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

Grigory Yaroslavtsev *

November 12, 2017

Abstract

We study the problem of constructing a linear sketch over $\mathbb{F}_2$ ($\mathbb{F}_2$-sketch) of the smallest dimension that allows to approximate a given real-valued function of the form $f\colon \mathbb{F}_2^n \to \mathbb{R}$ with small expected squared error. We analyze dimension of such sketches for most commonly studied types of valuation functions: additive, budget-additive, coverage, α-Lipschitz submodular and matroid rank functions. Our results are tight in most cases and extend to the distributional version of the problem where the input $x \in \mathbb{F}_2^n$ is drawn uniformly at random. Using known connections with dynamic streaming algorithms both upper and lower bounds obtained in our work extend to space complexity of dynamic streaming algorithms that process $\mathbb{F}_2$ updates.

1 Introduction

Approximate $\mathbb{F}_2$-sketching. In this paper we introduce a study of approximate linear sketching over $\mathbb{F}_2$ ($\mathbb{F}_2$-sketching). This extends the work of [KMSY17] 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 $f\colon \mathbb{F}_2^n \to \mathbb{R}$ that has to be evaluated over an input $x = (x_1, \ldots, x_n)$ we are looking for a distribution over k subsets $\mathbf{S}_1, \ldots, \mathbf{S}_k \subseteq [n]$ such that the following holds: for any input x given parities computed over these sets and denoted as $\chi_{\mathbf{S}_1}(x), \chi_{\mathbf{S}_2}(x), \ldots, \chi_{\mathbf{S}_k}(x)$, it should be possible to compute $f(x)$ with expected squared error at most ϵ.

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. If a fixed distribution D of x is assumed then the definition of error is modified to include an expectation over D in the error guarantee. See Section 2 for the formal definitions.

$\mathbb{F}_2$-sketching and streaming. In the dynamic data stream 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 changes the i-th coordinate of x by adding Δ_i to it. For $x \in \mathbb{F}_2^n$ only the coordinate i is specified and the corresponding bit is flipped. Dynamic stream algorithms aim to minimize space complexity of

*Indiana University, Bloomington, grigory@grigory.us

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 ([KMSY17]) give (almost) optimal space complexity for processing dynamic data streams in the respective update models. 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 applied to distributed algorithms as $\mathbb{F}_2$-sketches can be used for distributed inputs and all our communication lower bounds also hold in the simultaneous message passing communication model, which is stricter than one-way communication.

Our results for valuation functions. 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[1] 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.

- **Additive (linear).** $f(x) = \sum_{i=1}^{n} w_i x_i$ where $w_i \in \mathbb{R}$.
  Our results: For additive functions we show that dimension of $\mathbb{F}_2$-sketches is $O(\min(\|w\|_1^2/\epsilon, n))$ (Corollary 2.5) and give a matching communication lower bound (Theorem 2.11).
- **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 we show that dimension of $\mathbb{F}_2$-sketches is $O(\min(\|w\|_1^2/\epsilon, n))$ (Corollary 2.8) and give a matching communication bound for the "hockey stick" function under the uniform distribution for constant ϵ (Theorem C.1).
- **Coverage.** A function f is a *coverage function* on some universe U of size m if there exists a collection $A_1, \ldots, A_n$ of subsets of U and a vector of non-negative weights $(w_1, \ldots, w_m)$ such that:
  $$f(S) = \sum_{i \in \cup_{j \in S}} w_i.$$
  Our results: We show a simple upper bound of $O(1/\epsilon)$ (Corollary 2.6) 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:
  $$rank_M(S) := \max\{|I|\colon I \subseteq S, I \in \mathcal{I}\}.$$

[1] 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 E).

It follows from the definition that $rank_M$ is always a submodular 1-Lipschitz function.

Our results: For matroids of rank 2 we show an $O(\sqrt{n \log n})$ upper bound the dimension of exact $\mathbb{F}_2$-sketches, i.e for $\epsilon = 0$ (Theorem 3.1).

- **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 3.3). Our construction uses a (scaled) matroid rank function and hence the lower bound also applies to matroids. [2]

Other contributions and previous work. 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], approximation [GHIM09, BDF$^+$12] and property testing [CH12, SV14, BB16]. 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 [KMSY17]. 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 (Section 2.1), originally introduced by Grolmusz [Gro97] (see also [MO09]), turns out to be optimal in its dependence on both spectral norm and the error parameter (see Section 2.3). For Boolean functions a corresponding result is not known as Boolean functions with small spectral norm and necessary properties are hard to construct.

Another 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 in Section 3.2 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.

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

Definition 2.1 (Exact $\mathbb{F}_2$-sketching, [KMSY17])**.** *For a function $f\colon \mathbb{F}_2^n \to \mathbb{R}$ its* exact randomized $\mathbb{F}_2$-sketch complexity *with error δ (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 postprocessing function $g : \mathbb{F}_2^k \to \mathbb{R}$ which satisfies:*

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

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

The following fact is folklore (see e.g. Fact B.7 in [KMSY17]) and can be shown by considering a sketch that uses linear functions over $\mathbb{F}_2^n$ chosen uniformly at random.

[2] 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 without any sketching at all.

Fact 2.2. *For any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $\min_{z\in\{0,1\}} \Pr_x[f(x) = z] \leq \epsilon$ it holds that:*

$$R^{lin}_{\delta}(f) \leq \log \frac{\epsilon 2^{n+1}}{\delta}.$$

Definition 2.3 (Approximate $\mathbb{F}_2$-sketching)**.** *For a function $f\colon \mathbb{F}_2^n \to \mathbb{R}$ we define its ϵ-*approximate randomized $\mathbb{F}_2$-sketch complexity *(denoted as $\bar{R}^{lin}_{\epsilon}(f)$) 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 $\mathbb{F}_2^n$ and a postprocessing function $g : \mathbb{F}_2^k \to \mathbb{R}$ which satisfies:*

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

If g is an unbiased estimator of f then this corresponds to an upper bound on the variance of the estimator. For example, for functions with small spectral norm there exist such approximate $\mathbb{F}_2$-sketches as we show below. 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.

2.1 Fourier ℓ_1-Sampling

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

Proposition 2.4 (Fourier ℓ_1-sampling)**.** *For any $f\colon \mathbb{F}_2^n \to \mathbb{R}$ it holds that $\bar{R}^{lin}_{\epsilon}(f) = O(\|\hat{f}\|_1^2/\epsilon)$.*

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:

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

Variance of Z is:

$$\begin{aligned} Var[Z] &= \mathop{\mathbb{E}}_{\mathbf{S}}\left[\left(sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)\|\hat{f}\|_1 - f(x)\right)^2\right] \\ &= \|\hat{f}\|_1^2 + f(x)^2 - 2\|\hat{f}\|_1 f(x) \mathop{\mathbb{E}}_{\mathbf{S}}[sgn(\hat{f}(\mathbf{S}))\chi_{\mathbf{S}}(x)] \\ &= \|\hat{f}\|_1^2 - f(x)^2 \\ &\leq \|\hat{f}\|_1^2. \end{aligned}$$

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

It follows Proposition 2.4 that additive and coverage functions admit small approximate $\mathbb{F}_2$-sketches.

Corollary 2.5. *Let $\ell_w(x)\colon \{0,1\}^n \to \mathbb{R}$ be an additive function $\ell_w(x) = \sum_{i=1}^n w_i x_i$. Then*

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

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

Corollary 2.6. *If $f\colon \mathbb{F}_2^n \to [0,1]$ is a coverage function then $\bar{R}^{lin}_{\epsilon}(f) = O(1/\epsilon)$.*

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 2.4. ∎

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 $f\colon \mathbb{R}^n \to \mathbb{R}$ is α-Lipschitz if $|f(x) - f(y)| \leq \alpha\|x - y\|_2$ for any $x, y \in \mathbb{R}^n$ and some constant $\alpha > 0$[3].

Proposition 2.7 (Composition theorem)**.** *If $g\colon \mathbb{R}^t \to \mathbb{R}$ is an α-Lipschitz function then for any functions $f_1, \ldots, f_t$ where $f_i\colon \mathbb{F}_2^n \to \mathbb{R}$ it holds that:*

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

Proof. Let $f'_1, \ldots, f'_t$ be the sketches of $f_1, \ldots, f_t$ respectively. Applying g to their values we have: Then we have:

$$\mathbb{E}[(g(f'_1, \ldots, f'_t) - g(f_1, \ldots, f_t))^2] \leq \mathbb{E}[\alpha^2\|f' - f\|_2^2] = \alpha^2 \sum_{i=1}^{t} \mathbb{E}[(f'_i(x) - f(x))^2] \leq \epsilon. \quad ∎$$

From Corollary 2.5 and Proposition 2.7 the following bound on approximate $\mathbb{F}_2$-sketch complexity of budget-additive functions follows immediately by setting $t = \alpha = 1$:

Corollary 2.8. *For any budget additive function $f(x) = min(b, \sum_{i=1}^{n} w_i x_i)$ it holds that:*

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

2.2 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 $f^+\colon \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ 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, KMSY17] for related communication complexity results).

Definition 2.9 (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 $R^{\rightarrow}_{\delta}(f^+)$) of its XOR-function is defined as the smallest size[4] (in bits) of the (randomized using public randomness) message $M(x)$ from Alice to Bob which allows Bob to evaluate $f^+(x, y)$ for any $x, y \in \mathbb{F}_2^n$ with error probability at most δ.*

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

[4]Formally the minimum here is taken over all possible protocols where for each protocol the size of the message $M(x)$ refers to the largest size (in bits) of such message taken over all inputs $x \in \mathbb{F}_2^n$. See [KN97] for a formal definition.

It is easy to see that $R^{\rightarrow}_{\delta}(f^+) \leq R^{lin}_{\delta}(f)$ 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 2.10 (Randomized one-way communication complexity of approximating an XOR function)**.** *For a function $f\colon \mathbb{F}_2^n \to \mathbb{R}$ the* randomized one-way communication complexity *(denoted as $\bar{R}^{\rightarrow}_{\epsilon}(f^+)$) 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.*

Similarly to the observation above it holds that $\bar{R}^{\rightarrow}_{\epsilon}(f^+) \leq \bar{R}^{lin}_{\epsilon}(f)$. We use this fact to show tightness of Fourier ℓ_1-sampling below.

2.3 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 $w \in \mathbb{R}^n$. 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 2.11. *For any additive function ℓ_w if $\|w\|_2^2 = \epsilon$ then it holds that:*

$$\bar{R}^{\rightarrow}_{\epsilon}(\ell^+_w) = \Omega(\min(\|w\|_1^2/\epsilon, n)).$$

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 2.12 ([KNR99])**.** $R^{\rightarrow}_{1/3}(Index) = \Omega(n)$.

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 $sign(z) = 0$ 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 $a_t = -1$, 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 2.13 (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 $\Omega(k) = \Omega(\|w\|_1^2/\epsilon)$ follows.

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

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 2.14 (Approximate distributional $\mathbb{F}_2$-sketching)**.** *For a function $f\colon \mathbb{F}_2^n \to \mathbb{R}$ we define its* ϵ-approximate randomized distributional $\mathbb{F}_2$-sketch complexity *with respect to a distribution D over $\mathbb{F}_2^n$ (denoted as $\bar{\mathcal{D}}_\epsilon^{lin,\mathsf{D}}(f)$) 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 postprocessing function $g : \mathbb{F}_2^k \to \mathbb{F}_2$ which satisfies:*

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

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

Fourier analysis plays an important role in the analysis of distributional $\mathbb{F}_2$-sketch complexity over the uniform distribution. In particular, Fourier concentration on a low-dimensional subspace implies existence of a small sketch which satisfies this guarantee:

Definition 2.15 (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 [KMSY17], adapted for the case of real-valued functions.

Definition 2.16 (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 2.17. *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:

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

3 Sketching Matroid Rank Functions

In this section we analyze sketching complexity of matroid rank functions. We start by analyzing the simplest possible matroids (of rank 2) in Section 3.1 and then give a lower bound for general matroids in Section 3.2.

3.1 Exact $\mathbb{F}_2$-Sketching for Matroids of Rank 2

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

Proof. $\mathbb{F}_2$-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. W.l.o.g 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 $R^{lin}_\delta(rank_M) = R^{lin}_\delta(f_M) + O(\log 1/\delta)$ as by Fact 2.2 we can use $O(\log 1/\delta)$-bit sketch to check whether $x = 0^n$ first and then evaluate $rank_M$ using f_M.

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.*

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. W.l.o.g 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 thresholded Hamming weight function.

In the construction of the sketch we partition sets $S_1, \ldots S_t$ into two groups: large and small. Let $s = \sqrt{n \log n}$ and let $A_1, \ldots A_{t_1}$ be sets of size at least s and $B_1, \ldots, B_{t_2}$ be sets of size less than s. Let:

$$f_M^1(x) = Ham_{\geq 2}\left(\bigvee_{j\in A_1} x_j, \bigvee_{j\in A_2} x_j, \ldots, \bigvee_{j\in A_{t_1}} x_j\right)$$

$$f_M^2(x) = Ham_{\geq 2}\left(\bigvee_{j\in B_1} x_j, \bigvee_{j\in B_2} x_j, \ldots, \bigvee_{j\in B_{t_2}} x_j\right).$$

Our $\mathbb{F}_2$-sketch will consist of the following two parts:

- Sketches for f_M^1 and f_M^2.
- Sketches for $f_1 = \bigvee_{j\in\cup_{i=1}^{t_1} A_i} x_i$ and $f_2 = \bigvee_{j\in\cup_{i=1}^{t_2} B_i} x_i$.

Using the above sketch we can evaluate $f_M(x)$ as follows. If either $f_M^1(x)$ or $f_M^2(x)$ evaluates to 1 according to their sketches then we output 1. Otherwise $f_M(x)$ evaluates to 1 if and only if both f_1 and f_2 evaluate to 1 which we can check based on their sketches. Assuming values computed from sketches for all four functions f_M^1, f_M^2, f_1, f_2 are correct the output of this algorithm equals $f_M(x)$. Thus it suffices to ensure that each of these four sketches errs with probability at most $\frac{1}{12}$.

Functions f_1 and f_2 can be sketched using $O(1)$ bits by Fact 2.2. For each i the function $g_i = \bigvee_{j\in A_i} x_j$ can be sketched using $O(\log 1/\delta)$ bits with probability $1-\delta$ using Fact 2.2. Note that $t_1 \leq n/s$. Hence setting $\delta = \frac{s}{12n}$ by a union bound we have that sketches for all g_i are correct with probability at least $11/12$. Correctness of the sketch for f_M^1 then follows and the size of the sketch is $O(n/s \log(n/s)) = O(\sqrt{n \log n})$.

Thus it suffices to analyze sketch complexity of f_M^2. Note that the number of inputs on which f_M^2 takes value 0 is at most $\sum_{i=1}^{t_2} 2^{|B_i|} \leq n2^s$. Hence by Fact 2.2 the sketch complexity of f_M^2 is $O(s + \log n) = O(\sqrt{n \log n})$.

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

Theorem 3.3. *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 3.4 (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 3.5 ([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:*

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

Theorem 3.6 ([BH10])**.** *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:

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

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

Now consider the family of matroids $\mathcal{M}$ given by Theorem 3.5 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 doesn't 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 fails on a randomly chosen matroid from $\mathcal{M}$ with very high probability probability $1 - 2^{2^{-\Omega(n)}}$ and then take a union bound over all 2^{nd} sketches of dimension at most d.

Indeed, fix any deterministic sketch $\mathcal{S}$ of dimension $d = \frac{1}{2} \log k = n/2^{11}$. 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 3.7. *If $t = \frac{1}{4} 2^{n/2^{11}}$ then $\frac{1}{k} \sum_{b_i \in G} |S_{b_i}| \geq \frac{3}{4}$.*

Proof. We have:

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

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 3.8. *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 unformly 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:

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

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:

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

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

Note that the total number of deterministic sketches of dimension d is at most 2^{nd} 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. ∎

Fix the set $\mathcal{B}$ constructed in Lemma 3.8 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:

$$\begin{aligned}
\frac{1}{k}\sum_{b_i\in\{0,1\}^n}\min(|S^1_{b_i}|,|S^2_{b_i}|) &\geq \frac{1}{k}\sum_{b_i\in G}\min(|S^1_{b_i}|,|S^2_{b_i}|) \\
&\geq \frac{1}{k}\sum_{b_i\in G}\frac{1}{4}|S_{b_i}| && \text{(by Lemma 3.8)} \\
&\geq \frac{3}{4}\times\frac{1}{4} && \text{(by Proposition 3.7)} \\
&\geq \frac{1}{6}.
\end{aligned}$$

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 $(D - b)/D = \frac{1}{2}$.

4 Acknowledgments

We would like to thank Swagato Sanyal for multiple discussions leading to this paper, including the proof of Theorem C.2. We would also like to thank Amit Chakrabarti, Nikolai Karpov and Qin Zhang.

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.

[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.

[BB16] Eric Blais and Abhinav Bommireddi. Testing submodularity and other properties of valuation functions. *CoRR*, abs/1611.07879, 2016.

[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.

[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.

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

[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, RANDOM 2004, Cambridge, MA, USA, August 22-24, 2004, Proceedings*, pages 261–272, 2004.

[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.

[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.

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

[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.

[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.

[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.

[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.

[KMSY17] Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear sketching over $\mathbb{F}_2$. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:174, 2017.

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

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

[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.

[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.

[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.

[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.

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

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

A Fourier analysis

We consider functions[5] 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 $\|f\|_2 = \|\hat{f}\|_2 = 1$.*

B 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 B.1 (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 B.2 (Entropy)**.**

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

[5] 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.

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 B.3.

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

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

Definition B.4 (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 B.5. (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 B.6 (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 B.5 (1).

Fact B.7. *For any two random variables X and Y, $I(X;Y) \leq H(X)$.*

C Communication complexity under 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 C.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:*

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

The proof relies on the following characterization of communication complexity using approximate Fourier dimension that is based on an analogous theorem for Boolean functions from [KMSY17].

Theorem C.2. *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:*

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

Proof. The proof is largely based on a similar proof in [KMSY17] 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[6], 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.

Lemma C.3. *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,*

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

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

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

message is m and Bob's input is y. Also, define $\mu_m^{(y)} := \mathbb{E}_{x\sim \mathsf{D}_m}[f^+(x,y)]$. Then,

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

Now,

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

Taking expectation over y we have:

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

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,

$$\begin{aligned}
\epsilon &\geq \|f\|_2^2 - \sum_{\alpha\in\mathbb{F}_2^n} \widehat{f}(\alpha)^2 \cdot \mathop{\mathbb{E}}_{m\sim M}\left(\mathop{\mathbb{E}}_{x\sim \mathsf{D}_m}[\chi_\alpha(x)]\right)^2 \\
&= \sum_{\alpha\in\mathbb{F}_2^n} \widehat{f}(\alpha)^2 \cdot \left(1 - \mathop{\mathbb{E}}_{m\sim M}\left(\mathop{\mathbb{E}}_{x\sim \mathsf{D}_m}[\chi_\alpha(x)]\right)^2\right). \qquad (3)
\end{aligned}$$

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 C.4 (a) we prove, using Lemma C.3, that f is ξ-concentrated on that subspace where $\xi = \|f\|_2^2 - \epsilon(1+2\delta)$. In Lemma C.4 (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 C.4. *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 C.3 we have that

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

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 C.2. 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 C.4, 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 C.5 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 B. We will need the following fact about entropy of a binary random variable. The proof can be found in Appendix A of [KMSY17].

Fact C.5. *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,

$$\begin{aligned}
c_\Pi &\geq H(M) \\
&\geq I(M; \chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x)) \\
&= H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x)) - H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x) \mid M) \\
&= \ell - H(\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_\ell}(x) \mid M) \\
&\geq \ell - \sum_{i=1}^{\ell} H(\chi_{\alpha_i}(x) \mid M) \\
&\geq \ell - \sum_{i=1}^{\ell} (1 - \frac{1}{2}\left(\mathbb{E}[\chi_{\alpha_i}(x)|M]\right)^2\Big) && \text{(by Fact C.5)} \\
&\geq \ell - \ell\left(1 - \delta\cdot\frac{1}{2}\right) \\
&= \frac{\ell\delta}{2}.
\end{aligned}$$

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

We are now ready to prove Theorem C.1.

Proof. By direct calculation (see Appendix D) 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 isn't $\|hs_\alpha\|_2^2-\epsilon$-concentrated on $o(n)$-dimensional subspaces. Formally, this is proved in Theorem 4.6 in [KMSY17] 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 $\sum_{S\neq\emptyset,S\neq[n]}\hat{f}(S)^2=\Omega(1)$. ∎

D "Hockey stick" function

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

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

Proof. 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:

$$\begin{aligned}
\sum_{i=0}^{\lfloor n/2\rfloor}i^2\binom{n}{i}&=n\sum_{i=0}^{\lfloor n/2\rfloor-1}\binom{n-1}{i}+n\sum_{i=0}^{\lfloor n/2\rfloor-1}i\binom{n-1}{i}\\
&=n\left(2^{n-1}-\binom{n-1}{(n-1)/2}/2\right)+n(n-1)\sum_{i=0}^{\lfloor n/2\rfloor-2}\binom{n-2}{i}\\
&=n\left(2^{n-1}-\binom{n-1}{(n-1)/2}/2\right)+n(n-1)\left(2^{n-3}-\binom{n-2}{\lfloor n-2\rfloor}\right)\\
&\approx n(2^{n-1}-\frac{\sqrt{2}2^{n-2}}{\sqrt{\pi n}})+n(n-1)\left(2^{n-3}-\frac{\sqrt{2}2^{n-2}}{\sqrt{\pi n}}\right)\\
&=n^2 2^{n-3}-\frac{\sqrt{2}n^{3/2}2^{n-2}}{\sqrt{\pi}}+3n2^{n-3}+o(2^n n)
\end{aligned}$$

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. ∎

E 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 E.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 $AI(x, i)$. 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 E.2 ([MNSW98, BJKK04])**.** $R^{\rightarrow}_{1/3}(AI) = \Omega(n)$.

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