Optimality of Linear Sketching Under Modular Updates

Kaave Hosseini
University of California, San Diego, USA
http://cseweb.ucsd.edu/~skhossei/
skhossei@ucsd.edu

Shachar Lovett
University of California, San Diego, USA
https://cseweb.ucsd.edu/~slovett/home.html
slovett@ucsd.edu

Grigory Yaroslavtsev
Indiana University, Bloomington, USA
https://cseweb.ucsd.edu/~slovett/home.html
grigory.yaroslavtsev@gmail.com

Abstract

We study the relation between streaming algorithms and linear sketching algorithms, in the context of binary updates. We show that for inputs in n dimensions, the existence of efficient streaming algorithms which can process $\Omega(n^2)$ updates implies efficient linear sketching algorithms with comparable cost. This improves upon the previous work of Li, Nguyen and Woodruff [23] and Ai, Hu, Li and Woodruff [3] which required a triple-exponential number of updates to achieve a similar result for updates over integers. We extend our results to updates modulo p for integers $p \geq 2$, and to approximation instead of exact computation.

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Sketching and sampling

Keywords and phrases communication complexity, linear sketching, streaming algorithm

Digital Object Identifier 10.4230/LIPIcs.CCC.2019.13

Funding *Kaave Hosseini*: Supported by NSF grant CCF-1614023.
Shachar Lovett: Supported by NSF grant CCF-1614023.

1 Introduction

Linear sketching has emerged in the recent years as a fundamental primitive for algorithm design and analysis including streaming and distributed computing. Applications of linear sketching include randomized algorithms for numerical linear algebra (see survey [32]), graph sparsification (see survey [24]), frequency estimation [4], dimensionality reduction [19], various forms of sampling, signal processing, and communication complexity. In fact, linear sketching has been shown to achieve optimal space complexity [3, 23] for processing very long dynamic data streams, which allow elements to be both inserted and deleted. Linear sketching is also a frequently used tool in distributed computing – summaries communicated between processors in massively parallel computational settings are often linear sketches.

In this paper we focus on linear sketches for functions evaluated modulo p. Namely, functions of the form $f\colon \mathbb{Z}_p^n \to [0, 1]$. Informally, the main result of our work is that for computing such functions linear sketching modulo p achieves almost optimal space complexity in dynamic streaming and distributed simultaneous communication settings. In particular, the setting of p a power of two (say, 32 or 64) is relevant as CPUs perform computations modulo such powers of two.

Exact sketching for binary data

We start with presenting our result in the simplest setting, where $p = 2$ and where the output of f is binary. Namely, we are interested in computing a given Boolean function of the form $f(x)\colon \{0,1\}^n \to \{0,1\}$ using only a small sketch of the input. In this context it is natural to consider sketches which are linear functions over the finite field $\mathbb{F}_2$. Due to their prominence in design of dynamic streaming graph algorithms and other applications [1,2,5,6,9,11,12,14,15,21,22,25] a study of such $\mathbb{F}_2$-sketches has been initiated in [20].

▶ **Definition 1** (Exact $\mathbb{F}_2$-sketching, [20]). *The* exact randomized $\mathbb{F}_2$-sketch complexity *with error δ of a function $f\colon \mathbb{F}_2^n \to \{0,1\}$ is the smallest integer k such that there exists a distribution over linear functions $\ell_1, \dots, \ell_k : \mathbb{F}_2^n \to \mathbb{F}_2$ and a post-processing function $h : \mathbb{F}_2^k \to \{0,1\}$ that satisfies:*

$$\forall x \in \mathbb{F}_2^n\colon \Pr_{\ell_1,\dots,\ell_k,h}[h(\ell_1(x), \ell_2(x), \dots, \ell_k(x)) = f(x)] \geq 1 - \delta.$$

In particular, $\mathbb{F}_2$-sketches naturally allow one to design algorithms for processing data streams in the XOR update model [30] which we refer to as just XOR streams below. In this model the input $x \in \{0,1\}^n$ is generated via a sequence of additive updates to its coordinates $i_1, \dots, i_t$ where each $i_j \in [n]$. Formally, let $x_0 = 0^n$ and let $x_j = x_{j-1} \oplus e_{i_j}$ where e_k is the k-th unit vector. This corresponds to flipping the bit in position i_j in x at time j and after applying the sequence of updates the resulting input is $x = x_t$. The goal of the streaming algorithm is to output $f(x)$. It is easy to see that by flipping linear functions which depend on x_{i_j} when the update i_j arrives one can maintain an $\mathbb{F}_2$-sketch through the XOR stream. Hence the size of the $\mathbb{F}_2$-sketch gives an upper bound on the space complexity of streaming algorithms in XOR streams.[1]

Whether this simple approach in fact achieves optimal space complexity for streaming applications is one of the central questions in the field. Two structural results regarding space optimality $\mathbb{F}_2$-sketching for dynamic streaming are known:

1. $\mathbb{F}_2$-sketches achieve optimal space for streams of length $2^{2^{2^{\Omega(n)}}}$ [3,20,23].
2. $\mathbb{F}_2$-sketches achieve optimal space for streams of length $\tilde{O}(n)$ under the assumption that updates are uniformly random [20].

It is open whether optimality of $\mathbb{F}_2$-sketching holds for short streams without any assumptions about the distribution of updates. In fact, it was conjectured in [20] that such optimality might hold for streams of length only $2n$ (see also Open Problem 78 on `http://sublinear.info` from Banff Workshop on Communication Complexity and Applications, 2017).

In this paper we make major progress towards resolving the gap between the two results discussed above. In particular, we show the following theorem.

▶ **Theorem 2.** *Let $f\colon \{0,1\}^n \to \{0,1\}$. Assume that there exists a streaming algorithm for computing f over XOR streams of length $\Omega(n^2)$, which uses c bits of space. Then the exact randomized $\mathbb{F}_2$-sketch complexity of f is $O(c)$.*

Moreover, using some more advanced tools in additive combinatorics we prove the following.

[1] More precisely, one also needs to derandomize the randomness in the sketching algorithm. This follows from a standard application of Nisan's pseudorandom generator [27]. For completeness we explain this in Appendix A.

▶ **Theorem 3.** *Let $f\colon \{0,1\}^n \to \{0,1\}$. Assume that there exists a streaming algorithm for computing f over XOR streams of length $\Omega(n)$, which uses c bits of space. Then the exact randomized $\mathbb{F}_2$-sketch complexity of f is $O(c^4)$.*

The proof of Theorems 2 and 3 follows from a standard approach in this field, of proving lower bounds for one-way communication protocols. We refer the reader to Section 2 where the model is defined, and to Theorems 8 and 12 which are the formal versions of Theorems 2 and 3, respectively.

Extensions to updates modulo p

We now consider the more general streaming model where updates modulo p are allowed, where $p \geq 2$ is an integer. In this model an underlying n-dimensional vector x is initialized to 0^n and evolves through a sequence of additive updates to its coordinates. These updates are presented to the streaming algorithm as a sequence and have the form $x_i \leftarrow (x_i + \delta_t) \mod p$ changing the i-th coordinate by an additive increment δ_t modulo p in the t-th update. Here δ_t can be an arbitrary positive or negative integer. In this setting the streaming algorithm is required to output a given function f of $\{0, \dots, p-1\}^n$ in the end of the stream.

The definition of the exact randomized $\mathbb{Z}_p$-sketch complexity of f is the natural extension of the definition for $\mathbb{F}_2$.

▶ **Definition 4** (Exact $\mathbb{Z}_p$-sketching)**.** *The* exact randomized $\mathbb{Z}_p$-sketch complexity *with error δ of a function $f\colon \mathbb{Z}_p^n \to \{0,1\}$ is the smallest integer k such that there exists a distribution over linear functions $\ell_1, \dots, \ell_k : \mathbb{Z}_p^n \to \mathbb{Z}_p$ and a post-processing function $h : \mathbb{Z}_p^k \to \{0,1\}$ that satisfies:*

$$\forall x \in \mathbb{Z}_p^n\colon \Pr_{\ell_1,\dots,\ell_k,h}[h(\ell_1(x), \ell_2(x), \dots, \ell_k(x)) = f(x)] \geq 1 - \delta.$$

▶ **Theorem 5.** *Let $f\colon \mathbb{Z}_p^n \to \{0,1\}$. Assume that there exists a streaming algorithm for computing f over streams with modulo p updates of length $\Omega(n^2 \log p)$, which uses c bits of space. Then the exact randomized $\mathbb{Z}_p$-sketch complexity of f is $O(c)$.*

The proof of Theorem 5 for prime p is very similar to the proof of Theorem 2. However, for non-prime p the proof is a bit more involved. We define the relevant model in Section 4, and the relevant theorem that implies Theorem 5 is Theorem 17.

Extensions to approximation

It is also natural to consider real-valued functions $f\colon \{0,1\}^n \to [0,1]$ and to allow the streaming algorithm to compute f with error ϵ (see e.g. [34]). It turns out that technically, a convenient notion of approximation is ℓ_2 approximation. Namely, a randomized function $\mathbf{g}$ (computed by a streaming protocol, or a sketching protocol) ε-approximates f if

$$\mathbb{E}\left[|f(x) - \mathbf{g}(x)|^2\right] \leq \varepsilon \qquad \forall x \in \{0,1\}^n.$$

A similar definition holds for more general functions $f : \mathbb{Z}_p^n \to [0,1]$. The definition of approximate randomized $\mathbb{Z}_p$-sketch complexity is the natural extension of the previous definitions.

▶ **Definition 6** (Approximate $\mathbb{Z}_p$-sketching)**.** *The* approximate randomized $\mathbb{Z}_p$-sketch complexity *with error δ of a function $f\colon \mathbb{Z}_p^n \to [0,1]$ is the smallest integer k such that there exists a distribution over linear functions $\ell_1, \dots, \ell_k : \mathbb{Z}_p^n \to \mathbb{Z}_p$ and a post-processing function $h : \mathbb{Z}_p^k \to [0,1]$ that satisfies:*

$$\forall x \in \mathbb{Z}_p^n\colon \mathop{\mathbb{E}}_{\ell_1,\dots,\ell_k,h}\left[|h(\ell_1(x), \ell_2(x), \dots, \ell_k(x)) - f(x)|^2\right] \leq \delta.$$

▶ **Theorem 7.** *Let $f\colon \mathbb{Z}_p^n \to [0,1]$. Assume that there exists a streaming algorithm which ε-approximates f over streams with modulo p updates of length $\Omega(n^2 \log p)$, which uses c bits of space. Then the approximate randomized $\mathbb{Z}_p$-sketch complexity of f with error $O(\varepsilon)$ is $O(c)$.*

We develop the machinery needed to handle approximation in two steps. First, in Section 3 we prove it for $p = 2$, see in particular Theorem 14. For general p it is done in Section 4, where the relevant theorem which implies Theorem 7 is Theorem 18.

Techniques

The result in Theorem 2 starts with a standard connection between streaming algorithms and a multi-party one-way communication game. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and assume that we are given a streaming algorithm for f which can process streams of Nn updates. We model this as a game with N players, each holding an input in $\{0,1\}^n$ which captures the commulative effect of n binary updates. We denote these inputs as $x_1, \ldots, x_N$.

The players communicate sequentially in N rounds where in the i-th round the i-th player sends a message to the $(i+1)$-th player. The players have access to shared randomness and the i-th message can depend on the $(i-1)$-th message, the input x_i and the shared randomness. The message sent in the final N-th round should be equal to $f(x_1 + \cdots + x_N)$. In fact, our proof works in a more general model where the i-th message can depend on all messages sent by previous players. We refer to the above model as the *one-way broadcasting communication model.*

We show (Theorem 8) that in any protocol which computes f at least one of the messages sent by the players has to be of size $\Omega(k)$ where k is the smallest dimension of an $\mathbb{F}_2$-sketch for f. This immediately implies a space lower bound of $\Omega(k)$ for streaming algorithms in the XOR update model. Indeed, if a streaming algorithm with smaller space existed then the players could just pass its state as their message after applying updates corresponding to their local inputs.

Our proof of the communication lower bound proceeds as follows. First, it will be easier to present the argument for $N+1$ players instead of N players. Assume that there exists a communication protocol which succeeds with probability q and sends at most c bits in every round. This protocol has to succeed for any distribution of the inputs. So, fix a “hard” distribution D over inputs $x \in \{0,1\}^n$.

We sample the inputs $x_1, \ldots, x_{N+1} \in \{0,1\}^n$ to the N players as follows: first, sample $x \sim D$. Then, sample $x_1, \ldots, x_N \in \{0,1\}^n$ uniformly. Finally, set $x_{N+1} = x + x_1 + \ldots + x_N$, so that the sum (modulo two) of the inputs to the $N+1$ players equals x.

Next, an averaging argument then shows that there is a transcript (sequence of messages) $\pi = (m_1, \ldots, m_N)$ of the first N players such that:

(i) Conditioned on the transcript π, the protocol computes f correctly with probability $\approx q$.

(ii) The probability for the transcript π is not too tiny, concretely $\approx 2^{-cN}$.

Once we fixed the transcript π, note that the output of the protocol depends only on the last input x_{N+1}, which we denote by $F(x_{N+1})$. Recall that by our construction, $x_{N+1} = x + x_1 + \ldots + x_N$. Define sets $A_1, \ldots, A_N \subset \{0,1\}^n$ such that whenever $x_1 \in A_1, \ldots, x_N \in A_N$, the players send the transcript π. By (ii) above it holds that the density of a typical A_i is approximately 2^{-c}. Then, if we sample $y_i \in A_i$ uniformly then by (i) we have

$$\Pr_{x \sim D, y_i \in A_i} [F(x + y_1 + \ldots + y_N) = f(x)] \approx q.$$

The next step is to apply Fourier analysis. In particular, we rely of Chang's lemma [10]. This allows us to deduce that there exists a subspace $V \subset \mathbb{F}_2^n$ of co-dimension $O(c)$ such that, if we sample in addition $v \in V$ uniformly, then

$$\Pr_{x \sim D, y_i \in A_i, v \in V} [F(x + y_1 + \ldots + y_N + v) = f(x)] \approx q.$$

Concretely, V is chosen to be orthogonal to the common large Fourier coefficients of the indicator functions of $A_1, \ldots, A_N$. In order for this to hold, it is necessary to choose N large enough so that the sum $y_1 + \ldots + y_N$ "mixes" enough in the group $\mathbb{F}_2^n$. It turns out that $N = \Omega(n)$ is sufficient for this.

This allows us to define a randomized $\mathbb{F}_2$-sketching protocol. Consider the quantity

$$g(x) = \Pr_{y_i \in A_i, v \in V} [F(x + y_1 + \ldots + y_N + v) = 1] \, .$$

The function $g(x)$ depends only on the coset $x + V$, and hence can be computed by a randomized $\mathbb{F}_2$-sketching protocol with complexity equals to the co-dimension of V.

The results for updates modulo p (Theorem 5) and approximation (Theorem 7) follow the same general scheme, except that now players are holding inputs in $\mathbb{Z}_p^n$ and we convert any c-bit protocol with small error into a sketch modulo p of dimension $O(c)$ which has a similar error. In order to achieve mixing in this setting the required number of players is $N = \Omega(n \log p)$.

Distributed computing in the simultaneous communication model

Our results imply that lower bounds on linear sketches modulo p immediately lead to lower bounds for computing additive functions in the simultaneous communication complexity (SMP) model. In this model [7,8] there are N players and a coordinator, who are all aware of a function $f \colon \mathbb{Z}_p^n \to [0, 1]$. The players have inputs $x_1, \ldots, x_N \in \mathbb{Z}_p^n$ and must send messages of minimal size to the coordinator so that the coordinator can compute $f(x_1, \ldots, x_N)$ using shared randomness. If f is additive, i.e. of the form $f(x_1 + \cdots + x_N)$ then this is strictly harder than the one-way broadcasting model described above. Note that dimension of the best linear sketch modulo p for f still translates to a protocol for the SMP model.

Previous work

Most closely related to ours are results of [23] and [3] which stemmed from the work of [16]. In particular [23] shows that under various assumptions about the updates turnstile streaming algorithms can be turned into linear sketches over integers with only a $O(\log p)$ multiplicative loss in space. While this is similar to our results, these approaches inherently require extremely long streams of adversarial updates (of length triply exponential in n in [23]) as they essentially aim to fail any small space algorithm (modeled as a finite state automaton) using a certain sequence of updates. Furthermore, the results of [23] rely on a certain "box constraint" requirement. This requirement says that correctness of the streaming algorithm should be guaranteed for the resulting input $x \in \{-m, \ldots, m\}^n$ even if the intermediate values of the coordinates of x throughout the stream are allowed to be much larger than m in absolute value. While this requirement has been subsequently removed in [3], their results again impose a certain constraint on the class of streaming algorithms they are applicable to. In particular, their Theorem 3.4 which removes the "box constraint" is only applicable to algorithms which use space at most $\frac{c \log m}{n}$ which is only non-trivial for $m = \Omega(2^n)$.

It has been open since the work of [23] and [3] whether similar results can be obtained using the tools from communication complexity, as has been the case for most other streaming lower bounds. While our results do not apply directly to updates over integers, a key component of the [3, 23] technique is to first reduce general automata to linear sketching modulo fixed integers. Hence our result can be seen as an alternative to their reduction which is specific to modular updates and is obtained through communication complexity tools.

Another related line of work is on communication protocols for XOR-functions [17, 20, 26, 29, 31, 33, 34]. For inputs $x_1, \dots, x_N \in \mathbb{F}_2^n$ a multi-parity XOR-function is defined as $f(x_1 + \dots + x_N)$. For the case of $p = 2$ our results are using one-way broadcasting communication complexity for the corresponding XOR-function of interest. While the communication complexity of XOR-functions has been studied extensively in the two-party communication model, to the best of our knowledge prior to our work it has not been considered in the one-way multi-party setting.

2 Sketching for $f : \mathbb{F}_2^n \to \{0, 1\}$

2.1 Model

We use regular letters x, f for deterministic objects and bold letters $\mathbf{x}, \mathbf{f}$ for random variables.

Streaming protocol

Let $F : (\mathbb{F}_2^n)^N \to \{0, 1\}$ be an N-player function, where the players' inputs are $x_1, \dots, x_N \in \mathbb{F}_2^n$. We assume that the players have access to shared randomness $\mathbf{r} \in \{0, 1\}^r$.

A *streaming protocol* for F with c bits of communication is defined as follows. The players send messages in order, where the i-th player's message $\mathbf{m}_i$ depends on her input x_i, the previous player's message $\mathbf{m}_{i-1}$ and the shared randomness $\mathbf{r}$. That is,

$$\mathbf{m}_1 = M_1(x_1, \mathbf{r}),$$
$$\mathbf{m}_i = M_i(x_i, \mathbf{m}_{i-1}, \mathbf{r}), \qquad i = 2, \dots, N$$

where $M_i : \mathbb{F}_2^n \times \{0, 1\}^c \times \{0, 1\}^r \to \{0, 1\}^c$ for $1 \le i \le N-1$ and $M_N : \mathbb{F}_2^n \times \{0, 1\}^c \times \{0, 1\}^r \to \{0, 1\}$, where the output of the protocol is the last message sent $\mathbf{m}_N \in \{0, 1\}$. We may write it as $\mathbf{m}_N = G(x_1, \dots, x_N, \mathbf{r})$, where G respects the protocol structure:

$$G(x_1, \dots, x_N, \mathbf{r}) = M_n(\dots, M_2(x_2, M_1(x_1, \mathbf{r}), \mathbf{r}), \mathbf{r}).$$

The protocol computes F correctly with probability q, if for all possible inputs $x_1, \dots, x_N \in \mathbb{F}_2^n$, it holds that

$$\Pr\left[G(x_1, \dots, x_N, \mathbf{r}) = F(x_1, \dots, x_N)\right] \ge q.$$

One-way broadcasting

A one-way broadcasting protocol is a generalization of a streaming protocol. We introduce this model as our simulation theorem extends to this model seamlessly.

In this model, the message sent by the i-th player is seen by all the players coming after her. Equivalently, the i-th player's message may depend on $\mathbf{m}_1, \dots, \mathbf{m}_{i-1}$,

$$\mathbf{m}_i = M_i(x_i, \mathbf{m}_1, \dots, \mathbf{m}_{i-1}, \mathbf{r}).$$

The notion of a protocol computing F correctly with probability q is defined analogously.

Linear sketches

A function $g : \mathbb{F}_2^n \to \{0,1\}$ is a k-linear-junta if $g(x) = h(\ell_1(x), \ldots, \ell_k(x))$, where each $\ell_i : \mathbb{F}_2^n \to \mathbb{F}_2$ is a linear function, and $h : \mathbb{F}_2^k \to \{0,1\}$ is an arbitrary function. A linear sketch of cost k is a distribution $\mathbf{g}$ over k-linear-juntas. We think of $\mathbf{g}$ as a randomized function $\mathbf{g} : \mathbb{F}_2^n \to \{0,1\}$. It computes f with success probability q if, for every input $x \in \mathbb{F}_2^n$, it holds that

$$\Pr[\mathbf{g}(x) = f(x)] \geq q.$$

Simulation theorem

Let $f : \mathbb{F}_2^n \to \{0,1\}$ and $F : (\mathbb{F}_2^n)^N \to \{0,1\}$ be the N-player function defined by

$$F(x_1, \ldots, x_N) = f(x_1 + \ldots + x_N).$$

We show that if there are sufficiently many players (N is large enough) and F has an efficient one-way broadcasting protocol, then f also has an efficient linear sketch.

▶ **Theorem 8.** *Let $N \geq 10n$ and suppose that F has a one-way broadcasting protocol with c bits of communication per message and success probability q. Then there exists a linear sketch of cost k which computes f with success probability $q - 2^{-\Omega(N)}$, where $k = O(c)$.*

▶ Remark 9. We remark that the statement and proof of Theorem 8 generalizes straightforwardly to the case that $f : \mathbb{F}_p^n \to \{0,1\}$ for any prime p; the only difference is that we have to ensure that $N \geq 10n \log p$. We provide the proof over $\mathbb{F}_2$ to slightly simplify the notation.

2.2 Proof of Theorem 8

By Yao's minimax principle, it will suffice to show that for any distribution D over $\mathbb{F}_2^n$, there exists a k-linear-junta $g : \mathbb{F}_2^n \to \{0,1\}$ such that

$$\Pr_{\mathbf{x} \sim D}[g(\mathbf{x}) = f(\mathbf{x})] \geq q - 2^{-\Omega(N)}.$$

Fix a distribution D. We consider the following distribution over the inputs. It will be easier to assume we have $N+1$ players instead of N players. First, sample $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{F}_2^n$ uniformly, and let $\mathbf{x} \sim D$. Set $\mathbf{x}_{N+1} = \mathbf{x}_1 + \ldots + \mathbf{x}_N + \mathbf{x}$. Under this input distribution, there exists a fixed choice of the shared randomness which attains success probability $\geq q$. Namely, there is a fixed r^* such that for $\mathbf{m}_1 = M_1(\mathbf{x}_1, r^*), \mathbf{m}_2 = M_2(\mathbf{x}_2, \mathbf{m}_1, r^*)$, etc, it holds that

$$\Pr_{\mathbf{x}_1, \ldots, \mathbf{x}_{N+1}}[G(\mathbf{m}_1, \ldots, \mathbf{m}_{N+1}, r^*) = f(\mathbf{x})] \geq q. \tag{1}$$

Let $\boldsymbol{\pi} = (\mathbf{m}_1, \ldots, \mathbf{m}_N) \in \{0,1\}^{cN}$ denote the messages of the first N players. For every possible value π of $\boldsymbol{\pi}$, define

$$a(\pi) = \Pr[\boldsymbol{\pi} = \pi], \qquad b(\pi) = \Pr[G(\mathbf{m}_1, \ldots, \mathbf{m}_{N+1}, r^*) = f(\mathbf{x}) \mid \boldsymbol{\pi} = \pi].$$

Then we may rewrite Equation (1) as

$$\sum_{\pi} a(\pi) b(\pi) \geq q.$$

Let $\delta = 2^{-N}$. By averaging, there exists a choice of $\pi = (m_1, \ldots, m_N)$ such that

(i) $a(\pi) \geq \delta 2^{-cN} = 2^{-(c+1)N}$.
(ii) $b(\pi) \geq q - \delta$.

Define sets $A_i = \{x_i \in \mathbb{F}_2^n : M_i(x_i, m_1, \ldots, m_{i-1}, r^*) = m_i\}$ for $i = 1, \ldots, N$ so that

$$[\boldsymbol{\pi} = \pi] \quad \Leftrightarrow \quad [\mathbf{x}_1 \in A_1, \ldots, \mathbf{x}_N \in A_N].$$

Let $\alpha_i = \frac{|A_i|}{2^n}$ denote the density of A_i. Condition (i) translates to

$$\prod_{i=1}^{N} \alpha_i = a(\pi) \geq 2^{-(c+1)N}. \tag{2}$$

Next, conditioned on $\boldsymbol{\pi} = \pi$, the one-way broadcasting protocol simplifies. First, define $h : \mathbb{F}_2^n \to \{0, 1\}$ as

$$h(x_{N+1}) = G(m_1, \ldots, m_N, M_{N+1}(x_{N+1}, m_1, \ldots, m_N, r^*), r^*).$$

Let $\mathbf{y}_1 \in A_1, \ldots, \mathbf{y}_N \in A_N$ be uniformly and independently chosen. Condition (ii) translates to

$$\Pr\left[h(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N) = f(\mathbf{x})\right] = b(\pi) \geq q - \delta. \tag{3}$$

We next apply Fourier analysis. Let us quickly set some common notation in the following paragraph.

Fourier analysis

Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be a function. Then given $\gamma \in \mathbb{F}_2^n$, the Fourier coefficient $\widehat{f}(\gamma)$ is define as $\widehat{f}(\gamma) = \mathbb{E}_{x \in \mathbb{F}_2^n} f(x)(-1)^{\langle x, \gamma \rangle}$. The function f can be expressed in the Fourier basis as $f(x) = \sum_{\gamma \in \mathbb{F}_2^n} \widehat{f}(\gamma)(-1)^{\langle x, \gamma \rangle}$. Given two functions $f, g : \mathbb{F}_2^n \to \mathbb{R}$, their convolution $f * g : \mathbb{F}_2^n \to \mathbb{R}$, is defined by $f * g(x) = \mathbb{E}_{y \in \mathbb{F}_2^n} f(y)g(x+y)$. Moreover, we have the equality $\widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma)$ for all $\gamma \in \mathbb{F}_2^n$. Given a set $A \subset \mathbb{F}_2^n$ of density $\alpha = \frac{|A|}{2^n}$, define its normalized indicator function $\varphi_A : \mathbb{F}_2^n \to \mathbb{R}$ as

$$\varphi_A(x) = \begin{cases} 1/\alpha & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}.$$

Note that under this normalization, $\widehat{\varphi_A}(0) = \mathbb{E}[\varphi_A] = 1$ and $|\widehat{\varphi_A}(\gamma)| \leq 1$ for all $\gamma \in \mathbb{F}_2^n$.

Going back to the proof, for technical reasons we switch to $\{-1, 1\}$ instead of $\{0, 1\}$. Define the functions $h' = (-1)^h$ and $f' = (-1)^f$. We work with h', f' instead. Note that Equation (3) translates to

$$\Pr\left[h'(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N)f'(\mathbf{x}) = 1\right] = b(\pi) \geq q - \delta. \tag{4}$$

We use the following immediate consequence of Chang's lemma (lemma 3.12 in [10]).

▶ **Lemma 10.** *Let $A \subset \mathbb{F}_2^n$ of density $\alpha = \frac{|A|}{2^n}$. Let $\gamma_1, \ldots, \gamma_k \in \mathbb{F}_2^n$ be linearly independent. Then*

$$\sum_{i=1}^{k} |\widehat{\varphi_A}(\gamma_i)|^2 \leq 8 \log 1/\alpha.$$

Let $S \subset \mathbb{F}_2^n$ be a set of "noticeable" Fourier coefficients of $A_1, \ldots, A_N$, defined as follows. First, define

$$B = \left\{ i \in [N] : \frac{|A_i|}{2^n} \geq 2^{-2(c+1)} \right\}.$$

Equation (2) implies that $|B| \geq N/2$. Then, define

$$S = \left\{ \gamma \in \mathbb{F}_2^n : \sum_{i \in B} |\widehat{\varphi_{A_i}}(\gamma)|^2 \geq |B|/2 \right\}.$$

Let $\gamma_1, \ldots, \gamma_k$ be a maximal set of linearly independent elements in S. Lemma 10 implies that $k = O(c)$. Thus, there exists a subspace $U \subset \mathbb{F}_2^n$ of dimension k such that $S \subset U$. Let $V \subset \mathbb{F}_2^n$ be the orthogonal subspace to U.

▷ Claim 11. Let $\mathbf{y}_1 \in A_1, \ldots, \mathbf{y}_N \in A_N, \mathbf{v} \in V$ be chosen uniformly and independently. Then for every $x \in \mathbb{F}_2^n$ it holds that

$$|\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N)] - \mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})]| \leq 2^n 2^{-N/8}.$$

Proof. We rewrite both expressions using their Fourier expansion:

$$\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots \mathbf{y}_N)] = \varphi_{A_1} * \cdots * \varphi_{A_N} * h'(x) = \sum_{\gamma \in \mathbb{F}_2^n} (-1)^{\langle \gamma, x \rangle} \widehat{h'}(\gamma) \prod_{i=1}^N \widehat{\varphi_{A_i}}(\gamma)$$

and similarly

$$\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots \mathbf{y}_N + \mathbf{v})] = \sum_{\gamma \in \mathbb{F}_2^n} (-1)^{\langle \gamma, x \rangle} \widehat{h'}(\gamma) \widehat{\varphi_V}(\gamma) \prod_{i=1}^N \widehat{\varphi_{A_i}}(\gamma).$$

As V is a subspace, we have $\widehat{\varphi_V}(\gamma) = 1_U(\gamma)$. We can thus bound

$$|\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N)] - \mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})]| \leq \sum_{\gamma \notin U} \prod_{i=1}^N |\widehat{\varphi_{A_i}}(\gamma)|.$$

If $\gamma \in S$ then also $\gamma \in U$. Otherwise, by the construction of S, $|\widehat{\varphi_{A_i}}(\gamma)|^2 \leq 1/2$ for at least $|B|/2$ elements $i \in [N]$. We thus have

$$|\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N)] - \mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})]| \leq 2^n 2^{-|B|/4} \leq 2^n 2^{-N/8}. \qquad ◀$$

Moreover, using the assumption $N \geq 10n$ provides

$$|\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N)] - \mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})]| \leq 2^{-\Omega(N)}.$$

Now, by Equation (4) we already have

$$\mathbb{E}[h'(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N) f'(\mathbf{x})] \geq 2q - 1 - 2^{-\Omega(N)}$$

allowing us to obtain

$$\mathbb{E}[h'(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v}) f'(\mathbf{x})] \geq 2q - 1 - 2^{-\Omega(N)}$$

implying

$$\Pr\left[h(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v}) = f(\mathbf{x})\right] \geq q - 2^{-\Omega(N)}$$

To conclude the proof, define the function $w : \mathbb{F}_2^n \to [0,1]$ as

$$w(x) = \mathbb{E}\left[h(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})\right].$$

Note that $w(x) = W(\ell_1(x), \ldots, \ell_k(x))$, where $\ell_1, \ldots, \ell_k$ are a basis for U, and $W : \mathbb{F}_2^k \to [0,1]$. Define a randomized function $\mathbf{G} : \mathbb{F}_2^k \to \{0,1\}$, where $\Pr[\mathbf{G}(z) = 1] = W(z)$ independently for each $z \in \mathbb{F}_2^k$. Define $\mathbf{g}(x) = \mathbf{G}(\ell_1(x), \ldots, \ell_k(x))$, which is a randomized k-linear-junta and observe that $\mathbf{g}(x)$ and $h(x + \mathbf{y}_1 + \cdots + \mathbf{y}_N + \mathbf{v})$ have the same distribution for every $x \in \mathbb{F}_2^n$. Therefore, we have that

$$\Pr\left[\mathbf{g}(\mathbf{x}) = f(\mathbf{x})\right] = \Pr\left[h(\mathbf{x} + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v}) = f(\mathbf{x})\right] \geq q - 2^{-\Omega(N)},$$

Finally, note that by an averaging argument, we can fix the internal randomness of $\mathbf{g}$ to obtain a k-linear-junta g so that

$$\Pr_{\mathbf{x} \sim D}\left[g(\mathbf{x}) = f(\mathbf{x})\right] \geq q - 2^{-\Omega(N)}.$$

This concludes the proof.

2.3 Sketching for three players

In this section we revise the proof of Theorem 8 and use a tool from additive combinatorics to obtain a version of Theorem 8 that requires only three players. In particular we show the following.

▶ **Theorem 12.** *Let $N \geq 3$ and suppose that F has a one-way broadcasting protocol with c bits of communication per message and success probability q. Then there exists a linear sketch of cost k which computes f with success probability $q - 0.01$, where $k = O(c^4)$.*

Note that the only caveat is that here we obtain $k = O(c^4)$ instead of the bound $k = O(c)$ in Theorem 8. The main technical lemma that we use is an almost periodicity result due to Croot and Sisask [13]. We quote the suitable version of the result bellow.

▶ **Lemma 13** (Theorem 3.2 of [28]). *Let $h' : \mathbb{F}_2^n \to \mathbb{R}$ be a function with $\|h'\|_\infty \leq 1$. Let $A_1, A_2 \subset \mathbb{F}_2^n$ and $|A_1|, |A_2| \geq \alpha|\mathbb{F}_2^n|$ for some $\alpha > 0$. Also let $\varepsilon > 0$ be an error parameter. Then there is a subspace V of codimension $O(\varepsilon^{-2}\log^4(\varepsilon^{-1}\alpha^{-1}))$ so that for all $x \in \mathbb{F}_2^n$,*

$$|h' * \varphi_{A_1} * \varphi_{A_2} * \varphi_V(x) - h' * \varphi_{A_1} * \varphi_{A_2}(x)| \leq \varepsilon.$$

Proof of Theorem 12. The proof is similar to the proof of Theorem 8 so we just explain the point where it differs. Find a transcript π and sets A_1, A_2 and functions $h : \mathbb{F}_2^n \to \{0,1\}$,$h' : \mathbb{F}_2^n \to \{-1,1\}$ as before that satisfy:

1. $[\boldsymbol{\pi} = \pi] \quad \Leftrightarrow \quad [\mathbf{x}_1 \in A_1, \mathbf{x}_2 \in A_2]$.
2. $|A_1|, |A_2| \geq 2^{-O(c)}2^n$
3. $\Pr\left[h(\mathbf{x} + \mathbf{y}_1 + \mathbf{y}_2) = f(\mathbf{x})\right] = b(\pi) \geq q - 0.001$, for uniform and independent $\mathbf{y}_1 \in A_1, \mathbf{y}_2 \in A_2$.

Then instead of Claim 11, apply Lemma 13 to the function h', sets A_1, A_2, and $\varepsilon = 0.001$, to obtain the subspace V of codim $O(c^4)$ that we are looking for. Then continue the proof as before. ◀

3 Sketching for $f : \mathbb{F}_2^n \to [0, 1]$

In this section, we approximate a given function $f : \mathbb{F}_2^n \to [0, 1]$ with an additive error. Both the model and the proof are similar to the previous case.

3.1 Model

Protocols

The notions of *streaming protocol* and *one-way broadcasting protocol* with c bits of communication are defined similar as before, where the only difference is that the last message $\mathbf{m}_N$ takes values in $[0, 1]$ instead of $\{0, 1\}$.

The protocol G is said to *ε-approximate* $F : (\mathbb{F}_2^n)^N \to [0, 1]$ if for all possible inputs $x_1, \dots, x_N \in \mathbb{F}_2^n$, it holds that

$$\mathop{\mathbb{E}}_{\mathbf{r}}\left[|G(x_1, \dots, x_N, \mathbf{r}) - F(x_1, \dots, x_N)|^2\right] \le \varepsilon.$$

Linear sketches

A linear sketch of cost k is a distribution $\mathbf{g}$ over k-linear-juntas, where a k-linear-junta $g : \mathbb{F}_2^n \to [0, 1]$ is defined as before. We think of $\mathbf{g}$ as a randomized function $\mathbf{g} : \mathbb{F}_2^n \to [0, 1]$. The linear sketch $\mathbf{g}$ *ε-approximates* $f : \mathbb{F}_2^n \to [0, 1]$ if, for every $x \in \mathbb{F}_2^n$, it holds that

$$\mathbb{E}\left[|\mathbf{g}(x) - f(x)|^2\right] \le \varepsilon.$$

We prove the following theorem in the rest of the section.

▶ **Theorem 14.** *Let $f : \mathbb{F}_2^n \to [0, 1]$ and assume $N \ge 10n$ and $F : (\mathbb{F}_2^n)^N \to [0, 1]$ is defined by $F(x_1, \dots, x_N) = f(x_1 + \dots + x_N)$. Suppose that F is ε-approximated by a one-way broadcasting protocol with c bits of communication per message. Then there is a linear sketch $\mathbf{g} : \mathbb{F}_2^n \to [0, 1]$ of cost k that ε'-approximates f, where $k = O(c)$ and $\varepsilon' = 2\varepsilon + 2^{-\Omega(N)}$.*

▶ Remark 15. In this case as well, the proof directly generalizes to the case that $f : \mathbb{F}_p^n \to [0, 1]$ for prime p conditioned on $N \ge 10n \log p$.

3.2 Proof of Theorem 14

The proof is similar to the proof of Theorem 8. We point out the necessary modifications. Same as before, we fix an arbitrary distribution D over $\mathbb{F}_2^n$, and show that there exists a k-linear-junta $g : \mathbb{F}_2^n \to [0, 1]$ such that

$$\mathop{\mathbb{E}}_{\mathbf{x} \sim D}\left[|g(\mathbf{x}) - f(\mathbf{x})|^2\right] \le \varepsilon'.$$

To do so, obtain a function $h : \mathbb{F}_2^n \to [0, 1]$ and sets $A_1, \cdots, A_N$ as before, so that for $\mathbf{x} \sim D, \mathbf{y}_i \in A_i$, we have

$$\mathbb{E}\left[|h(\mathbf{x} + \mathbf{y}_1 + \dots + \mathbf{y}_N) - f(\mathbf{x})|^2\right] \le \varepsilon + 2^{-N}. \tag{5}$$

Now we switch to the exponential basis. Define functions $f', h' : \mathbb{F}_2^n \to \mathbb{C}$ by

$$f'(x) = e^{if(x)}, h'(x) = e^{-ih(x)}.$$

where $e = 2.71828\cdots$ is Euler's constant. Note that if $f(x) = h(x)$, then $\mathrm{Re}\left[f'(x)h'(x)\right] = 1$, where $\mathrm{Re}[z]$ is the real component of z. We need the following claim.

▷ Claim 16. Let $\mathbf{z}$ be a $[-1,1]$-valued random variable. Then,

$$1-\frac{1}{2}\mathbb{E}\left[\mathbf{z}^2\right] \leq \mathbb{E}\left[\operatorname{Re}[e^{i\mathbf{z}}]\right] \leq 1-\frac{1}{3}\mathbb{E}\left[\mathbf{z}^2\right].$$

Proof. The Taylor expansion of $\operatorname{Re} e^{iz} = cos(z)$ is $1-\frac{z^2}{2!}+\frac{z^4}{4!}-\cdots$. Therefore

$$1-\frac{z^2}{2} \leq \operatorname{Re}[e^{iz}] \leq 1-\frac{z^2}{3}$$

provided that $z \in [-1,1]$. ◁

Using the lower bound in Claim 16, by taking $\mathbf{z} = h(\mathbf{x}+\mathbf{y}_1+\ldots+\mathbf{y}_N) - f(\mathbf{x})$, we get

$$\mathbb{E}\left[\operatorname{Re}[h'(\mathbf{x}+\mathbf{y}_1+\ldots+\mathbf{y}_N)f'(\mathbf{x})]\right] \geq 1-\varepsilon/2-2^{-\Omega(N)}$$

We apply Claim 11 same as before to obtain subspaces V, U and

$$\mathbb{E}\left[\operatorname{Re}[h'(\mathbf{x}+\mathbf{y}_1+\ldots+\mathbf{y}_N+\mathbf{v})f'(\mathbf{x})]\right] \geq 1-\varepsilon/2-2^{-\Omega(N)} \tag{6}$$

Define a randomized k-linear-junta $\mathbf{r}: \mathbb{F}_2^n \to [0,1]$ as follows. Sample $\mathbf{y}_1, \cdots, \mathbf{y}_N, \mathbf{v}$. Then for every $u \in U$ and $v \in V+u$, set

$$\mathbf{r}(v) = h(u+\mathbf{y}_1+\cdots+\mathbf{y}_N+\mathbf{v}).$$

Observe that for every $x \in \mathbb{F}_2^n$, the randomized functions $\mathbf{r}(x)$ and $h(x+\mathbf{y}_1+\ldots+\mathbf{y}_N+\mathbf{v})$ have identical distributions, and therefore, $e^{-\pi i\mathbf{r}(x)}$ has the same distribution as $h'(x+\mathbf{y}_1+\cdots+\mathbf{y}_N+\mathbf{v})$. Combining this with Equation (6) implies

$$\mathbb{E}\left[\operatorname{Re}[e^{-\pi i\mathbf{r}(\mathbf{x})}f'(\mathbf{x})]\right] \geq 1-\varepsilon/2-2^{-\Omega(N)}$$

By an averaging argument, there is a k-linear-junta $g: \mathbb{F}_2^n \to [0,1]$ that

$$\mathbb{E}\left[\operatorname{Re}[e^{-\pi ig(\mathbf{x})}f'(\mathbf{x})]\right] \geq 1-\varepsilon/2-2^{-\Omega(N)}.$$

Finally, by using the upper bound in Claim 16, we get that

$$\mathbb{E}\left[|g(\mathbf{x})-f(\mathbf{x})|^2\right] \leq 1-4\,\mathbb{E}\left[\operatorname{Re}[e^{-\pi ig(\mathbf{x})}f'(\mathbf{x})]\right] \leq 2\varepsilon+2^{-\Omega(N)}$$

which finishes the proof.

4 Sketching over abelian groups of bounded exponent

Let G be a finite abelian group. We generalize Theorem 8 and Theorem 14 to the case where $f: G \to \{0,1\}$ and $f: G \to [0,1]$, respectively. Even though we are mostly interested in $G = \mathbb{Z}_p^n$, it turns out to be useful to consider this more general formulation. In particular, the proofs of Theorem 8 and Theorem 14 directly generalize to the case of prime p, but the case of composite p requires a more careful analysis. We introduce the required modifications to the definitions and the proofs.

Protocols

The concept of broadcasting protocol and streaming protocol for the function $F(x_1, \cdots, x_N) = f(x_1+\cdots+x_N)$ is defined as before.

Sketching

We modify the previous definition of sketching so that it will be meaningful for arbitrary abelian groups. Let H be an arbitrary subgroup of G. Let $Q = G/H$ be the quotient group. A function $g : G \to \{0, 1\}$ is *H-invariant* if it is constant on every coset $H + w$ for all $w \in Q$. Note that such g can be factored as $g(x) = h(q(x))$ where $q : H \to Q$ is defined by $q(x) = x + H$ and $h : Q \to \{0, 1\}$ is an arbitrary function. A function $g : G \to \{0, 1\}$ has *linear complexity* r if there is a subgroup $H \le G$ so that g is H-invariant and also $|G/H| \le r$. Note that for functions $g : \mathbb{F}_2^n \to \{0, 1\}$, the notion of k-linear-junta is equivalent to linear complexity 2^k.

A *linear sketch* of *complexity* r is a distribution $\mathbf{g} : G \to \{0, 1\}$ over functions $g : G \to \{0, 1\}$ of linear complexity r. We have the following two theorems for functions $g : G \to \{0, 1\}$ and $g : G \to [0, 1]$.

Simulation theorems

Before stating the theorems, we need to introduce one parameter of the group G that will be important here. Let the *exponent* of G, be the smallest m so that $m \cdot g = 0$ for all $g \in G$. Now, we can state the simulation theorems. Same as before, let $f : G \to \{0, 1\}$ (respectively, $f : G \to [0, 1]$) and define $F : G^N \to \{0, 1\}$ (respectively, $F : G^N \to [0, 1]$) by $F(x_1, \cdots, x_N) = f(x_1 + \cdots + x_N)$.

▶ **Theorem 17.** *Let G be an abelian group of exponent m. Let $N \ge 10n \log m$ and suppose that F has a one-way broadcasting protocol with c bits of communication per message and success probability q. Then there exists a linear sketch of complexity r which computes f with success probability $q - 2^{-\Omega(N)}$, where $r = m^{O(c)}$.*

And similarly, for bounded real-valued functions we have the following.

▶ **Theorem 18.** *Let G be an abelian group of exponent m and $N \ge 10n \log m$. Suppose that F is ε-approximated by a one-way broadcasting protocol with c bits of communication per message. Then there is a linear sketch $\mathbf{g} : \mathbb{F}_2^n \to [0, 1]$ of complexity r that ε'-approximates f, where $r = m^{O(c)}$ and $\varepsilon' = 2\varepsilon + 2^{-\Omega(N)}$.*

We need to provide suitable versions of Lemma 10 and Claim 11 here. The other parts of the proof are same as before. We first have to introduce some notation to do Fourier analysis.

Fourier analysis

A character $\gamma : G \to \mathbb{C}^*$ of G is a homomorphism to the group $\mathbb{C}^*$. That is, for every $x, y \in G$ we have $\gamma(x + y) = \gamma(x)\gamma(y)$ and $\gamma(0) = 1$. The dual group of G, denoted by $\widehat{G}$, is the group of all characters of G. $\widehat{G}$ has the group structure introduced by $(\gamma_1 + \gamma_2)(x) = \gamma_1(x)\gamma_2(x)$. In fact $\widehat{G}$ is isomorphic to G. Given any function $f : G \to \mathbb{C}$, we can write f in its Fourier basis as

$$f(x) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma(x)$$

where

$$\widehat{f}(\gamma) = \mathop{\mathbb{E}}_{x \in G} f(x)\overline{\gamma(x)}$$

and $\overline{\gamma(x)}$ is the complex conjugate of $\gamma(x)$. Moreover the convolution operator is defined as before. Given $f, g : G \to \mathbb{C}$, write $f * g(x) = \mathbb{E}_{y \in G} f(y)g(x-y)$ for $x \in G$, which leads to $\widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma)$ for all $\gamma \in \widehat{G}$. Again, given a set $A \subset G$, define its normalized function as $\varphi_A = \frac{|G|}{|A|}1_A$. We need two more definitions. Let $\Gamma \subset \widehat{G}$. Then $\Gamma^{\perp}$, called the *annihilator* of Γ is a subgroup of G defined by $\Gamma^{\perp} = \{x \in G : \gamma(x) = 1, \forall \gamma \in \Gamma\}$. A subset $\Gamma \subset \widehat{G}$ is called *dissociated* if there are no non-trivial solutions to the equation

$$\sum_{\gamma \in \Gamma} a_\gamma \cdot \gamma = 0$$

where each $a_\gamma \in \{-1, 0, 1\}$, $1 \cdot \gamma = \gamma$, $(-1) \cdot \gamma = -\gamma$, and $0 \cdot \gamma = 0$. Let us restate the general form of Chang's lemma [10].

▶ **Lemma 19.** *Let $A \subset G$ be a set of density $\alpha > 0$. Suppose that $\Gamma \subset \widehat{G}$ is a dissociated set. Then*

$$\sum_{\gamma \in \Gamma} |\widehat{\varphi_A}(\gamma)|^2 \leq O(\log \alpha^{-1}).$$

Note that if $\Gamma \subset \widehat{G}$ and $\Gamma' \subset \Gamma$ is the largest dissociated subset of Γ, then $\Gamma \subset \langle \Gamma' \rangle$, the span of Γ'. Since G has exponent m, we have $|\Gamma| \leq m^{|\Gamma'|}$. Moreover, one can show that $\Gamma^{\perp} \cong G/\langle \Gamma \rangle$, therefore, $|\langle \Gamma^{\perp} \rangle| \geq \frac{|G|}{m^{|\Gamma'|}}$.

Finally, the last part to modify in the proof of Theorems 17 and 18 is to obtain a suitable version of Claim 11. Using Chang's lemma as stated above and an analogous proof as before, we can find the function $h' : G \to \mathbb{C}$ taking values in the unit circle, and also the sets $A_1, \cdots, A_N \subset G$ as discussed in the proof of Theorem 8. Also using Lemma 19 we can find a maximal dissociated subset Γ' of size $k = O(c)$. Then take the subgroup $H = \Gamma'^{\perp} \leq G$ (as the analog of the subspace $V \leq \mathbb{F}_2^n$) so that $|G/H| \leq m^k$. The following claim about H is what we need.

▷ Claim 20. Let $\mathbf{y}_1 \in A_1, \ldots, \mathbf{y}_N \in A_N, \mathbf{v} \in H$ be chosen uniformly and independently. Then for every $x \in G$ it holds that

$$|\mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N)] - \mathbb{E}[h'(x + \mathbf{y}_1 + \ldots + \mathbf{y}_N + \mathbf{v})]| \leq 2^{-N/8}|G|.$$

The proof is analogous to the proof of Claim 11. By taking $N \geq 10n \log m$ we can make sure that $2^{-N/8}|G| \leq 2^{-\Omega(N)}$. This finishes the proofs of Theorems 17 and 18.

References

1 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.

2 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.

3 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. doi:10.4230/LIPIcs.CCC.2016.20.

4 Noga Alon, Yossi Matias, and Mario Szegedy. The Space Complexity of Approximating the Frequency Moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999. doi:10.1006/jcss.1997.1545.

5 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.

6 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. doi:10.1137/1.9781611974331.ch93.

7 László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication Complexity of Simultaneous Messages. *SIAM J. Comput.*, 33(1):137–166, 2003. doi:10.1137/S0097539700375944.

8 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.

9 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.

10 Mei-Chu Chang et al. A polynomial bound in Freiman's theorem. *Duke mathematical journal*, 113(3):399–419, 2002.

11 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.

12 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.

13 Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geometric and functional analysis*, 20(6):1367–1396, 2010.

14 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.

15 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.

16 Sumit Ganguly. Lower Bounds on Frequency Estimation of Data Streams (Extended Abstract). In *Computer Science - Theory and Applications, Third International Computer Science Symposium in Russia, CSR 2008, Moscow, Russia, June 7-12, 2008, Proceedings*, pages 204–215, 2008.

17 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. doi:10.1109/FOCS.2016.38.

18 Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. doi:10.1145/1147954.1147955.

19 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.

20 Sampath Kannan, Elchanan Mossel, Swagato Sanyal, and Grigory Yaroslavtsev. Linear Sketching over $\mathbb{F}_2$. In *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*, pages 8:1–8:37, 2018.

21 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. doi:10.1109/FOCS.2017.50.

22 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.

23 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. doi:10.1145/2591796.2591812.

24 Andrew McGregor. Graph stream algorithms: a survey. *SIGMOD Record*, 43(1):9–20, 2014. doi:10.1145/2627692.2627694.

25 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.

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

27 Noam Nisan. Psuedorandom Generators for Space-Bounded Computation. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*, pages 204–212, 1990.

28 Tomasz Schoen and Olof Sisask. Roth theorem for four variables and additive structures in sums of sparse sets. In *Forum of Mathematics, Sigma*, volume 4. Cambridge University Press, 2016.

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

30 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.

31 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. doi:10.1109/FOCS.2013.76.

32 David P. Woodruff. Sketching as a Tool for Numerical Linear Algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014. doi:10.1561/0400000060.

33 Penghui Yao. Parity Decision Tree Complexity and 4-Party Communication Complexity of XOR-functions Are Polynomially Equivalent. *Chicago J. Theor. Comput. Sci.*, 2016, 2016. URL: http://cjtcs.cs.uchicago.edu/articles/2016/12/contents.html.

34 Grigory Yaroslavtsev and Samson Zhou. Approximate $\mathbb{F}_2$-Sketching of Valuation Functions. *In submission to ITCS'19, available at http://grigory.us/files/approx-linsketch.pdf.*, 2018.

A Pseudorandom generators

Below we describe a standard application of pseudorandom generators for space bounded computation by Nisan [27] in the streaming setting following presentation given by Indyk [18]. Such pseudorandom generators can be used to fool any Finite State Machine which uses only $O(S)$ space or $2^{O(S)}$ states. Since a sketch consisting of s numbers modulo p can only take p^s values we can think of such sketches as being Finite State Machines with $O(s \log p)$ space.

Assume that a Finite State Machine Q which uses $O(S)$ bits of space uses at most k blocks of random bits where each block is of length b. The generator $G \colon \{0,1\}^m \to (\{0,1\}^b)^k$ expands a small number m of uniformly random bits into kb bits which "look random" for Q. Formally, let $U(\{0,1\}^t)$ be a uniform distribution over $\{0,1\}^t$. For any discrete random variable let $D[X]$ be the distribution of X interpreted is a vector of probabilities. Let $Q(x)$ denote the state of Q after using the random bits sequence x. Then G is a pseudorandom generator *with parameter* $\epsilon > 0$ for a class $\mathcal{C}$ of Finite State Machines, if for every $Q \in \mathcal{C}$:

$$|D[Q_{x\sim U(\{0,1\}^{bk})}] - D[Q_{x\sim U(\{0,1\}^m)}(G(x))]|_1 \leq \epsilon,$$

where $|y|_1$ denotes the ℓ_1-norm of a vector y.

▶ **Theorem 21** ([27]). *There exists a pseudorandom generator G with parameter $\epsilon = 2^{-O(S)}$ for Finite State Machines using space $O(S)$ such that:*

- *G expands $O(S \log R)$ bits into $O(R)$ bits.*
- *G requires only $O(S)$ bits of storage (in addition to its random input bits)*
- *Any length-$O(S)$ block of $G(x)$ can be computed using $O(\log R)$ arithmetic operations on $O(S)$-bit words.*

Using the above results we can reduce the amount of randomness used by a linear sketch modulo p as follows. Consider any state $\mathcal{S}$ of the linear sketch of dimension s. From the above discussion it follows that this state can be represented using $O(s \log p)$ bits. When a streaming update to coordinate i arrives we need only $O(s \log p)$ bits to generate the i-th row of the linear sketch matrix so that we can add it to the linear sketch. However, in order to ensure consistency, i.e. to make sure that when the i-th row is generate again we get the same result, we still need a lot of memory. Solution to this issue due to [18] follows below.

First, assume that the streaming updates (i, δ_t) are coming in the non-decreasing order of i. In this case we do not have to store the rows of the linear sketch matrix as we can generate them on the fly. Indeed, after the i-th row is generated we can apply it to all updates which contain i since such updates arrive sequentially one after another. This gives an algorithm which uses only $O(s \log p)$ storage and $O(n)$ blocks of random bits of size $O(s \log p)$ each. Hence there exists a pseudorandom generator G which given a random seed of size $O(s \log p \log(n/\delta))$ expands it to a pseudorandom sequence using which instead of the rows the sketch matrix only results in a negligible probability of error. I.e. the resulting state of the sketch can still be used to estimate the value of the function f of interest.

The key observation is that for every fixed random seed the resulting state doesn't depend on the order of updates (i, δ_t) by the commutativity of addition. Hence one can use G even if the updates come in any order.