness for two vectors x and y of weight d is $\Omega(d \log d)$, which reduces to the function $\mathsf{HAM}_{\leq d}(x \oplus y)$. 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 $g_S(x) = \bigvee_{i \in S} x_i$ 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 3.15. *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*

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

12

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

Proof. As by Lemma 3.11, 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 3.13. Thus if $f(x) = 0$, then $f_h(x) = 0$ with probability at least $1 - \frac{1}{200}$.

If $f(x) = 1$, then either $|S| \leq \frac{\theta}{l^n}$ or $|S| \geq \frac{\theta}{m}$. If $|S| \leq \frac{\theta}{m}$, then the probability there are collisions under h in S is at most $\frac{1}{50}$ by Fact 3.13, 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 $1 - \frac{1}{50}$. ∎

Theorem 3.16. *Let $g_S(x) = \bigvee_{i\in S} x_i$ with $w_S \geq 0$ 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 $O\left(\left(\frac{\theta}{m}\right)^3 \log^2 \frac{\theta}{m} \right)$, where m is the margin of q.

Proof. We first apply Lemma 3.9 and Lemma 3.11 to q so that weights w_i can be rounded to elements of a set W with $|W| = O\left(\frac{\theta}{m}\log\frac{\theta}{m}\right)$. By Lemma 3.15 and noting that $M = O\left(\left(\frac{\theta}{m}\right)^2\right)$, we can construct a randomized linear sketch with size $O\left(\left(\frac{\theta}{m}\right)^3 \log^2 \frac{\theta}{m} \right)$ by requiring that each sketch in M fails with probability $\frac{1}{50W}$ to ensure that the sketch errs with probability at most $\frac{1}{50}$ on each input. ∎

Proof of Theorem 3.3: 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 3.15. □

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 3.17. *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 [BH10], Theorem 8. The construction uses the following notion of lossless bipartite expanders:

Definition 3.18 (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 [BH10] so we restate their theorem as follows:

Theorem 3.19 ([BH10]). *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 [Vad12], see also [BH10].

Theorem 3.20 ([Vad12]). *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 3.19 using the expander construction above. The rest of the proof uses 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 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 2nd sketches of dimension at most d.

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

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

14

Proof. We have:

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

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

Lemma 3.22. *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* $b_i \in G$:

$$\min(|S^1_{b_i}|, |S^2_{b_i}|) \geq \frac{1}{4}|S_{b_i}|.$$

Proof. The proof uses probabilistic method to show 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:

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

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 $S^1_{b_i}$ and $S^2_{b_i}$ we have that:

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

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

Note that the total number of deterministic sketches of dimension d is at most 2nd 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. $\quad \blacksquare$

Fix the set $\mathcal{B}$ constructed in Lemma 3.22 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:

$$\begin{aligned}
\frac{1}{k}\sum_{b_i \in \{0,1\}^n}\min(|S^1_{b_i}|, |S^2_{b_i}|) &\geq \frac{1}{k}\sum_{b_i \in G}\min(|S^1_{b_i}|, |S^2_{b_i}|) && \\
&\geq \frac{1}{k}\sum_{b_i \in G}\frac{1}{4}|S_{b_i}| && \text{(by Lemma 3.22)} \\
&\geq \frac{3}{4} \times \frac{1}{4} \geq \frac{1}{6}. && \text{(by Proposition 3.21)}
\end{aligned}$$

Finally, the construction of [BH10] 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 $(D-b)/D = \frac{1}{2}$. $\quad \blacksquare$

15

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

Proof. 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 $\frac{1}{3}$. 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 $\frac{2}{3}$. ∎

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

4 Communication Under the Uniform Distribution

In this section we switch to lower bounds for the uniform distribution and show the following result for the "hockey stick" function:

Theorem 4.1. *For any odd n, constant $c > 0$ and $\alpha = c\sqrt{n}$ there exists a constant $\epsilon > 0$ such that for the "hockey stick" function $hs_\alpha(x) = \min(\alpha, \frac{2\alpha}{n}\sum_{i=1}^{n} x_i)$ it holds that:*

$$\bar{\mathcal{D}}_\epsilon^{\rightarrow, U}(hs_\alpha^+) = \Omega(n)$$

The proof relies on the characterization of communication complexity using approximate Fourier dimension (Theorem 2.7). We also require a structural result, whose proof we defer to Section 4.1, about the Fourier spectrum of the hockey stick function.

Lemma 4.2. *Let n be odd and let $hs_\alpha(x) = min(\alpha, \frac{2\alpha}{n}\sum_{i=1}^{n} x_i)$ then:*

$$\|\widehat{hs_\alpha}\|_2^2 - \widehat{hs_\alpha}(\emptyset)^2 - \widehat{hs_\alpha}([n])^2 = \Theta\left(\frac{\alpha^2}{n}\right),$$

where $[n]$ denotes the set $\{1, \ldots, n\}$.

We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1: By Lemma 4.2 it follows that $\sum_{S \neq \emptyset, S \neq [n]} \widehat{hs_\alpha}(S)^2 = \Omega\left(\frac{\alpha^2}{n}\right) = \Omega(1)$. Since hs_α is a symmetric function and hence its Fourier coefficients for all sets of the same size are the same one can show that it is not $\|hs_\alpha\|_2^2 - \epsilon$-concentrated on $o(n)$-dimensional subspaces. Formally, this is proved in Theorem 4.6 in [KMSY18] which shows that there exists $\epsilon > 0$ such that $\dim_{\|f\|_2^2 - \epsilon}(f) = \Omega(n)$ for any symmetric function which satisfies the condition $\sum_{S \neq \emptyset, S \neq [n]} \hat{f}(S)^2 = \Omega(1)$. $\qquad\square$

4.1 Fourier Spectrum of the "Hockey Stick" Function

In this section, we bound the fourier spectrum of the "hockey stick" function.

Proof of Lemma 4.2: We have:

$$\|hs_\alpha\|_2^2 = 2^{-n}\sum_{x \in \{0,1\}^n} hs_\alpha(x)^2 = 2^{-n}\left(\alpha^2 2^{n-1} + \frac{4\alpha^2}{n^2}\sum_{i=0}^{\lfloor n/2 \rfloor} i^2\binom{n}{i}\right)$$

We also have:

$$\widehat{hs_\alpha}(\emptyset)^2 = \left(2^{-n}\sum_{x \in \{0,1\}^n} hs_\alpha(x)\right)^2 = 2^{-2n}\left(\alpha 2^{n-1} + \frac{2\alpha}{n}\sum_{i=0}^{\lfloor n/2 \rfloor} i\binom{n}{i}\right)^2.$$

Hence:

$$\|\widehat{hs_\alpha}\|_2^2 - \widehat{hs_\alpha}(\emptyset)^2 = 4\alpha^2\left(\frac{1}{16} + \frac{1}{n^2 2^n}\sum_{i=0}^{\lfloor n/2 \rfloor} i^2\binom{n}{i} - \frac{1}{n2^{n+1}}\sum_{i=0}^{\lfloor n/2 \rfloor} i\binom{n}{i} - \frac{1}{n^2 2^{2n}}\left(\sum_{i=0}^{\lfloor n/2 \rfloor} i\binom{n}{i}\right)^2\right)$$

For $i \geq 1$ we have $i\binom{n}{i} = i\frac{n!}{i!(n-i)!} = n\frac{(n-1)!}{(i-1)!(n-i)!} = n\binom{n-1}{i-1}$. Hence:

$$\sum_{i=0}^{\lfloor n/2 \rfloor} i\binom{n}{i} = n\sum_{i=0}^{\lfloor n/2 \rfloor - 1}\binom{n-1}{i} = n(2^{n-2} - \binom{n-1}{(n-1)/2}/2) \approx n2^{n-2}(1 - \frac{\sqrt{2}}{\sqrt{\pi n}}).$$

Similarly we have $i^2\binom{n}{i} = ni\binom{n-1}{i-1} = n\binom{n-1}{i-1} + n(i-1)\binom{n-1}{i-1}$. Hence:

17

$$\sum_{i=0}^{\lfloor n/2 \rfloor} i^2 \binom{n}{i} = n \sum_{i=0}^{\lfloor n/2 \rfloor - 1} \binom{n-1}{i} + n \sum_{i=0}^{\lfloor n/2 \rfloor - 1} i \binom{n-1}{i}$$

$$= n \left(2^{n-1} - \binom{n-1}{(n-1)/2}/2 \right) + n(n-1) \sum_{i=0}^{\lfloor n/2 \rfloor - 2} \binom{n-2}{i}$$

$$= n \left(2^{n-1} - \binom{n-1}{(n-1)/2}/2 \right) + n(n-1) \left(2^{n-3} - \binom{n-2}{\lfloor n-2 \rfloor} \right)$$

$$\approx n \left(2^{n-1} - \frac{\sqrt{2}2^{n-2}}{\sqrt{\pi n}} \right) + n(n-1) \left(2^{n-3} - \frac{\sqrt{2}2^{n-2}}{\sqrt{\pi n}} \right)$$

$$= n^2 2^{n-3} - \frac{\sqrt{2}n^{3/2}2^{n-2}}{\sqrt{\pi}} + 3n2^{n-3} + o(2^n n)$$

Thus we have:

$$\|\widehat{hs_\alpha}\|_2^2 - \widehat{hs_\alpha}(\emptyset)^2 = \Theta \left(\frac{\alpha^2}{n} \right)$$

To complete the proof we will show that $\widehat{hs_\alpha}([n])^2 = 0$. It is well-known (see e.g. [FV15]) that for all $S \subseteq [n]$ such that $|S| \geq 2$ and $i \in S$ it holds that $\widehat{hs_\alpha}(S) = \alpha \frac{\widehat{Maj}(S \setminus \{i\})}{n}$. Using the fact that majority is an odd function its Fourier coefficients on sets of even size are 0. $\qquad \square$

Acknowledgments

We would like to thank Swagato Sanyal for multiple discussions leading to this paper, including the proof of Theorem 2.7. We would like to thank Nikolai Kaprov for his contributions to Section 3.1. We would also like to thank Amit Chakrabarti, Qin Zhang and anonymous reviewers for their comments.

References

[Ack78]   Dragan M Acketa. On the enumeration of matroids of rank-2. *Zbornik radova Prirodnomatematickog fakulteta–Univerzitet u Novom Sadu*, 8:83–90, 1978.

[AGM12a]   Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 459–467, 2012.

[AGM12b]   Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketches: sparsification, spanners, and subgraphs. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2012, Scottsdale, AZ, USA, May 20-24, 2012*, pages 5–14, 2012.

[AHLW16]   Yuqing Ai, Wei Hu, Yi Li, and David P. Woodruff. New Characterizations in Turnstile Streams with Applications. In Ran Raz, editor, *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:22, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[AKL16]   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 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 698–711, 2016.

[AKL17]   Sepehr Assadi, Sanjeev Khanna, and Yang Li. On estimating maximum matching size in graph streams. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1723–1742, 2017.

[AKLY16]   Sepehr Assadi, Sanjeev Khanna, Yang Li, and Grigory Yaroslavtsev. Maximum matchings in dynamic graph streams and the simultaneous communication model. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1345–1364, 2016.

[AMS99]   Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[BB17]   Eric Blais and Abhinav Bommireddi. Testing submodularity and other properties of valuation functions. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pages 33:1–33:17, 2017.

[BCIW12]   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.

[BDF⁺12]   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 2012, Kyoto, Japan, January 17-19, 2012*, pages 1025–1035, 2012.

[Bel92]   Mihir Bellare. A technique for upper bounding the spectral norm with applications to learning. In *Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory, COLT 1992, Pittsburgh, PA, USA, July 27-29, 1992.*, pages 62–70, 1992.

[BEM17]   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 2017, Washington DC, USA, July 24-26, 2017*, pages 13–23, 2017.

[BGKL03]   László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. Comput.*, 33(1):137–166, 2003.

[BH10]   Maria-Florina Balcan and Nicholas J. A. Harvey. Learning submodular functions. *CoRR*, abs/1008.2159, 2010.

[BHNT15]   Sayan Bhattacharya, Monika Henzinger, Danupon Nanongkai, and Charalampos E. Tsourakakis. Space- and time-efficient algorithm for maintaining dense subgraphs on one-pass dynamic streams. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 173–182, 2015.

[BJKK04]   Ziv Bar-Yossef, T. S. Jayram, Robert Krauthgamer, and Ravi Kumar. The sketching complexity of pattern matching. In *Approximation, Randomization, and Combinatorial Optimization, Algorithms and Techniques, 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX*

2004, and 8th International Workshop on Randomization and Computation, RAN-DOM 2004, Cambridge, MA, USA, August 22-24, 2004, Proceedings, pages 261–272, 2004.

[BK97] 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, Ulm, Germany, June 24-27, 1997*, pages 239–246, 1997.

[BMKK14] 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 '14, New York, NY, USA - August 24 - 27, 2014*, pages 671–680, 2014.

[BOSY13] Eric Blais, Krzysztof Onak, Rocco Servedio, and Grigory Yaroslavtsev. Concise representations of discrete submodular functions, 2013.

[BS92] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, ac^0 functions, and spectral norms. *SIAM J. Comput.*, 21(1):33–42, 1992.

[CCE$^+$16] Rajesh Chitnis, Graham Cormode, Hossein Esfandiari, MohammadTaghi Hajiaghayi, Andrew McGregor, Morteza Monemizadeh, and Sofya Vorotnikova. Kernelization via sampling with applications to finding matchings and related problems in dynamic graph streams. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1326–1344, 2016.

[CCHM15] Rajesh Hemant Chitnis, Graham Cormode, Mohammad Taghi Hajiaghayi, and Morteza Monemizadeh. Parameterized streaming: Maximal matching and vertex cover. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1234–1251, 2015.

[CGQ15] Chandra Chekuri, Shalmoli Gupta, and Kent Quanrud. Streaming algorithms for submodular function maximization. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 318–330, 2015.

[CH12] Deeparnab Chakrabarty and Zhiyi Huang. Testing coverage functions. In *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part I*, pages 170–181, 2012.

[CKKL12] 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 2012, Kyoto, Japan, January 17-19, 2012*, pages 1586–1592, 2012.

[CR12] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012.

[CW16] 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, Arlington, VA, USA, January 10-12, 2016*, pages 1365–1373, 2016.

[DIMV14] 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 2014, Austin, TX, USA, October 12-15, 2014. Proceedings*, pages 484–498, 2014.

[DKS12] 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 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, pages 517–528, 2012.

[EHW16] Hossein Esfandiari, MohammadTaghi Hajiaghayi, and David P. Woodruff. Brief announcement: Applications of uniform sampling: Densest subgraph and beyond. In *Proceedings of the 28th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA 2016, Asilomar State Beach/Pacific Grove, CA, USA, July 11-13, 2016*, pages 397–399, 2016.

[ER16] Yuval Emek and Adi Rosén. Semi-streaming set cover. *ACM Trans. Algorithms*, 13(1):6:1–6:22, 2016.

[FIS08] 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.

[FK14] Vitaly Feldman and Pravesh Kothari. Learning coverage functions and private release of marginals. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, pages 679–702, 2014.

[FKV13] 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, June 12-14, 2013, Princeton University, NJ, USA*, pages 711–740, 2013.

[FT16] Martin Farach-Colton and Meng-Tsung Tsai. Tight approximations of degeneracy in large graphs. In *LATIN 2016: Theoretical Informatics - 12th Latin American Symposium, Ensenada, Mexico, April 11-15, 2016, Proceedings*, pages 429–440, 2016.

[FV15] 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 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 923–942, 2015.

[FV16] Vitaly Feldman and Jan Vondrák. Optimal bounds on approximation of submodular and XOS functions by juntas. *SIAM J. Comput.*, 45(3):1129–1170, 2016.

[GHIM09] 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 2009, New York, NY, USA, January 4-6, 2009*, pages 535–544, 2009.

[GHRU13] 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.

[GMR12] Parikshit Gopalan, Raghu Meka, and Omer Reingold. DNF sparsification and a faster deterministic counting algorithm. In *Proceedings of the 27th Conference on Computational Complexity, CCC*, pages 126–135, 2012.

[GMT15] Sudipto Guha, Andrew McGregor, and David Tench. Vertex and hyperedge connectivity in dynamic graph streams. In *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 241–247, 2015.

[Gro97] Vince Grolmusz. On the power of circuits with gates of low l_1 norms. *Theor. Comput. Sci.*, 188(1-2):117–128, 1997.

[HHL16]   Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 282–288, 2016.

[HIMV16]   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 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, pages 371–383, 2016.

[HSZZ06]   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.

[IW03]   Piotr Indyk and David P. Woodruff. Tight lower bounds for the distinct elements problem. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 283–288, 2003.

[Jay10]   T. S. Jayram. Information complexity: a tutorial. In *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, June 6-11, 2010, Indianapolis, Indiana, USA*, pages 159–168, 2010.

[JKS08]   T. S. Jayram, Ravi Kumar, and D. Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.

[JL84]   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.

[KLM+17]   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.

[KMSY18]   Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear sketching over $\mathbb{F}_2$. *To appear in the 33rd Conference on Computational Complexity (CCC'18)*, 2018.

[KN97]   Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[KNP+17]   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 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 475–486, 2017.

[KNR99]   Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

[Kon15]   Christian Konrad. Maximum matching in turnstile streams. In *Algorithms - ESA 2015 - 23rd Annual European Symposium, Patras, Greece, September 14-16, 2015, Proceedings*, pages 840–852, 2015.

[LLN06]   Benny Lehmann, Daniel J. Lehmann, and Noam Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, 55(2):270–296, 2006.

[LLZ11]   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.

[LNW14]   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 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 174–183, 2014.

[Lov14]   Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *Bulletin of the EATCS*, 112, 2014.

[LZ10]   Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Automata, Languages and Programming, 37th International Colloquium, ICALP 2010, Bordeaux, France, July 6-10, 2010, Proceedings, Part I*, pages 475–489, 2010.

[LZ13]   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.

[McG14]   Andrew McGregor. Graph stream algorithms: a survey. *SIGMOD Record*, 43(1):9–20, 2014.

[MNSW98] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *J. Comput. Syst. Sci.*, 57(1):37–49, 1998.

[MO09]   Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *CoRR*, abs/0909.3392, 2009.

[MTVV15] Andrew McGregor, David Tench, Sofya Vorotnikova, and Hoa T. Vu. Densest subgraph in dynamic graph streams. In *Mathematical Foundations of Computer Science 2015 - 40th International Symposium, MFCS 2015, Milan, Italy, August 24-28, 2015, Proceedings, Part II*, pages 472–482, 2015.

[O'D14]   Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[RY13]   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 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1356–1368, 2013.

[SG09]   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 2009, April 30 - May 2, 2009, Sparks, Nevada, USA*, pages 697–708, 2009.

[SV14]   C. Seshadhri and Jan Vondrák. Is submodularity testable? *Algorithmica*, 69(1):1–25, 2014.

[SW12]   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 2012, February 29th - March 3rd, 2012, Paris, France*, pages 477–488, 2012.

[SZ08]   Yaoyun Shi and Zhiqiang Zhang. Communication complexities of symmetric xor functions. *Quantum Inf. Comput*, pages 0808–1762, 2008.

[Tha16]   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.

[TWXZ13] 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 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 658–667, 2013.

[Vad12]   Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[Von10]   Jan Vondrák.   A note on concentration of submodular functions.   *CoRR*, abs/1005.2791, 2010.

[Woo14]   David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014.

[ZS10]   Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of boolean functions. *Theor. Comput. Sci.*, 411(26-28):2612–2618, 2010.

A   Background

A.1   Fourier Analysis

We consider functions[6] from $\mathbb{F}_2^n$ to $\mathbb{R}$. For any fixed $n \geq 1$, the space of these functions forms an inner product space with the inner product $\langle f, g \rangle = \mathbb{E}_{x \in \mathbb{F}_2^n}[f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)g(x)$. The ℓ_2 norm of $f : \mathbb{F}_2^n \to \mathbb{R}$ is $\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathbb{E}_x[f(x)^2]}$ and the ℓ_2 distance between two functions $f, g : \mathbb{F}_2^n \to \mathbb{R}$ is the ℓ_2 norm of the function $f - g$. In other words, $\|f - g\|_2 = \sqrt{\langle f - g, f - g \rangle} = \sqrt{\frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (f(x) - g(x))^2}$.

For $x, y \in \mathbb{F}_2^n$ we denote the inner product as $x \cdot y = \sum_{i=1}^n x_i y_i$. For $\alpha \in \mathbb{F}_2^n$, the *character* $\chi_\alpha : \mathbb{F}_2^n \to \{+1, -1\}$ is the function defined by $\chi_\alpha(x) = (-1)^{\alpha \cdot x}$. Characters form an orthonormal basis as $\langle \chi_\alpha, \chi_\beta \rangle = \delta_{\alpha\beta}$ where δ is the Kronecker symbol. The *Fourier coefficient* of $f : \mathbb{F}_2^n \to \mathbb{R}$ corresponding to α is $\hat{f}(\alpha) = \mathbb{E}_x[f(x)\chi_\alpha(x)]$. The *Fourier transform* of f is the function $\hat{f} : \mathbb{F}_2^n \to \mathbb{R}$ that returns the value of each Fourier coefficient of f. The Fourier ℓ_1 norm, or the *spectral norm* of f, is defined as $\|\hat{f}\|_1 := \sum_{\alpha \in \mathbb{F}_2^n} |\hat{f}(\alpha)|$.

Fact A.1 (Parseval's identity). *For any $f : \mathbb{F}_2^n \to \mathbb{R}$ it holds that $\|f\|_2 = \|\hat{f}\|_2 = \sqrt{\sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)^2}$. Moreover, if $f : \mathbb{F}_2^n \to \{+1, -1\}$ then $\|f\|_2 = \|\hat{f}\|_2 = 1$.*

A.1.1   Fourier ℓ_1-Sampling

The following Fourier ℓ_1-sampling primitive is based on the work of Bruck and Smolensky [Bel92] (see also [Gro97, MO09]). Here we need to analyze its properties for approximating real-valued functions instead of computing Boolean functions as in [Gro97, MO09].

Proposition A.2 (Fourier ℓ_1-sampling). *For any $f : \mathbb{F}_2^n \to \mathbb{R}$ it holds that $\bar{R}_\epsilon^{lin}(f) = O(\|\hat{f}\|_1^2/\epsilon)$.*

Proof. Sample $\mathbf{S} \in \{0, 1\}^n$ from the following distribution: $\Pr[\mathbf{S} = S] = \frac{|\hat{f}(S)|}{\|\hat{f}\|_1}$. Let $Z = sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)\|\hat{f}\|_1$. Then:

$$
\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}_{\mathbf{S}}[sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)\|\hat{f}\|_1] \\
&= \sum_{S \in \{0,1\}^n} sgn(\hat{f}(S))\frac{|\hat{f}(S)|}{\|\hat{f}\|_1}\chi_S(x)\|\hat{f}\|_1 \\
&= \sum_{S \in \{0,1\}^n} \hat{f}(S)\chi_S(x) \\
&= f(x).
\end{aligned}
$$

---

[6] In all Fourier-analytic arguments Boolean functions are treated as functions of the form $f : \mathbb{F}_2^n \to \{+1, -1\}$ where 0 is mapped to 1 and 1 is mapped to -1. Otherwise we use these two notations interchangeably.

Variance of Z is:

$$Var[Z] = \underset{\mathbf{S}}{\mathbb{E}}\left[\left(sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)\|\hat{f}\|_1 - f(x)\right)^2\right]$$

$$= \|\hat{f}\|_1^2 + f(x)^2 - 2\|\hat{f}\|_1 f(x) \underset{\mathbf{S}}{\mathbb{E}}[sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)]$$

$$= \|\hat{f}\|_1^2 - f(x)^2$$

$$\leq \|\hat{f}\|_1^2.$$

Thus averaging Z over $\frac{\|\hat{f}\|_1^2}{\epsilon}$ repetitions reduces variance to at most ϵ as desired. ∎

It follows from Proposition A.2 that additive and coverage functions admit small approximate $\mathbb{F}_2$-sketches.

Corollary A.3. *Let $\ell_w(x): \{0,1\}^n \to \mathbb{R}$ be an additive function $\ell_w(x) = \sum_{i=1}^n w_i x_i$. Then*

$$\bar{R}_\epsilon^{lin}(\ell_w) = O(\min(\|w\|_1^2/\epsilon, n)).$$

Proof. Note that $\|\hat{\ell}_w\|_1 = O(\|w\|_1)$ and hence the bound follows. ∎

Corollary A.4. *If $f: \mathbb{F}_2^n \to [0,1]$ is a coverage function then $\bar{R}_\epsilon^{lin}(f) = O(1/\epsilon)$.*

Proof. It is known (see Lemma 3.1 in [FK14]) that for such coverage functions $\|\hat{f}\|_1 \leq 2$ and hence the desired bound follows from Proposition A.2. ∎

However, direct Fourier ℓ_1-sampling can fail even in some fairly basic situations, e.g. even for budget-additive functions. Consider, for example, the "hockey stick" function: $hs_{\frac{1}{2}}(x) = \min\left(\frac{1}{2}, \frac{1}{n}\sum_{i=1}^n x_i\right)$. Fourier spectrum of this function is well-understood (see. e.g. [FV15]) and in particular $\|\hat{f}\|_1 = 2^{\Omega(n)}$. Nevertheless small sketches for budget-additive functions can be constructed using the following composition theorem.

A function $f: \mathbb{R}^n \to \mathbb{R}$ is α-Lipschitz if $|f(x) - f(y)| \leq \alpha\|x-y\|_2$ for any $x, y \in \mathbb{R}^n$ and some constant $\alpha > 0$[7].

Proposition A.5 (Composition theorem)**.** *If $h: \mathbb{R}^t \to \mathbb{R}$ is an α-Lipschitz function then for any functions $f_1, \ldots, f_t$ where $f_i: \mathbb{F}_2^n \to \mathbb{R}$ it holds that:*

$$\bar{R}_\epsilon^{lin}(h(f_1, \ldots, f_t)) \leq \sum_{i=1}^t \bar{R}_{\epsilon/\alpha^2 t}^{lin}(f_i).$$

Proof. Let $f_1', \ldots, f_t'$ be the sketches of $f_1, \ldots, f_t$ respectively. Applying h to their values we have:

$$\mathbb{E}[(h(f_1', \ldots, f_t') - h(f_1, \ldots, f_t))^2] \leq \mathbb{E}[\alpha^2\|f' - f\|_2^2] = \alpha^2 \sum_{i=1}^t \mathbb{E}[(f_i'(x) - f_i(x))^2] \leq \epsilon. \quad ∎$$

From Corollary A.3 and Proposition A.5 the following bound on approximate $\mathbb{F}_2$-sketch complexity of budget-additive functions follows immediately.

Corollary A.6. *For any budget additive function $f(x) = \min(b, \sum_{i=1}^n w_i x_i)$ it holds that:*

$$\bar{R}_\epsilon^{lin}(f) = O(\min(\|w\|_1^2/\epsilon, n)).$$

Proof. In the composition theorem above we set $h = \min(b, z)$ and hence h is a 1-Lipschitz function of z. Hence we can set $\alpha = 1$ and $t = 1$ by treating $\sum_{i=1}^n w_i x_i$ as f_1 and the proof follows. ∎

---

[7]Note that this definition is slightly different from the corresponding definition for functions over the Boolean hypercube

A.1.2 Optimality of Fourier ℓ_1-Sampling

Let $\ell_w(x)\colon \{0,1\}^n \to \mathbb{R}$ be an additive function $\ell_w(x) = \sum_{i=1}^n w_i x_i$ parametrized by $w \in \mathbb{R}^n$ with non-negative weights $w_1, w_2 \ldots, w_n$. The corresponding XOR-function $\ell_w^+(x,y)$ gives weighted Hamming distance between vectors x and y. The following result can be seen as a generalization of the unweighted Gap Hamming lower bound due to Jayram, Kumar and Sivakumar [JKS08] (see also [IW03, CR12]).

Theorem A.7. *For any additive function ℓ_w of the form $\ell_w(x) = \sum_{i=1}^n w_i x_i$ where $w_i \geq 0$ it holds that for any $\epsilon \geq \|w\|_2^2$:*

$$\bar{R}_\epsilon^{\rightarrow}(\ell_w^+) = \Omega\left(\frac{\|w\|_1^2}{\epsilon}\right).$$

Proof. We use reduction from the standard communication problem INDEX. In this problem Alice is given $a \in \{0,1\}^n$ and Bob is given $t \in [n]$. Alice needs to send one message to Bob so that he can compute a_t. It is well-known that this requires linear communication:

Theorem A.8 ([KNR99]). $R_{1/3}^{\rightarrow}(\mathsf{INDEX}) = \Omega(n)$.

Let n be odd and k be a parameter to be chosen later. Consider an instance of indexing where Alice has an input $a \in \{-1,1\}^k$ and Bob has an index $t \in [k]$. Draw n random vectors $r_1, \ldots, r_n$ where each r_i is uniform over $\{-1,1\}^k$. Construct vectors $x, y \in \{-1,1\}^n$ as follows:

$$x_i = sign(\langle a, r_i \rangle), \qquad y_i = sign(r_{i,t}),$$

where we define $sign(z) = -1$ if $z \leq 0$ and $sign(z) = 1$ if $z > 0$.

Note that if $a_t = 1$ then $\Pr[x_i = y_i] \geq \frac{1}{2} + \frac{c}{\sqrt{k}}$, otherwise $\Pr[x_i = y_i] \leq \frac{1}{2} - \frac{c}{\sqrt{k}}$ for some absolute constant $c > 0$. Now consider the function $\ell_w^+(x,y) = \sum_{i=1}^n w_i(x_i + y_i)$. We will show that for a suitable choice of k with a large constant probability $\ell_w^+(x,y) > \frac{1}{2}\|w\|_1 + 2\sqrt{\epsilon}$ if $a_t = 1$ and $\ell_w^+(x,y) < \frac{1}{2}\|w\|_1 - 2\sqrt{\epsilon} =$ if $a_t = -1$. By Markov's inequality a communication protocol for ℓ_w^+ with expected squared error ϵ has squared error at most 4ϵ (and hence absolute error at most $2\sqrt{\epsilon}$) with probability at least $3/4$. Hence, such a protocol can distinguish these two cases with probability $3/4 - \xi$ where ξ is the error probability introduced by the reduction. If $\xi < 1/12$ then it can solve indexing on strings of length k with probability at least $2/3$ and so a lower bound of $\Omega(k)$ follows.

Indeed, consider the case $a_t = -1$, as the case $a_t = 1$ is symmetric. Let Z_i be a random variable defined as $Z_i = w_i I[x_i = y_i]$. We have $\mathbb{E}[Z_i] \leq w_i\left(\frac{1}{2} - \frac{c}{\sqrt{k}}\right)$. Let $Z = \sum_{i=1}^n Z_i$, then:

$$\mathbb{E}[Z] \leq \sum_{i=1}^n w_i\left(\frac{1}{2} - \frac{c}{\sqrt{k}}\right) = \|w\|_1\left(\frac{1}{2} - \frac{c}{\sqrt{k}}\right).$$

Let $X_i = Z^{\leq i} - \mathbb{E}[Z^{\leq i}]$ where $Z^{\leq i} = \sum_{j=1}^i Z_j$. We have

$$\begin{aligned}
\mathbb{E}[X_{i+1}|X_1, \ldots, X_i] &= \mathbb{E}[Z^{\leq i+1} - \mathbb{E}[Z^{\leq i+1}]|X_1, \ldots, X_i] \\
&= \mathbb{E}[Z_{i+1} - \mathbb{E}[Z_{i+1}] + X_i|X_1, \ldots, X_i] \\
&= \mathbb{E}[Z_{i+1} - \mathbb{E}[Z_{i+1}]] + X_i \\
&= X_i,
\end{aligned}$$

and hence X_i is a martingale. Furthermore, for every i it holds that:

$$|X_i - X_{i-1}| = |Z^{\leq i} - \mathbb{E}[Z^{\leq i}] - Z^{\leq i-1} + \mathbb{E}[Z^{\leq i-1}]| = |Z_i - \mathbb{E}[Z_i]| < |w_i|.$$

We can now use the following form of Azuma's inequality:

Theorem A.9 (Azuma's inequality)**.** *If X_i for $i = 0, 1, \ldots$ is a martingale such that $X_0 = 0$ and $|X_i - X_{i-1}| < c_i$ almost surely then for every integer m and positive real θ it holds that:*

$$\Pr[X_m \geq \theta] \leq e^{-\frac{\theta^2}{2\sum_{i=1}^{m} c_i^2}}.$$

Applying Azuma's inequality we have: $\Pr[X_n \geq \theta] \leq e^{-\frac{\theta^2}{2\|w\|_2^2}}$. Recall that $\mathbb{E}[Z^{\leq n}] \leq \|w\|_1 \left(\frac{1}{2} - \frac{c}{\sqrt{k}} \right)$ and hence:

$$\Pr\left[Z \geq \|w\|_1 \left(\frac{1}{2} - \frac{c}{\sqrt{k}} \right) + \theta \right] \leq e^{-\theta^2/2\|w\|_2^2}.$$

Setting $\theta = \frac{c\|w\|_1}{2\sqrt{k}}$ we have $\Pr\left[Z \geq \frac{\|w\|_1}{2}(1 - c/\sqrt{k}) \right] \leq e^{-\frac{c^2\|w\|_1^2}{8k\|w\|_2^2}}$. If $k = \frac{c^2\|w\|_1^2}{36\|w\|_2^2}$ then:

$$\Pr\left[Z \geq \frac{\|w\|_1}{2} - 3\|w\|_2 \right] \leq e^{-4}.$$

Using similar analysis for the case $a_t = 1$ we conclude that with probability at least $1 - 2e^{-4} > 1 - 1/12$ in this case it holds that $\Pr\left[Z \leq \frac{\|w\|_1}{2} + 3\|w\|_2 \right] \leq e^{-4}$ and hence error probability ξ introduced by the reduction is at most $1/12$. Thus using this reduction we obtain a protocol for solving indexing on strings of length k with probability at least $2/3$ and the lower bound of $\Omega(k) = \Omega(\|w\|_1^2/\|w\|_2^2) = \Omega(\|w\|_1^2/\epsilon)$ follows where we used the fact that $\epsilon \geq \|w\|_2^2$

A.2 Information Theory

Let X be a random variable supported on a finite set $\{x_1, \ldots, x_s\}$. Let $\mathcal{E}$ be any event in the same probability space. Let $\mathbb{P}[\cdot]$ denote the probability of any event. The *conditional entropy* $H(X \mid \mathcal{E})$ of X conditioned on $\mathcal{E}$ is defined as follows.

Definition A.10 (Conditional entropy)**.**

$$H(X \mid \mathcal{E}) := \sum_{i=1}^{s} \mathbb{P}[X = x_i \mid \mathcal{E}] \log_2 \frac{1}{\mathbb{P}[X = x_i \mid \mathcal{E}]}$$

An important special case is when $\mathcal{E}$ is the entire sample space. In that case the above conditional entropy is referred to as the *Shannon entropy* $H(X)$ of X.

Definition A.11 (Entropy)**.**

$$H(X) := \sum_{i=1}^{s} \mathbb{P}[X = x_i] \log_2 \frac{1}{\mathbb{P}[X = x_i]}$$

Let Y be another random variable in the same probability space as X, taking values from a finite set $\{y_1, \ldots, y_t\}$. Then the conditional entropy of X conditioned on Y, $H(X \mid Y)$, is defined as follows.

Definition A.12.

$$H(X \mid Y) = \sum_{i=1}^{t} \mathbb{P}[Y = y_i] \cdot H(X \mid Y = y_i)$$

We next define the binary entropy function $H_b(\cdot)$.

Definition A.13 (Binary entropy)**.** *For $p \in (0, 1)$, the binary entropy of p, $H_b(p)$, is defined to be the Shannon entropy of a random variable taking two distinct values with probabilities p and $1 - p$.*

$$H_b(p) := p \log_2 \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}.$$

The following properties of entropy and conditional entropy will be useful.

Fact A.14. (1) *Let X be a random variable supported on a finite set $\mathcal{A}$, and let Y be another random variable in the same probability space. Then $0 \leq H(X \mid Y) \leq H(X) \leq \log_2 |\mathcal{A}|$.*

(2) *(Sub-additivity of conditional entropy). Let $X_1, \ldots, X_n$ be n jointly distributed random variables in some probability space, and let Y be another random variable in the same probability space, all taking values in finite domains. Then,*

$$H(X_1, \ldots, X_n \mid Y) \leq \sum_{i=1}^{n} H(X_i \mid Y).$$

(3) *Let $X_1, \ldots, X_n$ are independent random variables taking values in finite domains. Then,*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i).$$

(4) *(Taylor expansion of binary entropy in the neighborhood of $\frac{1}{2}$).*

$$H_b(p) = 1 - \frac{1}{2 \log_e 2} \sum_{n=1}^{\infty} \frac{(1 - 2p)^{2n}}{n(2n - 1)}$$

Definition A.15 (Mutual information)**.** *Let X and Y be two random variables in the same probability space, taking values from finite sets. The mutual information between X and Y, $I(X;Y)$, is defined as follows.*

$$I(X;Y) := H(X) - H(X \mid Y).$$

It can be shown that $I(X;Y)$ is symmetric in X and Y, i.e. $I(X;Y) = I(Y;X) = H(Y) - H(Y \mid X)$.

The following observation follows immediately from the first inequality of Fact A.14 (1).

Fact A.16. *For any two random variables X and Y, $I(X;Y) \leq H(X)$.*

B   Missing Proofs

Proof of Theorem 2.7:

Proof. The proof is largely based on a similar proof in [KMSY18] except that here we work with real-valued functions with unbounded norm. In the next two lemmas, we look into the structure of a one-way communication protocol for f^+, and analyze its performance when the inputs are uniformly distributed. We give a lower bound on the number of bits of information that any correct randomized one-way protocol reveals about Alice's input[8], in terms of the linear sketching complexity of f for uniform distribution.

The next lemma bounds the probability of error of a one-way protocol from below in terms of the Fourier coefficients of f, and the conditional distributions of different parities of Alice's input conditioned on Alice's random message.

---

[8]We thus prove an *information complexity* lower bound. See, for example, [Jay10] for an introduction to information complexity.

Lemma B.1. *Let $\epsilon \in [0, \frac{1}{2})$. Let Π be a deterministic one-way protocol for f^+ such that $\mathbb{E}_{x,y \sim U(\mathbb{F}_2^n)}[\Pi(x, y) - f^+(x, y)]^2 \leq \epsilon$. Let M denote the distribution of the random message sent by Alice to Bob in Π. For any fixed message m sent by Alice, let D_m denote the distribution of Alice's input x conditioned on the event that $M = m$. Then,*

$$\epsilon \geq \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \cdot \left(1 - \mathop{\mathbb{E}}_{m \sim M} \left(\mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2 \right).$$

Proof. For any fixed input y of Bob, define $\epsilon_m^{(y)} := \mathbb{E}_{x \sim D_m}(\Pi(x, y) - f^+(x, y))^2$. Thus,

$$\epsilon \geq \mathop{\mathbb{E}}_{m \sim M} \mathop{\mathbb{E}}_{y \sim U(\mathbb{F}_2^n)}[\epsilon_m^{(y)}]. \tag{1}$$

Note that the output of the protocol is determined by Alice's message and y. Hence for a fixed message and Bob's input, if the restricted function has high variance, the protocol is forced to commit error with high probability. Formally, let $a_m^{(y)}$ be the output of the protocol when Alice's message is m and Bob's input is y. Also, define $\mu_m^{(y)} := \mathbb{E}_{x \sim D_m}[f^+(x, y)]$. Then,

$$
\begin{aligned}
\epsilon_m^{(y)} &= \mathop{\mathbb{E}}_{x \sim D_m} \left[(a_m^{(y)} - f^+(x, y))^2 \right] \\
&= \mathop{\mathbb{E}}_{x \sim D_m} \left[((\mu_m^{(y)} - f^+(x, y)) + (a_m^{(y)} - \mu_m^{(y)}))^2 \right] \\
&= \mathop{\mathbb{E}}_{x \sim D_m} \left[((\mu_m^{(y)} - f^+(x, y))^2 + (a_m^{(y)} - \mu_m^{(y)})^2) \right] + 2(a_m^{(y)} - \mu_m^{(y)}) \mathop{\mathbb{E}}_{x \sim D_m} \left[(\mu_m^{(y)} - f^+(x, y)) \right] \\
&\geq \mathop{\mathbb{E}}_{x \sim D_m} \left[(\mu_m^{(y)} - f^+(x, y))^2 \right] \\
&= \mathsf{Var}_{x \sim D_m} \left[f^+(x, y) \right]. \tag{2}
\end{aligned}
$$

Now,

$$
\begin{aligned}
\mathsf{Var}_{x \sim D_m}[f^+(x, y)] &= \mathop{\mathbb{E}}_{x \sim D_m}[f^+(x, y)^2] - \left(\mathop{\mathbb{E}}_{x \sim D_m}[f^+(x, y)] \right)^2 \\
&= \mathop{\mathbb{E}}_{x \sim D_m}[f^+(x, y)^2] - \left(\sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha) \chi_\alpha(y) \mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2 \\
&= \mathop{\mathbb{E}}_{x \sim D_m}[f^+(x, y)^2] - \left(\sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \left(\mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2 \right. \\
&\qquad \left. + \sum_{(\alpha_1, \alpha_2) \in \mathbb{F}_2^n \times \mathbb{F}_2^n : \alpha_1 \neq \alpha_2} \widehat{f}(\alpha_1) \widehat{f}(\alpha_2) \chi_{\alpha_1 + \alpha_2}(y) \mathop{\mathbb{E}}_{x \sim D_m}[\chi_{\alpha_1}(x)] \mathop{\mathbb{E}}_{x \sim D_m}[\chi_{\alpha_2}(x)] \right).
\end{aligned}
$$

Taking expectation over y we have:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{y \sim U(\mathbb{F}_2^n)} \left[\mathsf{Var}_{x \sim D_m}[f^+(x, y)] \right] &= \mathop{\mathbb{E}}_{y \sim U(\mathbb{F}_2^n)} \mathop{\mathbb{E}}_{x \sim D_m}[f^+(x, y)^2] - \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \left(\mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2 \\
&= \mathop{\mathbb{E}}_{x \sim D_m} \mathop{\mathbb{E}}_{y \sim U(\mathbb{F}_2^n)}[f^+(x, y)^2] - \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \left(\mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2 \\
&= \|f\|_2^2 - \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \left(\mathop{\mathbb{E}}_{x \sim D_m}[\chi_\alpha(x)] \right)^2,
\end{aligned}
$$

29

where in the last step we used the fact that for any fixed x we have $\mathbb{E}_{y \sim U(\mathbb{F}_2^n)}[f^+(x,y)^2] = \mathbb{E}_{z \sim U(\mathbb{F}_2^n)}[f^2(z)] = \|f\|_2^2$. Taking expectation over messages it follows using (1), (2) that,

$$
\begin{aligned}
\epsilon &\geq \|f\|_2^2 - \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \cdot \underset{m \sim M}{\mathbb{E}} \left(\underset{x \sim \mathsf{D}_m}{\mathbb{E}} [\chi_\alpha(x)] \right)^2 \\
&= \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 \cdot \left(1 - \underset{m \sim M}{\mathbb{E}} \left(\underset{x \sim \mathsf{D}_m}{\mathbb{E}} [\chi_\alpha(x)] \right)^2 \right).
\end{aligned}
$$

(3)

The second equality above follows from Parseval's identity. The lemma follows. ∎

Let Π be a deterministic protocol such that $\mathbb{E}_{x,y \sim U(\mathbb{F}_2^n)}[(\Pi(x,y) - f^+(x,y))^2] \leq \epsilon$, with optimal cost $c_\Pi := \mathcal{D}_\epsilon^{\rightarrow, U}(f^+)$. To prove our theorem, we use the protocol Π to come up with a subspace of $\mathbb{F}_2^n$. Next, in Lemma B.2 (a) we prove, using Lemma B.1, that f is ξ-concentrated on that subspace where $\xi = \|f\|_2^2 - \epsilon(1 + 2\delta)$. In Lemma B.2 (b) we upper bound the dimension of that subspace in terms of c_Π.

Let $\mathcal{A}_\delta := \{\alpha \in \mathbb{F}_2^n : \mathbb{E}_{m \sim M} \left(\mathbb{E}_{x \sim \mathsf{D}_m} \chi_\alpha(x) \right)^2 \geq \delta\} \subseteq \mathbb{F}_2^n$.

Lemma B.2. *Let* $\delta \in [0, 1/2]$ *and* $\xi = \|f\|_2^2 - \epsilon(1 + 2\delta)$, *then* $\sum_{\alpha \notin \mathsf{span}(\mathcal{A}_\delta)} \widehat{f}(\alpha)^2 \leq \|f\|_2^2 - \xi$.

Proof. We show that f is ξ-concentrated on $\mathsf{span}(\mathcal{A}_\delta)$. By Lemma B.1 we have that

$$
\begin{aligned}
\epsilon &\geq \sum_{\alpha \in \mathsf{span}(\mathcal{A}_\delta)} \widehat{f}(\alpha)^2 \cdot \left(1 - \underset{m \sim M}{\mathbb{E}} \left(\underset{x \sim \mathsf{D}_m}{\mathbb{E}} \chi_\alpha(x) \right)^2 \right) + \sum_{\alpha \notin \mathsf{span}(\mathcal{A}_\delta)} \widehat{f}(\alpha)^2 \cdot \left(1 - \underset{m \sim M}{\mathbb{E}} \left(\underset{x \sim \mathsf{D}_m}{\mathbb{E}} \chi_\alpha(x) \right)^2 \right) \\
&> (1 - \delta) \cdot \sum_{\alpha \notin \mathsf{span}(\mathcal{A}_\delta)} \widehat{f}(\alpha)^2.
\end{aligned}
$$

Thus $\sum_{\alpha \notin \mathsf{span}(\mathcal{A}_\delta)} \widehat{f}(\alpha)^2 < \frac{\epsilon}{1 - \delta} \leq \epsilon \cdot (1 + 2\delta) = \|f\|_2^2 - \xi$ (since $\delta \leq 1/2$). ∎

Now we are ready to complete the proof of Theorem 2.7. Let $\ell = dim(\mathsf{span}(\mathcal{A}_\delta))$. Then it suffices to show that $\ell \leq \frac{2c_\Pi}{\delta}$. Note that $\chi_\alpha(x)$ is a unbiased random variable taking values in $\{1, -1\}$. For each α in the set $\mathcal{A}_\delta$ in Proposition B.2, the value of $\mathbb{E}_{m \sim M} \left(\mathbb{E}_{x \sim \mathsf{D}_m} \chi_\alpha(x) \right)^2$ is bounded away from 0. This suggests that for a typical message m drawn from M, the distribution of $\chi_\alpha(x)$ conditioned on the event $M = m$ is significantly biased. Fact B.3 enables us to conclude that Alice's message reveals $\Omega(1)$ bit of information about $\chi_\alpha(x)$. However, since the total information content of Alice's message is at most c_Π, there can be at most $O(c_\Pi)$ independent vectors in $\mathcal{A}_\delta$. Now we formalize this intuition.

In the derivation below we use several standard facts about properties of entropy and mutual information which can be found in Appendix A.2. We will need the following fact about entropy of a binary random variable. The proof can be found in Appendix A of [KMSY18].

Fact B.3. *For any random variable X supported on $\{1, -1\}$, $H(X) \leq 1 - \frac{1}{2}(\mathbb{E}X)^2$.*

Let $\mathcal{T} = \{\alpha_1, \ldots, \alpha_\ell\}$ be a basis of $\mathsf{span}(\mathcal{A}_\delta)$. Then,

$$
\begin{aligned}
c_\Pi &\geq H(M) \\
&\geq I(M; \chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x)) \\
&= H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x)) - H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x) \mid M) \\
&= \ell - H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x) \mid M) \\
&\geq \ell - \sum_{i=1}^{\ell} H(\chi_{\alpha_i}(x) \mid M) \\
&\geq \ell - \sum_{i=1}^{\ell} (1 - \frac{1}{2}\left(\mathbb{E}[\chi_{\alpha_i}(x) | M])^2\right) && \text{(by Fact B.3)} \\
&\geq \ell - \ell\left(1 - \delta \cdot \frac{1}{2}\right) \\
&= \frac{\ell\delta}{2}.
\end{aligned}
$$

Thus $\ell \leq \frac{2c_\Pi}{\delta}$. ∎

C Lower bound for XS functions

The class of XS functions introduced in [LLN06] corresponds to unit demand functions $f(S) = \max_{i \in S} w_i$.

Theorem C.1. *If f is an XS function corresponding to a collection of distinct weights then $R^{\rightarrow}_{1/3}(f^+) = \Omega(n)$.*

Proof. Let $w_1 > w_2 > \cdots > w_n$. We use a reduction from a standard communication problem called Augmented Indexing, denoted $\mathsf{AI}(x, i)$. In this problem Alice's input is $x \in \mathbb{F}_2^n$ and Bob's input is $i \in [n]$ and the bits $x_1, \ldots, x_{i-1}$.

Theorem C.2 ([MNSW98, BJKK04]). $R^{\rightarrow}_{1/3}(\mathsf{AI}) = \Omega(n)$.

In order to solve $\mathsf{AI}(x, i)$ using a protocol for f^+ set $x' = x$ and $y' = (x_1, \ldots, x_{i-1}, 0, \ldots, 0)$. If $\mathsf{AI}(x, i) = 1$ then $f^+(x' + y') = w_i$, otherwise $f^+(x' + y') \leq w_{i+1}$. Hence an $\Omega(n)$ lower bound follows.

D Sketches under Uniform Distribution

Recall that a function $f : \{0, 1\}^n \to \{0, 1\}$ is ϵ-approximated by a function $g : \{0, 1\}^n \to \{0, 1\}$ if

$$\mathbf{Pr}_{x \in \{0,1\}^n}[f(x) \neq g(x)] \leq \epsilon.$$

Theorem D.1. *[GMR12] Every DNF with width w can be ϵ-approximated by a $\left(w \log \frac{1}{\epsilon}\right)^{O(w)}$-junta.*

Theorem D.2. *Let f be a (θ, m)-LTF of width w DNFs then $\mathcal{D}^{lin, U}_{1-\epsilon}(f) \leq u\left(\frac{1}{\epsilon}, \frac{\theta}{m}, w\right)$ for some function u.*

Proof. Observe that by Lemma 3.15 and standard probability amplification techniques, there exists an $\frac{\epsilon}{2}$-approximation of f, denoted f_h, that is a threshold function of $q = O\left(\left(\frac{\theta}{m}\right)^3 \log^2 \frac{\theta}{m} \log \frac{2}{\epsilon}\right)$ width w disjunctions. Thus, it suffices to take a $\frac{\epsilon}{2q}$-approximation for each of the q width w disjunctions, using Theorem D.1. By a simple union bound, the resulting linear sketch differs from f on at most ϵ fraction of the inputs. The resulting sketch complexity is $\left(w \log \frac{1}{\epsilon}\right)^{O(w)}$ for each of the $O\left(\left(\frac{\theta}{m}\right)^3 \log^2 \frac{\theta}{m}\right)$ disjunctions. ∎

Corollary D.3. *If f can be represented as a (monotone) linear threshold function of (θ_i, m_i)-linear threshold functions, then $\mathcal{D}_{1-\epsilon}^{lin,U}(f) \leq u\left(\frac{1}{\epsilon}, \frac{\theta}{m}\right)$, where $\frac{\theta}{m} = \max_i \frac{\theta_i}{m_i}$.*

Note that any matroid of rank r can be expressed as a linear threshold function of DNFs, where each clause contains the r independent basis elements. Therefore, matroid rank functions can be sketched succinctly under the uniform distribution:

Theorem D.4. *Given $0 < \epsilon < 1$ to be the probability of failure, matroids of rank r have a randomized linear sketch of size $g\left(r, \frac{1}{\epsilon}\right)$ under the uniform distribution, where g is some function.*

In fact, the function $f(\cdot)$ can be improved using the following observation about using juntas to approximate monotone submodular functions.

Theorem D.5 ([BOSY13], Theorem 6). *Let $f : \{0,1\}^n \to \{a_1, \ldots, a_r\}$ be a monotone submodular function. For any $\epsilon \in \left(0, \frac{1}{2}\right)$, f is ϵ-close to a $(z+1)^{r+1}$-junta, where $z = O\left(r \log r + \log \frac{1}{\epsilon}\right)$.*

Corollary D.6. *Matroids of rank r under the uniform distribution have a sketch of size $O\left(\left(r \log r + \log \frac{1}{\epsilon}\right)^{r+1}\right)$, where ϵ is the probability of failure.*

Finally, we note the concentration of matroid rank functions on uniform distributions (see also [Von10]):

Theorem D.7 ([Von10],[GHRU13] Lemma 2.1). *Let $f : 2^U \to \mathbb{R}$ be a 1-Lipschitz submodular function. Then for any product distribution $\mathcal{D}$ over 2^U,*

$$\Pr_{S \sim \mathcal{D}}[|f(x) - \mathbb{E}[f(S)]| \geq t] \leq 2 \exp\left(-\frac{t^2}{2\left(\mathbb{E}[f(S)] + 5t/6\right)}\right).$$

Corollary D.8. *For matroids of rank $\Omega(n)$ and $\epsilon > \frac{1}{\sqrt{n}}$, the expectation of the matroid rank function $rank_M$ suffices for a ϵ-approximation to the matroid rank.*

Proof. Recall that the matroid rank function $rank_M$ is always a submodular 1-Lipschitz function. Thus, matroids of rank $\Omega(n)$ are highly concentrated around their expectation and so to get an ϵ-approximation to the matroid rank, it suffices to simply output the expectation of f, provided $\epsilon > \frac{1}{\sqrt{n}}$. ∎