A Tight Lower Bound for High Frequency Moment Estimation with Small Error

Yi Li
Department of EECS
University of Michigan, Ann Arbor
leeyi@umich.edu

David P. Woodruff
IBM Research, Almaden
dpwoodru@us.ibm.com

Abstract

We show an $\Omega((n^{1-2/p}\log M)/\epsilon^2)$ bits of space lower bound for $(1+\epsilon)$-approximating the p-th frequency moment $F_p = \|x\|_p^p = \sum_{i=1}^n |x_i|^p$ of a vector $x \in \{-M, -M+1, \ldots, M\}^n$ with constant probability in the turnstile model for data streams, for any $p > 2$ and $\epsilon \geq 1/n^{1/p}$ (we require $\epsilon \geq 1/n^{1/p}$ since there is a trivial $O(n \log M)$ upper bound). This lower bound matches the space complexity of an upper bound of Ganguly for any $\epsilon < 1/\log^{O(1)} n$, and is the first of any bound in the long sequence of work on estimating F_p to be shown to be optimal up to a constant factor for any setting of parameters. Moreover, our technique improves the dependence on ϵ in known lower bounds for cascaded moments, also known as mixed norms. We also continue the study of tight bounds on the dimension of linear sketches (drawn from some distribution) required for estimating F_p over the reals. We show a dimension lower bound of $\Omega(n^{1-2/p}/\epsilon^2)$ for sketches providing a $(1+\epsilon)$-approximation to $\|x\|_p^p$ with constant probability, for any $p > 2$ and $\epsilon \geq 1/n^{1/p}$. This is again optimal for $\varepsilon < 1/\log^{O(1)} n$.

1 Introduction

In the standard turnstile model of data streams [Ind07, Mut05], there is an underlying n-dimensional vector x, which we sometimes refer to as the frequency vector, which is initialized to the zero vector and which evolves through a sequence of additive updates to its coordinates. These updates are fed into a streaming algorithm, and have the form $x_i \leftarrow x_i + \delta$, changing the i-th coordinate by the value δ. Here δ is an arbitrary positive or negative integer, and x is guaranteed to satisfy the promise that at all times $x \in \{-M, -M+1, \ldots, M\}^n$. The goal of the streaming algorithm is to make a small number of passes over the data and to use limited memory to compute statistics of x, such as the frequency moments [AMS99], the number of distinct elements [FM83], the empirical entropy [CDM06], and the heavy hitters [CCFC02, CM05a]. Since computing these statistics exactly or deterministically requires a prohibitive $\Omega(n)$ bits of space [AMS99], these algorithms are both randomized and approximate. For most of these problems in the turnstile model, they are quite often studied in the model in which the data stream algorithm can only make a single pass over the data. This is critical in many online applications, such as network traffic monitoring, and when most of the data resides on an external disk, for which multiple passes over it is too costly. In this paper we focus on one-pass streaming algorithms.

We show new lower bounds for approximating the p-th frequency moment F_p, $p > 2$, in a data stream. In this problem the goal is to estimate $\sum_{i=1}^n |x_i|^p$ up to a factor of $1+\epsilon$ with constant probability, where $x \in \{-M, -M+1, \ldots, M\}^n$ and we make the standard assumption that $\log(Mn) = \Theta(\log M)$ and $p > 2$ is a constant. We summarize the sequence of work on this problem in Table 1.

F_p Algorithm	Space Complexity
[IW05]	$O(n^{1-2/p}\epsilon^{-O(1)} \log^{O(1)} n \log(M))$
[BGKS06]	$O(n^{1-2/p}\epsilon^{-2-4/p} \log n \log^2(M))$
[MW10]	$O(n^{1-2/p}\epsilon^{-O(1)} \log^{O(1)} n \log(M))$
[AKO11]	$O(n^{1-2/p}\epsilon^{-2-6/p} \log n \log(M))$
[BO10]	$O(n^{1-2/p}\epsilon^{-2-4/p} \log n \cdot g(p,n) \log(M))$
[And]	$O(n^{1-2/p} \log n \log(M) \epsilon^{-O(1)})$
[Gan11], **Best upper bound**	$O(n^{1-2/p}(\epsilon^{-2} + \epsilon^{-4/p} \log n) \cdot \log(M))$
[AMS99]	$\Omega(n^{1-5/p})$
[Woo04]	$\Omega(\epsilon^{-2})$
[BYJKS04]	$\Omega(n^{1-2/p-\gamma}\epsilon^{-2/p})$, any constant $\gamma > 0$
[CKS03]	$\Omega(n^{1-2/p}\epsilon^{-2/p})$
[WZ12]	$\Omega(n^{1-2/p}\epsilon^{-4/p} / \log^{O(1)} n)$
[Gan12]	$\Omega(n^{1-2/p}\epsilon^{-2} / \log n)$
This paper	$\Omega(n^{1-2/p}\epsilon^{-2} \log(M))$

Table 1: Results are in terms of bits and for constant $p > 2$. Here, $g(p,n) = \min_{c \text{ constant}} g_c(n)$, where $g_1(n) = \log n$, $g_c(n) = \log(g_{c-1}(n))/(1-2/p)$. For brevity, we only list those results which work in the general turnstile model, and for which bounds for general ϵ have been derived. For other recent interesting work, we refer the reader to [BO12], which requires the insertion-only model and does not have bounds for general $\epsilon > 0$. We also start the upper bound timeline with [IW05], since that is the first work which achieved an exponent of $1-2/p$ for n. For earlier works which achieved worse exponents for n, see [AMS99, CK04, Gan04a, Gan04b]. We note that [AMS99] initiated the problem and obtained an $O(n^{1-1/p}\epsilon^{-2} \log(M))$ bound in the insertion-only model. We also omit from the table previous lower bounds which hold for linear sketches rather than for the turnstile model [ANPW13, PW12], though these are discussed in the Introduction.

The previous best upper bound is due to Ganguly [Gan11], and is $O(n^{1-2/p}\epsilon^{-2} \log n \log M / \min(\log n, \epsilon^{4/p-2}))$. Notice that for $\epsilon < 1/\log^{O(1)} n$, this bound simplifies to $O(n^{1-2/p}\epsilon^{-2} \log M)$. The previous best lower bound is due to [CKS03, Gan12], and is $\Omega(n^{1-2/p}\epsilon^{-2}/\log n + n^{1-2/p}\epsilon^{-2/p})$. We improve the space complexity lower bound, in bits, to $\Omega(n^{1-2/p}\epsilon^{-2}\log M)$ for any $\epsilon > 1/n^{1/p}$ (we require $\epsilon > 1/n^{1/p}$ since there is a trivial $O(n \log M)$ upper bound). In light of the upper bound given above, our lower bound is optimal for any $\epsilon < 1/\log^{O(1)} n$ and constant $p > 2$. This is an important range of parameters; even in applications with 1% error, i.e., $\epsilon = .01$, we have that for, e.g., $n = 2^{32}$, $\epsilon < 1/\log n$. Understanding the limitations of streaming algorithms in terms of ϵ is also the focus of a body of work in the streaming literature, see, for example, [BYJK$^+$02, CCM10, CW09, Gan08, Gan09, GC07, IW03, KNPW11, PT07, Woo04]. Our lower bound gives the first asymptotically optimal bound for any setting of parameters in the long line of work on estimating F_p in a data stream.

A few recent works [ANPW13, PW12] also study the "sketching model" of F_p-estimation in which the underlying vector x is in $\mathbb{R}^n$, rather than in the discrete set $\{-M, -M+1, \ldots, M\}^n$. One seeks a distribution over linear maps $A : \mathbb{R}^n \to \mathbb{R}^s$, for some $s \ll n$, so that for any fixed vector $x \in \mathbb{R}^n$, one can $(1+\epsilon)$-approximate $\|x\|_p^p$ with constant probability by applying an estimation procedure $E : \mathbb{R}^s \to \mathbb{R}$ to Ax. One seeks the smallest possible s for a given ϵ and n. Lower bounds in the turnstile model do not imply lower bounds in the sketching model. Indeed, if the input vector $x \in \{-M, -M+1, \ldots, M\}^n$, then the inner product of x with the single vector $(1, 1/(M+1), 1/(M+1)^2, \ldots, 1/(M+1)^{n-1})$ is enough to recover x, so a sketching dimension of $s = 1$ suffices. Previously, it was known that $s = \Omega(n^{1-2/p})$ [PW12], which for constant $p > 2$ and constant $\epsilon > 0$ was recently improved to $s = \Omega(n^{1-2/p}\log n)$ [ANPW13]. We note that the upper bound of [Gan11] is a linear sketch with $s = O(n^{1-2/p}\epsilon^{-2})$ dimensions for any $\epsilon < 1/\log^{O(1)} n$. We improve the lower bound on s for general ϵ, obtaining an $s = \Omega(n^{1-2/p}\epsilon^{-2})$ lower bound. Our lower bound matches the upper bound of [Gan11] for $\epsilon < 1/\log^{O(1)} n$ up to a constant factor, and improves the lower bound of [ANPW13] for $\epsilon < 1/\log^{O(1)} n$.

Our Approach: To prove our lower bound in the turnstile model, we define a variant of the ℓ_∞^k communication problem [BYJKS04]. In this problem there are two parties, Alice and Bob, holding vectors $x, y \in \{-M, -M+1, \ldots, M\}^n$ respectively, and their goal is to decide if $\|x-y\|_\infty = \max_{i \in [n]} |(x-y)_i| \leq 1$ or there exists a unique $i \in [n]$ for which $|(x-y)_i| \geq k$ and for all $j \neq i$, $|(x-y)_j| \leq 1$. The standard reduction to frequency moments is to set $k = \epsilon^{1/p} n^{1/p}$, from which one can show that any streaming algorithm for outputting a $(1+\epsilon)$-approximation to F_p can be used to build a communication protocol for solving ℓ_∞^k with communication proportional to the algorithm's space complexity. Using the communication lower bound of $\Omega(n/k^2)$ for the ℓ_∞^k problem, this gives the bound $\Omega(n^{1-2/p}\epsilon^{-2/p})$.

Our first modification is to instead set $k = \epsilon n^{1/p}$, which gives a communication lower bound of $\Omega(n^{1-2/p}\epsilon^{-2})$. However, the reduction from approximating F_p no longer works. To remedy this, we introduce a third player Charlie whose input is $z \in \{0^n, n^{1/p}e_1, \ldots, n^{1/p}e_n\}$, where e_i denotes the i-th standard unit vector, and we seek a $(1+\epsilon)$-approximation to $\|x-y+z\|_\infty$. The main point is that if $|x_i - y_i| = \epsilon n^{1/p}$, then $\|x-y+z\|_\infty$ differs by a factor of $1+\epsilon$ depending on whether or not Charlie's input is $n^{1/p}e_i$, 0^n, or $n^{1/p}e_j$ for some $j \neq i$. Note that Charlie has no information as to whether $|x_i - y_i| = k$ or $|x_i - y_i| \leq 1$, which is determined by Alice and Bob's inputs. One can think of this as an extension to the classical indexing problem, which involves two players, in which the first player has a string $x \in \{0,1\}^n$, the second player an index $i \in [n]$, and the second player needs to output x_i. Now, we have Alice and Bob solving multiple single-coordinate problems and Charlie is indexing into one of these problems.

This modification allows us to strengthen the problem for use in applications. We choose the legal inputs x, y, z to the three-player problem to have the following promise: (1) $\|x-y+z\|_\infty \leq 1$, (2) there is a unique i for which $|(x-y+z)_i| = \epsilon n^{1/p}$ and all other $j \neq i$ satisfy $|(x-y+z)_j| \leq 1$, or (3) there is a unique i for which $|(x-y+z)_i|$ is either $(1+\epsilon)n^{1/p}$ or $(1-\epsilon)n^{1/p}$ and all other $j \neq i$ satisfy $|(x-y+z)_j| \leq 1$. Using the 1-way property of a communication protocol, we can adapt the argument in [BYJKS04] for the ℓ_∞^k problem to show an $\Omega(n^{1-2/p}\epsilon^{-2})$ lower bound for this 3-player problem. Here we use the intuitive fact that Alice and Bob need to solve the ℓ_∞^k problem with $k = \epsilon n^{1/p}$ because if Charlie has the input $z = n^{1/p}e_i$, then $\|x-y+z\|_\infty$ differs by a factor of $(1+\epsilon)$ depending on whether $|(x-y)_i| = \epsilon n^{1/p}$ or $|(x-y)_i| \leq 1$. Moreover, Alice and Bob have no information about z since the protocol is 1-way.

We show that a streaming algorithm providing a $(1+\epsilon)$-approximation to F_p can decide which of the three cases the input is in by invoking it twice in the reduction to the communication problem. Here, Alice, Bob, and Charlie create local streams σ_A, σ_B, and σ_C, Alice sends the state of the algorithm on σ_A to Bob, who computes the state of the algorithm on $\sigma_A \circ \sigma_B$ who sends it to Charlie. Charlie then queries a $(1+\epsilon)$-approximate F_p value of $\sigma_A \circ \sigma_B$, together with a $(1+\epsilon)$-approximate F_p value of $\sigma_A \circ \sigma_B \circ \sigma_C$. Assuming both query responses are correct, we can solve this new communication problem, yielding an $\Omega(n^{1-2/p}\epsilon^{-2})$ bits of space lower bound.

To improve the space further, we define an augmented version of this 3-player problem, in which Alice, Bob, and Charlie have $r = \Theta(\log M)$ independent instances of this problem, denoted x^i, y^i, z^i, for $i \in [r]$. Charlie additionally has an index $I \in [r]$ together with (x^i, y^i) for all $i > I$. His goal is to solve the I-th instance of the communication problem. This problem can be seen as an extension to the classical augmented indexing problem, which involves two players, in which the first player has a string $x \in \{0,1\}^n$, the second player an index $i \in [n]$ together with $x_{i+1}, \ldots, x_n$, and the second player needs to output x_i. We now have a "functional" version of augmented indexing, in which Alice and Bob solve multiple instances of a problem, and Charlie's input indexes one of these problems. Via a direct sum argument [BYJKS04, CSWY01], we show our problem has randomized communication complexity $\Omega(n^{1-2/p}\epsilon^{-2}\log M)$. Finally, we show how a streaming algorithm for $(1+\epsilon)$-approximating F_p can be used to solve this augmented problem.

We believe our technique will improve the dependence on ϵ in space lower bounds for other problems in the data stream literature. For example, we can improve the dependence on ϵ in known lower bounds for estimating cascaded moments, also known as mixed norms [AKO11, CM05b, JW09]. Here there is an underlying $n \times d$ matrix A, and the goal is to estimate $\ell_p(\ell_q)(A) = (\sum_{i=1}^n \|A_i\|_q^p)^{1/p}$, where A_i is the i-th row of A. In [JW09] (beginning of Section 2) a lower bound of $\Omega(n^{1-2/p}d^{1-2/q})$ is shown for constant ϵ and $p, q \geq 2$ via a reduction to the so-called t-player set disjointness problem for $t = 2n^{1/p}d^{1/q}$. Straightforwardly setting $t = \Theta(\epsilon^{1/k}n^{1/p}d^{1/q})$, their proof establishes a lower bound of $\Omega(n^{1-2/p}d^{1-2/q}\epsilon^{-2/p})$ for general ϵ. Our technique also applies to the t-player set disjointness problem, by introducing a $(t+1)$-st player Charlie with an input $z \in \{0^{nd}, n^{1/p}d^{1/q}e_ie_j^T \text{ for } i \in [n], j \in [d]\}$, and applying analogous ideas to those given above. This results in a new lower bound of $\Omega(n^{1-2/p}d^{1-2/q}\epsilon^{-2})$. The same ideas apply to improving the $\Omega(n^{1/2})$ lower bound for $\ell_2(\ell_0)(A)$ given in [JW09] (here $\|A_i\|_0$ denotes the number of non-zero entries of A_i). A straightforward adaptation of the arguments in [JW09] for general ϵ gives a lower bound of $\Omega(n^{1/2}\epsilon^{-1/2})$, while our technique strengthens this to $\Omega(n^{1/2}\epsilon^{-1})$. We sketch these improvements in Appendix D.

Our lower bound in the sketching model is simpler and perhaps surprising. We consider two cases: the input $x \in \mathbb{R}^n$ is equal to $g + n^{1/p}e_i$ for a vector g of i.i.d. standard normal random variables and a random standard unit vector e_i, or the input x is equal to $g' + n^{1/p}(1+\epsilon)e_i$ for a vector g' of i.i.d. standard normal random variables. By Yao's minimax principle, there exists a fixed $s \times n$ sketching matrix A for which the variation distance between distributions $A(g + n^{1/p}e_i)$ and $A(g' + n^{1/p}(1+\epsilon)e_i)$ is large. Since we can, w.l.o.g., assume the rows of A are orthonormal (given Ax, one can always compute LAx for any change of basis matrix L for the rowspace of A), this implies the variation distance between $h + n^{1/p}A_i$ and $h' + n^{1/p}(1+\epsilon)A_i$ is large, where h, h' are s-dimensional vectors of i.i.d. standard normal random variables and A_i is the i-th column of A. However, for a random i, since the rows of A are orthonormal, $\|A_i\|_2$ is only about $O(\sqrt{s/n})$. For such i, this contradicts a standard variation distance upper bound between two shifted s-dimensional Gaussian vectors unless $s = \Omega(n^{1-2/p}/\epsilon^2)$.

2 Preliminaries

Notations. We denote the canonical basis of $\mathbb{R}^n$ by $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Let $[n]$ denote the set $\{1, \ldots, n\}$. For a vector $v \in \mathbb{R}^n$ and an index set $K \subset [n]$, define a vector in $\mathbb{R}^n$, denoted by $v|_K$, such that $(v|_K)_i = v_i$ for all $i \in K$ and $(v|_K)_i = 0$ for all $i \notin K$.

Probability. For a random variable X and a probability distribution $\mathcal{D}$, we write $X \sim D$ for X being subject to the distribution $\mathcal{D}$. We denote the multivariate Gaussian with mean μ and covariance matrix Σ by $N(\mu, \Sigma)$. Let I_n denote the identity matrix of size $n \times n$.

We shall need the following lemma regarding concentration of Gaussian measure, see Chapter 1 of [LT91].

Lemma 1. *Suppose that $X \sim N(0, I_n)$ and the function $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz, i.e., $|f(x) - f(y)| \leq \|x - y\|_2$ for all $x, y \in \mathbb{R}^n$. Then for any $t > 0$ it holds that $\Pr_x\{|f(x) - \mathbb{E}f(x)| > t\} \leq 2e^{-t^2/2}$.*

Definition 1. *Suppose μ and ν are two probability measures over some Borel algebra $\mathcal{B}$ on $\mathbb{R}^n$. Then the total variation distance between μ and ν is defined as*

$$d_{TV}(\mu, \nu) = \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)| \left(= \frac{1}{2} \int_x |f(x) - g(x)| dx \right),$$

where the second equality holds when μ and ν have probability density functions $f(x)$ and $g(x)$ respectively.

The following is a result ([Das08]) that bounds the total variation distance between two multivariate Gaussian distributions.

Proposition 1. $d_{TV}(N(\mu_1, I_n), N(\mu_2, I_n)) \leq \|\mu_1 - \mu_2\|_2 / \sqrt{2}$.

Communication Model. We briefly summarize the notions from communication complexity that we will need. For more background on communication complexity, we refer the reader to [KN97]. In this paper we consider a one-way communication model. There are three players Alice, Bob and Charlie with private random coins. Alice is given an input x, Bob y and Charlie z, and their goal is to compute a function $f(x, y, z)$. Alice sends exactly one message to Bob and Bob sends exactly one message to Charlie, according to a protocol Π, and then Charlie outputs an answer. We say the protocol Π is δ-error if for every legal triple (x, y, z) of inputs, the answer equals $f(x, y, z)$ with probability at least $1 - \delta$, where the probability is taken over the random coins of the players. The concatenation of the message sent from Alice to Bob with the message from Bob to Charlie, as well as Charlie's output, is called the *transcript* of Π. The maximum length of the transcript (in bits) is called the *communication cost* of Π. The *communication complexity* of f is the minimal communication cost of a δ-error protocol for f, and is denoted $R_\delta(f)$.

Mutual Information. Let (X, Y) be a pair of discrete random variables with joint distribution $p(x, y)$. The mutual information $I(X; Y)$ is defined as $I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x,y)}{p(x)p(y)}$, where $p(x)$ and $p(y)$ are marginal distributions. The following are basic properties regarding mutual information.

Proposition 2. *Let X, Y, Z be discrete random variables defined on $\Omega_X, \Omega_Y, \Omega_Z$, respectively, and let f be a function defined on Ω. Then*

1. *$I(X; Y) \geq 0$ and the equality is attained iff X and Y are independent;*
2. *Chain rule for mutual information: $I(X, Y; Z) = I(X; Z) + I(X; Y|Z)$;*
3. *Data processing inequality: $I(f(X); Y) \leq I(X; Y)$.*

2.1 Direct-sum Technique

The following definitions and results are from [BYJKS04]. See also Section 6 of [BY02].

Definition 2. *Let Π be a randomized protocol with inputs belonging to a set $\mathcal{K}$. We shall abuse notation and also use $\Pi(X, Y, Z)$ to denote the transcript of protocol Π, which is a random variable which also depends on the private coins of the players. When X, Y, Z are understood from context, we sometimes further abbreviate $\Pi(X, Y, Z)$ as Π. Let μ be a distribution on $\mathcal{K}$ and suppose that $(X, Y, Z) \sim \mu$. The* information cost *of Π with respect to μ is defined to be $I(X, Y, Z; \Pi(X, Y, Z))$.*

Definition 3. *The δ-error information complexity of f with respect to a distribution μ, denoted by $IC_{\mu,\delta}(f)$, is defined to be the minimum information cost of a δ-error protocol for f with respect to μ.*

Using this definition, it follows immediately that (see [BYJKS04]):

Proposition 3. *$R_\delta(f) \geq IC_{\mu,\delta}(f)$ for any distribution μ and $\delta > 0$.*

Definition 4 (Conditional information cost)**.** *Let Π be a randomized protocol whose inputs belong to some set $\mathcal{K}$ of valid inputs and that ζ is a mixture of product distributions on $\mathcal{K} \times \mathcal{W}$. Suppose that $((X, Y, Z), W) \sim \zeta$. The conditional information cost of Π with respect to ζ is defined as $I(X, Y, Z; \Pi(X, Y, Z)|W)$.*

Definition 5 (Conditional information complexity)**.** *The δ-error conditional information complexity of f with respect to ζ, denoted $CIC_{\zeta,\delta}(f)$, is defined to be the minimum conditional information cost of a δ-error protocol for f with respect to ζ.*

Definition 6 (Decomposable functions)**.** *Suppose that f is a function defined on $\mathcal{L}^n$. We say that f is g-decomposable with primitive h if it can be written as $f(\mathbf{x}, \mathbf{y}, \mathbf{z}) = g(h(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1), \ldots, h(\mathbf{x}_n, \mathbf{y}_n, \mathbf{z}_n))$ for some function h defined on $\mathcal{L} \to \mathcal{Q}$ and g on $\mathcal{Q}^n$. Sometimes we simply say that f is decomposable with primitive h.*

Definition 7 (Embedding)**.** *For a vector $w \in \mathcal{L}^n$, $j \in [n]$ and $u \in \mathcal{L}$, we define $embed(w, j, u)$ to be the n-dimensional vector over $\mathcal{L}$ whose i-th component is defined as follows: $embed(w, j, u)_i = w_i$ if $i \neq j$; and $embed(w, j, u)_i = u$ if $i = j$.*

Definition 8 (Collapsing distribution)**.** *Suppose f is g-decomposable with primitive h. We call $(x, y, z) \in \mathcal{L}^n$ a collapsing input for f, if for every j and $(u, v, w) \in \mathcal{L}$, it holds that*

$$f(embed(x, j, u), embed(y, j, v), embed(z, j, w)) = h(u, v, w).$$

We call a distribution μ on $\mathcal{L}^n$ collapsing for f if every (x, y, z) in the support of μ is a collapsing input.

Lemma 2 (Information cost decomposition)**.** *Let Π be a protocol whose inputs belong to $\mathcal{L}^n$ for some set $\mathcal{L}$. Let ζ be a mixture of product distributions on $\mathcal{L} \times \mathcal{D}$ and suppose that $((X, Y, Z), D) \sim \zeta^n$. Then, $I(X, Y, Z; \Pi(X, Y, Z)|D) \geq \sum_i I(X_i, Y_i, Z_i; \Pi(X, Y, Z)|D)$.*

Lemma 3 (Reduction lemma)**.** *Let Π be a δ-error protocol for a decomposable function f defined on $\mathcal{L}^n$ with primitive h. Let ζ be a mixture of product distributions on $\mathcal{L} \times \mathcal{D}$, let $\eta = \zeta^n$, and suppose that $((X, Y, Z), D) \sim \eta$. If the distribution of (X, Y, Z) is a collapsing distribution for f, then for all $j \in [n]$, it holds that $I((X_j, Y_j, Z_j); \Pi(X, Y, Z)|D) \geq CIC_{\zeta,\delta}(h)$.*

2.2 Hellinger Distance

Definition 9. *The Hellinger distance $h(P, Q)$ between probability distributions P and Q on a domain Ω is defined by*

$$h^2(P, Q) = 1 - \sum_{\omega \in \Omega} \sqrt{P(\omega)Q(\omega)} = \frac{1}{2} \sum_{\omega \in \Omega} (\sqrt{P(\omega)} - \sqrt{Q(\omega)})^2.$$

One can verify that the Hellinger distance is a metric satisfying the triangle inequality, see, e.g., [BYJKS04]. The following proposition connects the Hellinger distance and the total variation distance.

Proposition 4. *(see, e.g., [BY02]) $h^2(P, Q) \leq d_{TV}(P, Q) \leq \sqrt{2}h(P, Q)$.*

In connection with mutual information, we have that

Lemma 4 ([BYJKS04])**.** *Let F_{z_1} and F_{z_2} be two random variables. Let Z denote a random variable with uniform distribution in $\{z_1, z_2\}$ Suppose $F(z)$ is independent of Z for each $z \in \{z_1, z_2\}$. Then, $I(Z; F(Z)) \geq h^2(F_{z_1}, F_{z_2})$.*

In [BY02], it is shown that a randomized private-coin three-party protocol exhibits the rectangle property in the following sense: there exist functions q_1, q_2, q_3 such that for all legal inputs (x, y, z) and transcripts τ, it holds that

$$\Pi_{x,y,z}(\tau) = q_1(x, \tau)q_2(y, \tau)q_3(z, \tau).$$

The following is a variant of the inverse triangle inequality in [BYJKS04] that we need to accommodate our setting of three players. The proof is similar to that for two players and thus we postpone it to Appendix A.

Lemma 5 (Inverse triangle inequality)**.** *For any randomized protocol Π and for any inputs x, y, z and x', y', z it holds that*

$$h^2(\Pi_{x,y,z}, \Pi_{x',y,z}) + h^2(\Pi_{x,y',z}, \Pi_{x',y',z}) \leq 2h^2(\Pi_{x,y,z}, \Pi_{x',y',z}).$$

3 Augmented L_∞ Promise Problem

In this section we define the Augmented L_∞ Promise Problem. First, though, we consider a slightly different gap problem than that considered in [BYJKS04] for a problem which we refer to as the L_∞ Promise problem.

Definition 10 ($L_\infty(k, \epsilon)$)**.** *Assume that $\epsilon k \geq 1$. There are three players Alice, Bob and Charlie in the one-way communication model with private coins. Alice receives a vector $\mathbf{a} \in \{0, \ldots, \epsilon k\}^n$, Bob a vector $\mathbf{b} \in \{0, \ldots, \epsilon k\}^n$ and Charlie both an index $j \in [n]$ and a bit $c \in \{0, 1\}$. The input is guaranteed to satisfy $|\mathbf{a}_i - \mathbf{b}_i| \leq 1$ for all $j \neq i$. Charlie is asked to decide which three of the following cases happen, provided we are promised that the input is indeed in one of them:*

(1) $|(\mathbf{a} - \mathbf{b})_j + ck| \leq \epsilon k$*;*

(2) $(\mathbf{a} - \mathbf{b})_j + ck \in \{k - 1, k, k + 1\}$*;*

(3) $(\mathbf{a} - \mathbf{b})_j + ck \in \{(1 - \epsilon)k, (1 + \epsilon)k\}$*;*

Charlie's output must be correct with probability $\geq 9/10$.

In the definition above, the index j is referred to as the *spike* position.

We consider the following distribution μ on the input. Let $c = 0$. Define the random variable $((X, Y), D)$ as follows. The random variable is uniform on $\{0, \ldots, k\} \times \{0, 1\} \setminus \{(0, 1), (k, 0)\}$. If $D = (d, 0)$ then $X = d$ and Y is uniform on $\{d, d + 1\}$; if $D = (d, 1)$ then $Y = d$ and X is uniformly distributed on $\{d - 1, d\}$.

Theorem 1. $R(L_\infty(k, \epsilon)) = \Omega(n/(k^2\epsilon^2))$.

Proof. Making $c = 0$ in Charlie's input, we see that μ^n is a collapsing distribution for $L_\infty(k, \epsilon)$ so we can apply the direct sum technique. Letting $\mathbf{x}_i = (\mathbf{a}_i, \mathbf{b}_i)$, it follows that

$$R(L_\infty(k, \epsilon)) \geq \sum_{i=1}^{n} I(\mathbf{x}_1, \ldots, \mathbf{x}_n; \Pi(\mathbf{x}_1, \ldots, \mathbf{x}_n)|D_1, \ldots, D_n) \geq nCIC_\mu(L^1_\infty(k, \epsilon)),$$

where $L^1_\infty(k, \epsilon)$ is the single coordinate problem of $L_\infty(k, \epsilon)$, that is, the $L_\infty(k, \epsilon)$ problem with $n = 1$. Therefore, it suffices to show that

$$CIC_\mu(L^1_\infty(k, \epsilon)) = \Omega\left(\frac{1}{k^2\epsilon^2}\right). \tag{1}$$

This is a single-coordinate problem, and we shall drop the index i henceforth in the proof. Let U_d denote a random variable with uniform distribution on $\{d, d + 1\}$.

$$CIC_\mu(L^1_\infty(k, \epsilon)) = I(\mathbf{x}; \Pi(\mathbf{x})|D) = \frac{1}{2\epsilon k}\left(\sum_{d=0}^{\epsilon k - 1} I(U_d; \Pi(d, U_d)) + \sum_{d=1}^{\epsilon k} I(U_{d-1}; \Pi(U_{d-1}, d))\right)$$

$$\geq \frac{1}{2\epsilon k}\left(\sum_{d=0}^{\epsilon k-1} h^2(\Pi_{d,d,0},\Pi_{d,d+1,0}) + \sum_{d=0}^{\epsilon k-1} h^2(\Pi_{d-1,d,0},\Pi_{d,d,0})\right) \tag{2}$$

$$\geq \frac{1}{4\epsilon^2 k^2}\left(\sum_{d=0}^{\epsilon k-1} h(\Pi_{d,d,0},\Pi_{d,d+1,0}) + \sum_{d=0}^{\epsilon k-1} h(\Pi_{d-1,d,0},\Pi_{d,d,0})\right)^2 \tag{3}$$

$$\geq \frac{1}{4\epsilon^2 k^2} h^2(\Pi_{0,0,0},\Pi_{\epsilon k,\epsilon k,0}) \tag{4}$$

where we used Lemma 4 for (2), the Cauchy-Schwarz inequality for (3) and the triangle inequality for (4), By the three-player version of the inverse triangle inequality (Lemma 5),

$$\begin{aligned} h^2(\Pi_{0,0,0},\Pi_{\epsilon k,\epsilon k,0}) &\geq \frac{1}{2}(h^2(\Pi_{0,0,0},\Pi_{\epsilon k,0,0}) + h^2(\Pi_{0,\epsilon k,0},\Pi_{\epsilon k,\epsilon k,0})) \\ &\geq \frac{1}{2} h^2(\Pi_{0,\epsilon k,0},\Pi_{\epsilon k,\epsilon k,0}) \\ &\geq \frac{1}{4} d_{TV}^2(\Pi_{0,\epsilon k,0},\Pi_{\epsilon k,\epsilon k,0}), \end{aligned} \tag{5}$$

where we used Proposition 4 for the last inequality. We now claim that

$$d_{TV}(\Pi_{0,\epsilon k,0},\Pi_{\epsilon k,\epsilon k,0}) = \Omega(1). \tag{6}$$

Consider the message sent from Alice to Bob, together with the message sent from Bob to Charlie. Let us denote the concatenation of these two messages by $T = T(x,y)$. Notice that the messages do not depend on Charlie's input. We in fact claim a stronger statement than (6), namely that $d_{TV}(T(0,\epsilon k), T(\epsilon k,\epsilon k)) = \Omega(1)$.

To see this, suppose that Charlie's input bit equals 1. Then he needs to decide if the players inputs are in case (3) or in case (4). Let $\mathcal{T}$ be the set of messages from Alice and Bob and from Bob to Charlie that make Charlie output "case (3)" with probability $\geq 3/4$, over his private coins. Then by the correctness of the protocol, $\Pr\{T(0,\epsilon k) \in \mathcal{T}\} \geq \frac{3}{5}$ and $\Pr\{T(\epsilon k,\epsilon k) \in \mathcal{T}\} \leq \frac{2}{15}$. Indeed, otherwise if $\Pr\{T(0,\epsilon k) \in \mathcal{T}\} < 3/5$ then Charlie outputs "case (3)" with probability $< 3/5 + 3/4 \cdot 2/5 = 9/10$, contradicting the correctness of the protocol, while if $\Pr\{T(\epsilon k,\epsilon k) \in \mathcal{T}\} > 2/15$ then Charlie outputs "case (3)" with probability $> 2/15 \cdot 3/4 = 1/10$, again contradicting the correctness of the protocol. Therefore

$$d_{TV}(T(0,\epsilon k), T(\epsilon k,\epsilon k)) \geq |\Pr(T(0,\epsilon k) \in \mathcal{T}) - \Pr(T(\epsilon k,\epsilon k) \in \mathcal{T})| \geq \frac{3}{5} - \frac{2}{15} = \Omega(1),$$

whence (6) follows since $d_{TV}(\Pi_{0,\epsilon k,0},\Pi_{\epsilon k,\epsilon k,0}) \geq d_{TV}(T(0,\epsilon k), T(\epsilon k,\epsilon k))$.

Plugging (6) into (5) and then (5) into (4), we have that (1) follows immediately. □

Now we define a stronger problem called the Augmented L_∞ Promise problem, and denoted by AUG-$L_\infty(r,k,\epsilon)$. We further abbreviate this by AUG-$L_\infty(r,k)$ when ϵ is clear from the context.

Definition 11 (AUG-$L_\infty(r,k,\epsilon)$). *Consider r instances of $L_\infty(k,\epsilon)$, denoted $(\mathbf{a}_1,\mathbf{b}_1,j_1,c_1),\ldots,(\mathbf{a}_r,\mathbf{b}_r,j_r,c_r)$. In addition to these inputs, Charlie has an index $I \in [r]$, together with $\mathbf{a}_j$ and $\mathbf{b}_j$ for all $j > I$. The goal is to decide for the I-th $L_\infty(k)$ instance, which of the three cases the input is in, with probability $\geq 5/8$.*

Now we define a distribution ν on the inputs to the AUG-$L_\infty(r,k)$ problem: the r instances of $L_\infty(k)$ are independent hard instances (i.e., drawn from μ) of $L_\infty(k)$. The index I is uniformly random on the set $[r]$.

Theorem 2. $R(L_\infty(r,k,\epsilon)) = \Omega(nr/(k^2\epsilon^2))$.

Proof. Write $\mathbf{x}_i = (\mathbf{a}_i,\mathbf{b}_i)$. It suffices to show that $I(\mathbf{x}_1,\ldots,\mathbf{x}_r;\Pi|\mathbf{Z}_1,\ldots,\mathbf{Z}_r) = \Omega\left(\frac{nr}{k^2\epsilon^2}\right)$, where $\mathbf{Z}_i = (\mathbf{D}_i,j_i,0)$ (letting $c_i = 0$ for all i). We claim that $I(\mathbf{x}_t;\Pi|\mathbf{Z}_t,\mathbf{Z}_{-t},\mathbf{x}_{>t}) \geq CIC_{\mu^n}(L_\infty(k,\epsilon))$. Indeed, the players can hardwire $\mathbf{x}_{>t}$ into the protocol, and Charlie can set $I = t$. Conditioned on $\mathbf{Z}_{-t}$, the inputs to the instances $\mathbf{x}_{<t}$ are independent, and so the players can generate these inputs using their private randomness. Then, for the input of $L_\infty(k, \epsilon)$, the players can embed it as the t-th input to the protocol for the AUG-$L_\infty(k)$ problem. It follows that the output of AUG-$L_\infty(k)$ agrees with the output of $L_\infty(k)$. Moreover, since the distribution on the t-th input instance is μ, we have that

$$I(\mathbf{x}_t; \Pi | \mathbf{Z}_t, \mathbf{Z}_{-t}, \mathbf{x}_{>t}) \geq CIC_{\mu^n}(L_\infty(k, \epsilon)) = \Omega\left(\frac{n}{\epsilon^2 k^2}\right)$$

by Theorem 1. It follows that

$$\begin{aligned} I(\mathbf{x}_1, \ldots, \mathbf{x}_r; \Pi | \mathbf{Z}_1, \ldots, \mathbf{Z}_r) &= \sum_t I(\mathbf{x}_t; \Pi | \mathbf{Z}_1, \ldots, \mathbf{Z}_r, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_r) \\ &= \sum_t \sum_{z,x} I(\mathbf{x}_t; \Pi | \mathbf{Z}_t, \mathbf{Z}_{-t} = z, \mathbf{x}_{>t} = x) \Pr\{\mathbf{Z}_{-t} = z, \mathbf{x}_{>t} = x\} \\ &\geq \sum_t \sum_{z,x} \Omega\left(\frac{n}{\epsilon^2 k^2}\right) \Pr\{\mathbf{Z}_{-t} = z, \mathbf{x}_{>t} = x\} \\ &= \Omega\left(\frac{nr}{\epsilon^2 k^2}\right) \end{aligned}$$

as desired. □

4 Frequency Moments

Suppose that $x \in \mathbb{R}^n$. We say that a data stream algorithm solves the (ϵ, p)-NORM problem if its output X satisfies $(1-\epsilon)\|x\|_p^p \leq X \leq (1+\epsilon)\|x\|_p^p$ with probability $\geq 1 - \delta$. Our main result is the following.

Theorem 3. *For any constant $p > 2$, there exist absolute constants $c > 0$, $\alpha > 1$ and a constant $\epsilon_0 = \epsilon_0(p)$ which depends only on p such that for any $\epsilon \in [c/n^{1/p}, \epsilon_0]$, any randomized streaming algorithm that solves the (ϵ, p)-NORM problem for $x \in \{-M, -M+1, \ldots, M\}^n$ with probability $\geq 19/20$ and n sufficiently large, where $M = \Omega(n^{\alpha/p})$, requires $\Omega(n^{1-2/p}(\log M)/\epsilon^2)$ bits of space.*

Proof. Suppose that a randomized streaming algorithm $\mathcal{A}$ solves the (ϵ, p)-NORM problem with probability $\geq 19/20$. Let $k = \Theta(n^{1/p})$ and $r = (1 - 1/\alpha)\log_{10} M$. We shall reduce the (ϵ, p)-NORM problem to AUG-$L_\infty(r, k, \epsilon)$. Note that with our choice of parameters, $\epsilon k = \Omega(1)$.

Alice generates a stream σ_1 with underlying frequency vector $-\sum_j 10^{j-1} A^j$ and sends the state of $\mathcal{A}$ on σ_1 to Bob. Then Bob generates a stream σ_2 with underlying frequency vector $\sum_j 10^{j-1} B^j$ and continues running $\mathcal{A}$ on σ_2, starting from the state sent by Alice. The streaming algorithm then reaches a state corresponding to an underlying frequency vector $\sum_j 10^{j-1}(B^j - A^j)$. Bob sends this state to Charlie. Charlie, given I and (A^j, B^j) for all $j > I$, generates a stream σ_3 with underlying frequency vector $\sum_{j>I} 10^{j-1}(A^j - B^j)$ and continues running $\mathcal{A}$ on σ_3 to obtain an output V for the execution of $\mathcal{A}$ on a stream with underlying frequency vector $v = \sum_{j=1}^{I} 10^{j-1}(B^j - A^j)$. Finally, Charlie generates a stream σ_4 with underlying frequency vector $10^{I-1} c_I k \mathbf{e}_{j_I}$, where $j_I \in [n]$ and $c_I \in \{0, 1\}$ are the inputs to Charlie in the I-th instance of the $L_\infty(k)$ Promise problem, and continues running $\mathcal{A}$ on σ_4 to obtain an output W for a stream with underlying frequency vector equal to $w = v + 10^{I-1} c_I k \mathbf{e}_{j_I}$.

For notational convenience, let $X^j = B^j - A^j$, then $\|X^j\|_\infty \leq \epsilon k$. Notice that $\|v\|_\infty \leq \epsilon k \sum_{j=1}^{I} 10^{j-1} \leq \epsilon k \cdot 10^I/9 \leq \epsilon k \cdot 10^r/9 \leq \epsilon k M^{1-1/\alpha} \leq M$ and $\|w\|_\infty \leq \|v\|_\infty + k 10^{I-1} \leq k(1+\epsilon)10^r/9 \leq M$ by our assumption of M and choice of k and r. This implies that $\mathcal{A}$ outputs a correct approximation to $\|w\|_\infty$ with probability $\geq 19/20$ and a correct approximation to $\|v\|_\infty$ with probability $\geq 19/20$. Define the event

$$\mathcal{E}_1 = \left\{ \left|W - \|w\|_p^p\right| \leq \epsilon \|w\|_p^p \text{ and } \left|V - \|v\|_p^p\right| \leq \epsilon \|v\|_p^p \right\}.$$

By a union bound, $\Pr\{\mathcal{E}_1\} \geq 9/10$. We show next how to use W and V to solve AUG-$L_\infty(k, r, \epsilon)$, conditioned on the event $\mathcal{E}_1$. Charlie needs only to distinguish case (2) from case (3) in Definition 10, since if $c_I = 0$ he can output directly that it is case (1). Henceforth we assume that $c_I = 1$.

Let $L = \left(\sum_{j=1}^{I} 10^{j-1}\right)^p$, then $L = \left(\frac{10^I-1}{9}\right)^p \leq \left(\frac{10}{9}\right)^p 10^{(I-1)p}$ so $nL/(10^{I-1}k)^p \leq \beta^p$ as $n \to \infty$ if we choose $k \geq 10/(9\beta) \cdot n^{1/p}$. The value of β is a constant (depending only on p) which will be determined later. We also assume that $\epsilon \leq 2\beta$. Furthermore, for notational convenience we define

$$\begin{aligned} UB(x) &= (1+\epsilon)(1+x)^p - (1-\epsilon)|x|^p + 4\epsilon\beta^p \\ LB(x) &= (1-\epsilon)(1+x)^p - (1+\epsilon)|x|^p - 4\epsilon\beta^p. \end{aligned}$$

Now we are ready for the analysis.

Case (2). $|X^I_{j_I}| \leq 1$. Note that $\|\sum_j 10^{j-1}X^j\|_\infty \leq \epsilon k \sum_{i=1}^{I-1} 10^{i-1} = \epsilon k 10^{I-1}/9$. Then

$$\begin{aligned} W - V &\leq (1+\epsilon)\|w\|_p^p - (1-\epsilon)\|v\|_p^p \\ &= (1+\epsilon)(\|v|_{[n]\setminus\{j_I\}}\|_p^p + |w_{j_I}|_p^p) - (1-\epsilon)\|v\|_p^p \\ &= 2\epsilon\|v|_{[n]\setminus\{j_I\}}\|_p^p + (1+\epsilon)|w_{j_I}|^p - (1-\epsilon)|v_{j_I}|^p. \end{aligned}$$

Observe that $\|v|_{[n]\setminus\{j_I\}}\|_p^p$ is maximized when

(i) $j_1 = \cdots = j_{I-1} \neq j_I$;

(ii) $A^j_{j_1} - B^j_{j_1} = \epsilon k$ for $j < I$;

(iii) $A^I_{j_1} - B^I_{j_1} = 1$;

(iv) $A^t_i - B^t_i = 1$ for all $i \neq j_t$ and $t \leq I$,

so

$$\begin{aligned} \|v|_{[n]\setminus\{j_I\}}\|_p^p &\leq (n-2)L + \left(\sum_{j=1}^{I-1} 10^{j-1}\epsilon k + 10^{I-1}\right)^p \\ &\leq (n-2)L + 10^{(I-1)p}\left(\frac{\epsilon k}{9} + 1\right)^p \\ &\leq (n-2)L + 10^{(I-1)p}\left(\frac{\epsilon k}{2}\right)^p \\ &\leq 2\beta^p \cdot 10^{(I-1)p}k^p \end{aligned}$$

where the last inequality uses the assumption that $\epsilon k = \Omega(1)$. It follows that

$$\begin{aligned} W - V &\leq 2\epsilon \cdot 2\beta^p \cdot 10^{(I-1)p}k^p + 10^{(I-1)p}k^p\left\{(1+\epsilon).(1+\eta_1)^p - (1-\epsilon)\eta_1^p\right\} \\ &= 10^{(I-1)p}k^p\left\{(1+\epsilon)(1+\eta_1)^p - (1-\epsilon)|\eta_1|^p + 4\epsilon\beta^p\right\} \\ &= 10^{(I-1)p}k^p \cdot UB(\eta_1), \end{aligned}$$

where

$$\eta_1 = \frac{v_{j_I}}{10^{I-1}k}.$$

In this case we have that

$$|\eta_1| \leq \frac{\epsilon k \frac{10^{I-1}}{9} + 10^{I-1}}{10^{I-1}k} \leq \frac{\epsilon}{9} + o(1), \quad n \to \infty. \tag{7}$$

Similarly

$$\begin{aligned} W - V &\geq 10^{(I-1)p} k^p \left\{ (1-\epsilon)(1+\eta_1)^p - (1+\epsilon)|\eta_1|^p - 4\epsilon\beta^p \right\} \\ &= 10^{(I-1)p} k^p \cdot LB(\eta_1), \end{aligned}$$

Case (3)a. $X^I_{j_I} = -\epsilon k$. Similarly to the above

$$W - V \leq 10^{(I-1)p} k^p \cdot UB(\eta_2),$$

but now

$$\eta_2 = \frac{v_{j_I}}{10^{I-1}k} = \frac{-10^{I-1}\epsilon k + S}{10^{I-1}k}$$

for some $|S| \leq \epsilon k 10^{I-1}/9$. Thus

$$-\frac{10}{9}\epsilon + o(1) \leq \eta_2 \leq -\frac{8}{9}\epsilon + o(1), \quad n \to \infty. \tag{8}$$

Case (3)b. $X^I_{j_I} = \epsilon k$. Similarly to the above

$$W - V \geq 10^{(I-1)p} k^p \cdot LB(\eta_3),$$

with

$$\frac{8}{9}\epsilon + o(1) \leq \eta_3 \leq \frac{10}{9}\epsilon + o(1), \quad n \to \infty. \tag{9}$$

Charlie can distinguish case (2) from case(3) if it holds that

$$LB(\eta_1) > UB(\eta_2) \tag{10}$$

$$UB(\eta_1) < LB(\eta_3). \tag{11}$$

Note that $UB(x)$ is increasing around $x = 0$ and $LB(x)$ is decreasing around $x = 0$, it suffices to have that

$$\lim_{\eta \to -\frac{1}{9}\epsilon} LB(\eta) > \lim_{\eta \to -\frac{8}{9}\epsilon} UB(\eta) \tag{12}$$

$$\lim_{\eta \to \frac{1}{9}\epsilon} UB(\eta) < \lim_{\eta \to \frac{8}{9}\epsilon} LB(\eta), \tag{13}$$

then there would exist some constant δ such that whenever the $o(1)$ in (7), (8) and (9) are all at most δ (which holds when n is sufficiently large), the inequalities (10) and (11) are satisfied.

Now, the conditions (12) and (13) are equivalent to

$$(1-\epsilon)(1+\frac{1}{9}\epsilon)^p - (1+\epsilon)(\frac{1}{9}\epsilon)^p - 4\epsilon\beta^p > (1+\epsilon)(1-\frac{8}{9}\epsilon)^p - (1-\epsilon)(\frac{8}{9}\epsilon)^p + 4\epsilon\beta^p \tag{14}$$

$$(1+\epsilon)(1-\frac{1}{9}\epsilon)^p - (1-\epsilon)(\frac{1}{9}\epsilon)^p + 4\epsilon\beta^p < (1-\epsilon)(1+\frac{8}{9}\epsilon)^p - (1+\epsilon)(\frac{8}{9}\epsilon)^p - 4\epsilon\beta^p \tag{15}$$

respectively. We claim that both are satisfied when ϵ is small enough and $p > 2$ is a constant. For (14) to hold for small ϵ it suffices to have that

$$\liminf_{\epsilon \to 0^+} \frac{(1-\epsilon)(1+\frac{1}{9}\epsilon)^p - (1+\epsilon)(\frac{1}{9}\epsilon)^p - \left\{(1+\epsilon)(1-\frac{8}{9}\epsilon)^p - (1-\epsilon)(\frac{8}{9}\epsilon)^p\right\}}{\epsilon} > 8\beta^p.$$

Evaluating the limit on the left-hand side gives that[1]

$$p - 2 > 8\beta^p,$$

which is satisfied by choosing a small β depending on p. One can show similarly that (15) is also satisfied when ϵ is small.

In conclusion, Charlie can solve the AUG-$L_\infty(r, k, \epsilon)$ problem with probability $\geq 9/10$. The lower bound for the (ϵ, p)-NORM problem follows from Theorem 2. $\square$

5 Lower Bound for Linear Sketches

Given $\eta \geq 0$, define a distribution $\mathcal{D}_{k,\eta}$ on $\mathbb{R}^n$ as follows. Consider $x \sim N(0, I_n)$. Let j be uniformly random in $\{1, \ldots, n\}$. The distribution $\mathcal{D}_{k,\eta}$ is defined to be $\mathcal{L}(x + (1 + \eta)ke_j)$. Suppose that A is an $m \times n$ matrix of orthonormal rows. When operated on vectors $x \sim \mathcal{D}_{k,\eta}$, the product Ax induces a distribution, denoted by $\mathcal{F}_{A,k,\eta}$.

Lemma 6. *Let $\epsilon > 0$. It holds that $d_{TV}(\mathcal{F}_{A,k,0}, \mathcal{F}_{A,k,\epsilon}) \leq \epsilon k\sqrt{m/n}$.*

Proof. Let $y_1 \sim F_{A,k,0}$ and $y_2 \sim F_{A,k,\epsilon}$. By rotational invariance of the Gaussian distribution and the fact that A has orthonormal rows, y_1 is distributed as $x + kA_j$ and y_2 as $x + (1 + \epsilon)kA_j$, where $x \sim N(0, I_m)$, A_j is the j-th column of A, and j is uniform on $\{1, \ldots, n\}$.

Suppose the density functions of y_1 and y_2 are $p_1(x)$ and $p_2(x)$ respectively, then

$$p_1(x) = \frac{1}{n}\sum_i p(x - kA_i), \quad p_2(x) = \frac{1}{n}\sum_i p(x - (1 + \epsilon)kA_i),$$

where $p(x)$ is the density function of $N(0, I_m)$. It follows that

$$\begin{aligned}
d_{TV}(\mathcal{F}_{A,k,0}, \mathcal{F}_{A,k,\epsilon}) &= \frac{1}{2}\int_{x\in\mathbb{R}^m} |p_1(x) - p_2(x)|dx = \frac{1}{2}\int_{x\in\mathbb{R}^m} \left|\frac{1}{n}\sum_i p(x - kA_i) - \frac{1}{n}\sum_i p(x - (1 + \epsilon)kA_i)\right| dx \\
&\leq \frac{1}{2}\int_{x\in\mathbb{R}^m} \left(\frac{1}{n}\sum_i |p(x - kA_i) - p(x - (1 + \epsilon)kA_i|\right) dx \\
&= \frac{1}{n}\sum_i \frac{1}{2}\int_{x\in\mathbb{R}^m} |p(x - kA_i) - p(x - (1 + \epsilon)kA_i)|dx \\
&= \frac{1}{n}\sum_i d_{TV}(N(kA_i, I_n) - N((1 + \epsilon)kA_i, I_n)) \\
&\leq \frac{1}{n}\sum_i \|kA_i - (1 + \epsilon)kA_i\|_2 \qquad \text{(by Proposition 1)} \\
&= \frac{\epsilon k}{n}\sum_i \|A_i\|_2 = \epsilon k\mathbb{E}_j\|A_j\|_2,
\end{aligned}$$

Since $\sum_j \|A_j\|_2^2 = m$, $\mathbb{E}_j\|A_j\|_2^2 = m/n$ and thus $\mathbb{E}\|A_j\|_2 \leq (\mathbb{E}\|A_j\|_2^2)^{1/2} = \sqrt{m/n}$. It follows that $d_{TV}(\mathcal{F}_{A,k,0}, \mathcal{F}_{A,k,\eta}) \leq \epsilon k\sqrt{\frac{m}{n}}$. $\square$

Theorem 4. *Let $p > 2$ be a constant. Consider a distribution over $m \times n$ matrices A for which for every $x \in \mathbb{R}^n$, from Ax one can solve the (ϵ, p)-NORM problem, on input x with probability $\geq 3/4$ over the choice of A, where $\epsilon = \Omega(1/n^{1/p})$ is small enough. Then $m = \Omega(n^{1-2/p}/\epsilon^{-2})$.*

[1]The constant 2 in the formula below could be made arbitrarily small by adjusting the approximation ratio from $1 + \epsilon$ to $1 + \epsilon/10$, $1 + \epsilon/100$, etc.

Proof. W.l.o.g., A has orthonormal rows, since we can apply a change of basis to the vector space spanned by the rows of A in post-processing. Let $k = C_p^{1/p} n^{1/p}$, where C_p is the constant in

$$\mathbb{E}\|z\|_p^p = C_p n, \quad z \sim N(0, I_n).$$

Consider the input x drawn from $\mathcal{D}_0 := \mathcal{D}_{k,0}$ and $\mathcal{D}_1 := \mathcal{D}_{k,2\epsilon}$. Let $b \in \{0,1\}$ indicate that $x \sim \mathcal{D}_b$. We have that $Ax \sim \mathcal{F}_{A,k,0}$ when $b = 0$ and $Ax \sim \mathcal{F}_{A,k,2\epsilon}$ when $b = 1$. Suppose the algorithm outputs W.

Now we compute $\|x\|_p^p$ in each case. When $b = 0$, $\|x\|_p^p = \|x'\|_p^p + |g + k|^p$, where $x' \sim N(0, I_{n-1})$ and $g \sim N(0,1)$ are independent. Since $\|x\|_p$ is a 1-Lipschitz function, by concentration of measure (Lemma 1),

$$\Pr\{\left|\|x'\|_p - \mathbb{E}\|x'\|_p\right| \geq 5\} \leq 0.001.$$

Also, $|g| \leq 5$ with probability $\geq 1 - 0.001$. Note that $\mathbb{E}\|x'\|_p^p = C_p(n-1)$. It follows that with probability $\geq 1 - 0.002$, we have

$$\|x\|_p^p \leq 2(1 + o(1))C_p n. \tag{16}$$

Similarly, when $b = 1$, with probability $\geq 1 - 0.002$, it holds that

$$\|x\|_p^p \geq ((1 + 2\epsilon)^p + 1)(1 - o(1))C_p n. \tag{17}$$

The $o(1)$ in (16) and (17) are of the form $c_p/n^{1/p}$ for some (small) constant $c_p > 0$ that depends only on p. We condition on the event that (16) and (17) hold. With probability $\geq 3/4$, we have

$$\begin{aligned} W &\leq (1 + \epsilon)\|x\|_p^p, \quad b = 0 \\ W &\geq (1 - \epsilon)\|x\|_p^p, \quad b = 1 \end{aligned}$$

and thus

$$\begin{aligned} W &\leq 2(1 + \epsilon)(1 + o(1))C_p n, \quad b = 0 \\ W &\geq (1 - \epsilon)((1 + 2\epsilon)^p + 1)(1 - o(1))C_p n, \quad b = 1 \end{aligned}$$

So we can recover b from W with probability $\geq 3/4 - 0.002$ provided that (see the full derivation in Appendix C)

$$\begin{aligned} & 2(1 + \epsilon)(1 + o(1)) < (1 - \epsilon)((1 + 2\epsilon)^p + 1)(1 - o(1)) \\ \Longleftarrow & \left(2 + \frac{p+2}{4}\right)\frac{c_p}{n^{1/p}} < \frac{p-2}{2}\epsilon \qquad \text{(recall that } o(1) \text{ is actually } c_p/n^{1/p}) \end{aligned}$$

which holds for ϵ small enough while satisfying that $\epsilon = \Omega(1/n^{1/p})$. Consider the event $\mathcal{E}$ that the algorithm's output indicates $b = 1$. Then $\Pr(\mathcal{E}|x \sim \mathcal{D}_0) \leq 1/4 + 0.002$ while $\Pr(\mathcal{E}|x \sim \mathcal{D}_1) \geq 3/4 - 0.002$. By definition of total variation distance,

$$d_{TV}(\mathcal{F}_{A,k,0}, \mathcal{F}_{A,k,2\epsilon}) \geq |\Pr(\mathcal{E}|x \sim \mathcal{D}_1) - \Pr(\mathcal{E}|x \sim \mathcal{D}_0)| \geq \frac{1}{2} + 0.004.$$

On the other hand, by the preceding lemma, $d_{TV}(\mathcal{F}_{A,k,0}, \mathcal{F}_{A,k,2\epsilon}) \leq 2\epsilon k\sqrt{\frac{m}{n}}$. Therefore it must hold that $m = \Omega\left(\frac{n}{k^2\epsilon^2}\right) = \Omega\left(\frac{n^{1-2/p}}{\epsilon^2}\right)$. □

References

[AKO11] A. Andoni, R. Krauthgamer, and K. Onak. Streaming algorithms via precision sampling. In *FOCS*, pages 363–372, 2011.

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. *JCSS*, 58(1):137–147, 1999.

[And] A. Andoni. High frequency moment via max stability. Available at http://web.mit.edu/andoni/www/papers/fkStable.pdf.

[ANPW13] A. Andoni, H. L. Nguyen, Y. Polyanskiy, and Y. Wu. Tight lower bound for linear sketches of moments. In *ICALP*, 2013.

[BGKS06] L. Bhuvanagiri, S. Ganguly, D. Kesh, and C. Saha. Simpler algorithm for estimating frequency moments of data streams. In *SODA*, pages 708–713, 2006.

[BO10] V. Braverman and R. Ostrovsky. Recursive sketching for frequency moments. *CoRR*, abs/1011.2571, 2010.

[BO12] V. Braverman and R. Ostrovsky. Approximating large frequency moments with pick-and-drop sampling. *CoRR*, abs/1212.0202, 2012.

[BY02] Z. Bar-Yossef. *The Complexity of Massive Data Set Computations*. PhD thesis, University of California, Berkeley, 2002.

[BYJK$^+$02] Z. Bar-Yossef, T. S. Jayram, R. Kumar, D. Sivakumar, and L. Trevisan. Counting distinct elements in a data stream. In *RANDOM*, pages 1–10, 2002.

[BYJKS04] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[CCFC02] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. In *ICALP*, pages 693–703, 2002.

[CCM10] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for estimating the entropy of a stream. *ACM Transactions on Algorithms*, 6(3), 2010.

[CDM06] A. Chakrabarti, K. Do Ba, and S. Muthukrishnan. Estimating Entropy and Entropy Norm on Data Streams. In *STACS*, pages 196–205, 2006.

[CK04] D. Coppersmith and R. Kumar. An improved data stream algorithm for frequency moments. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 151–156, 2004.

[CKS03] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *CCC*, pages 107–117, 2003.

[CM05a] G. Cormode and S. Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *J. Algorithms*, 55(1):58–75, 2005.

[CM05b] G. Cormode and S. Muthukrishnan. Space efficient mining of multigraph streams. In *PODS*, pages 271–282, 2005.

[CSWY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. C.-C. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *FOCS*, pages 270–278, 2001.

[CW09] K. L. Clarkson and D. P. Woodruff. Numerical linear algebra in the streaming model. In *STOC*, pages 205–214, 2009.

[Das08] A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer, 2008.

[FM83] P. Flajolet and G. N. Martin. Probabilistic counting. In *Proceedings of the 24th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 76–82, 1983.

[Gan04a] S. Ganguly. Estimating frequency moments of data streams using random linear combinations. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM)*, pages 369–380, 2004.

[Gan04b] S. Ganguly. A hybrid algorithm for estimating frequency moments of data streams, 2004. Manuscript.

[Gan08] S. Ganguly. Lower bounds on frequency estimation of data streams (extended abstract). In *CSR*, pages 204–215, 2008.

[Gan09] S. Ganguly. Deterministically estimating data stream frequencies. In *COCOA*, pages 301–312, 2009.

[Gan11] S. Ganguly. Polynomial estimators for high frequency moments. *CoRR*, abs/1104.4552, 2011.

[Gan12] S. Ganguly. A lower bound for estimating high moments of a data stream. *CoRR*, abs/1201.0253, 2012.

[GC07] S. Ganguly and G. Cormode. On estimating frequency moments of data streams. In *APPROX-RANDOM*, pages 479–493, 2007.

[Ind07] P. Indyk. Sketching, streaming and sublinear-space algorithms, 2007. Graduate course notes available at `http://stellar.mit.edu/S/course/6/fa07/6.895/`.

[IW03] P. Indyk and D. P. Woodruff. Tight lower bounds for the distinct elements problem. In *FOCS*, pages 283–288, 2003.

[IW05] P. Indyk and D. P. Woodruff. Optimal approximations of the frequency moments of data streams. In *STOC*, pages 202–208, 2005.

[Jay09] T. S. Jayram. Hellinger strikes back: A note on the multi-party information complexity of and. In *APPROX-RANDOM*, pages 562–573, 2009.

[JW09] T. S. Jayram and D. P. Woodruff. The data stream space complexity of cascaded norms. In *FOCS*, pages 765–774, 2009.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KNPW11] D. M. Kane, J. Nelson, E. Porat, and D. P. Woodruff. Fast moment estimation in data streams in optimal space. In *STOC*, pages 745–754, 2011.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[Mut05] S. Muthukrishnan. Data Streams: Algorithms and Applications. *Foundations and Trends in Theoretical Computer Science*, 1(2):117–236, 2005.

[MW10] M. Monemizadeh and D. P. Woodruff. 1-pass relative-error l_p-sampling with applications. In *SODA*, 2010.

[PT07] A. Pavan and S. Tirthapura. Range-efficient counting of distinct elements in a massive data stream. *SIAM J. Comput.*, 37(2):359–379, 2007.

[PW12] E. Price and D. P. Woodruff. Applications of the shannon-hartley theorem to data streams and sparse recovery. In *ISIT*, pages 2446–2450, 2012.

[Woo04] D. P. Woodruff. Optimal space lower bounds for all frequency moments. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 167–175, 2004.

[WZ12] D. P. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 941–960, 2012.

A Proof of Lemma 5

Proof. Using the rectangle property and the arithmetic-geometric mean inequality, we have that

$$\begin{aligned}
h^2(\Pi_{x,y,z},\Pi_{x',y,z}) + h^2(\Pi_{x,y',z},\Pi_{x',y',z}) &= 2 - \sum_\tau \sqrt{\Pi_{x,y,z}(\tau)\Pi_{x',y,z}(\tau)} - \sum_\tau \sqrt{\Pi_{x,y',z}(\tau)\Pi_{x',y',z}(\tau)} \\
&= 2 - \sum_\tau \sqrt{q_1(x,\tau)q_2(y,\tau)q_3(z,\tau)q_1(x',\tau)q_2(y,\tau)q_3(z,\tau)} \\
&\quad - \sum_\tau \sqrt{q_1(x,\tau)q_2(y',\tau)q_3(z,\tau)q_1(x',\tau)q_2(y',\tau)q_3(z,\tau)} \\
&= 2 - \sum_\tau \sqrt{q_1(x,\tau)q_1(x',\tau)}q_3(z,\tau)(q_2(y,\tau)+q_2(y',\tau)) \\
&\le 2 - 2\sum_\tau \sqrt{q_1(x,\tau)q_1(x',\tau)}q_3(z,\tau)\sqrt{q_2(y,\tau)q_2(y',\tau)} \\
&= 2 - 2\sum_\tau \sqrt{\Pi_{x,y,z}(\tau)\Pi_{x',y',z}(\tau)} \\
&= h^2(\Pi_{x,y,z},\Pi_{x',y',z}).
\end{aligned}$$

□

B Omitted details in the proof of Theorem 3

Writing the inequalities in full,

$$\begin{aligned}
& LB_2 > UB_1 \\
\iff & (1-\epsilon)10^{(I-1)p}(1-\epsilon)^p k^p - 2\epsilon(n-1)L - (1+\epsilon)\epsilon^p 10^{(I-1)p}k^p > 2\epsilon nL \\
\iff & (1-\epsilon)10^{(I-1)p}(1-\epsilon)^p k^p - (1+\epsilon)\epsilon^p 10^{(I-1)p}k^p > 2\epsilon nL + 2\epsilon(n-1)L \\
\Longleftarrow & \left((1-\epsilon)(1-\epsilon)^p - (1+\epsilon)\epsilon^p\right) 10^{(I-1)p}k^p > 4\epsilon nL \\
\iff & (1-\epsilon)^{p+1} - (1+\epsilon)\epsilon^p > 4\epsilon \frac{n}{k^p}\left(\frac{10}{9}\right)^p \\
\Longleftarrow & (1-\epsilon)^{p+1} - (1+\epsilon)\epsilon^p > 4\epsilon \frac{n}{k^p}\left(\frac{10}{9}\right)^p.
\end{aligned}$$

Let $p' = p/2+1$, then $2 < p' < p$. Note that $(1-x)^p = 1 - px + o(x^2)$ as $x \to 0^+$, we see that

$$(1-\epsilon)^p \le 1 - p'\epsilon \tag{18}$$

for ϵ small enough. Now,

$$\begin{aligned}
& LB_3 > UB_2 \\
\iff & (1-\epsilon)10^{(I-1)p}k^p - 2\epsilon(n-1)L - (1+\epsilon)\epsilon^p 10^{(I-1)p}k^p \\
& \qquad > 2\epsilon(n-1)L + (1+\epsilon)10^{(I-1)p}(1-\epsilon)^p k^p + (1-\epsilon)\epsilon^p k^p 10^{(I-1)p} \\
\iff & \left(1-\epsilon-(1+\epsilon)(1-\epsilon)^p - (1+\epsilon)\epsilon^p - (1-\epsilon)\epsilon^p\right) 10^{(I-1)p}k^p \\
& \qquad > 4\epsilon(n-1)L \\
\Longleftarrow & 1-\epsilon-(1+\epsilon)(1-\epsilon)^p - 2\epsilon^p > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p \\
\Longleftarrow & 1-\epsilon-(1+\epsilon)(1-p'\epsilon) - 2\epsilon^p > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p \quad \text{(invoking (18))}
\end{aligned}$$

$$\iff (p'-2)\epsilon + p\epsilon^2 - 2\epsilon^p > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p$$
$$\Longleftarrow \epsilon(p'-2-2\epsilon^{p-1}) > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p$$
$$\iff \epsilon\left(\frac{p-2}{2} - 2\epsilon^{p-1}\right) > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p$$

Therefore, Charlie can solve AUG-$L_\infty(r,k,\epsilon)$ provided that

$$(1-\epsilon)^{p+1} - (1+\epsilon)\epsilon^p > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p$$
$$\epsilon\left(\frac{p-2}{2} - 2\epsilon^{p-1}\right) > 4\epsilon\frac{n}{k^p}\left(\frac{10}{9}\right)^p.$$

C Omitted details in the proof of Theorem 4

$$2(1+\epsilon)(1+o(1)) < (1-\epsilon)((1+2\epsilon)^p+1)(1-o(1))$$
$$\Longleftarrow 2(1+\epsilon)(1+o(1)) < (1-\epsilon)(1+2p\epsilon+1)(1-o(1)) \quad \text{(since } (1+x)^p \geq 1+px\text{)}$$
$$\iff 2(1+\epsilon)(1+o(1)) < 2(1-\epsilon)(1+p\epsilon)(1-o(1))$$
$$\Longleftarrow (1+\epsilon)(1+o(1)) < \left(1+\frac{p}{2}\epsilon\right)(1-o(1)) \qquad \left(p\epsilon^2 \leq (\frac{p}{2}-1)\epsilon \text{ when } \epsilon \leq \frac{1}{2}-\frac{1}{p}\right)$$
$$\Longleftarrow \left(2+\frac{p+2}{2}\epsilon\right)o(1) < \frac{p-2}{2}\epsilon$$
$$\Longleftarrow \left(2+\frac{p+2}{4}\right)\frac{c_p}{n^{1/p}} < \frac{p-2}{2}\epsilon \qquad \text{(recall that } o(1) \text{ is actually } c_p/n^{1/p} \text{ and } \epsilon \leq 1/2\text{)}$$

D Application to Cascaded Moments

We sketch our improvement to estimating $\ell_p(\ell_q)(A)$ for $p, q \geq 2$, as outlined in the Introduction. In the t-player set disjointness problem, there are t players, holding subsets $S^1, \ldots, S^t \subseteq [N]$ respectively. They are promised that either: (1) for all $i \in [N]$, there is at most one $j \in [t]$ for which $i \in S^j$, or (2) there is a unique $i \in [N]$ for which $i \in S^j$ for all $j \in [t]$, and for all $i' \neq i$, there is at most one $j \in [t]$ for which $i' \in S^j$. In a 1-way protocol the k-th player needs to output which of the two cases the input is in. If we define the N-dimensional vector x so that x_i is the number of $j \in [t]$ for which $i \in S^j$, then in case (1) we have $\|x\|_\infty \leq 1$, while in case (2) we have that there is a unique $i \in [N]$ for which $x_i = t$ and for all $j \neq i$ we have $x_j \in \{0,1\}$.

Let μ^N be the input distribution which for each $i \in [N]$, chooses a random player $D_i \in [t]$, and with probability $1/2$ we have $i \in S^{D_i}$ while with probability $1/2$ we have $i \notin S^{D_i}$. For all $j \neq D_i$, it holds that $i \notin S^j$. Then μ^N is a collapsing distribution.

It is known [**?**, CKS03] that for any δ-error 1-way randomized protocol Π, and inputs distributed according to μ^N, that

$$I(S^1, \ldots, S^t; \Pi(S^1, \ldots, S^t) | D_1, \ldots, D_N) \geq N \cdot CIC_\mu(AND_t),$$

where AND_t is a single-coordinate problem in which the t players each have a single bit, they are promised that either all of their bits equal 1, or there is at most a single bit which is equal to 1, and they need to decide which case they are in.

As in Section 3, we introduce a $(t+1)$-st player Charlie who holds a bit $c \in \{0,1\}$ and an index $j \in [N] = [n] \times [d]$. In the distribution μ^N, Charlie's input c is always set to 0, so that μ^N is collapsing. For our application to cascaded moments of $n \times d$ matrices A, we fix $N = nd$ and $t = 2\varepsilon n^{1/p} d^{1/q}$. The problem is to decide whether (1) $x_j + c2n^{1/p}d^{1/q} \leq 1$, (2) $x_j + c2n^{1/p}d^{1/q} \in \{2n^{1/p}d^{1/q}, 2n^{1/p}d^{1/q} + 1\}$, or (3) $x_j + c2n^{1/p}d^{1/q} \geq 2(1+\varepsilon)n^{1/p}d^{1/q}$.

We can use the same derivation as in equation (3) of [Jay09] to lower bound mutual information by Hellinger distance, provided we fix Charlie's input bit c to 0 throughout the derivation. This results in the derivation:

$$CIC_\mu(AND_t) \geq \frac{1}{t} h^2(\Pi_{0^{t+1}}, \Pi_{1^t 0}) \geq \frac{1}{2t} d_{TV}^2(\Pi_{0^{t+1}}, \Pi_{1^t 0}).$$

Now, as in (6), we claim that $d_{TV}(\Pi_{0^{t+1}}, \Pi_{1^t 0}) = \Omega(1)$. This follows from the 1-way property of the protocol and the same derivation after (6), by considering T to be the concatenation of the first k player messages, and using the correctness of the protocol when Charlie's input $c = 1$, to show that $d_{TV}(T(0^t), T(1^t)) = \Omega(1)$. We thus arrive at the lower bound of $\Omega(N/t) = \Omega(n^{1-1/k} d^{1-1/p} \varepsilon^{-1})$ for our $(t+1)$-player modification to the t-player disjointness problem.

Finally, it suffices to show that a streaming algorithm providing a $(1+\Theta(\varepsilon))$-approximation to $\ell_p(\ell_q)(A)$ can decide whcih of the three cases above we are in. We again invoke it twice, once on the stream before the insertion of $2cn^{1/p}d^{1/q}$ into the j-th position of the $n \times d$ matrix, and once after the insertion of this item. In case (1) we have that $c = 0$, and so $\ell_p(\ell_q)(A) \leq n^{1/p}d^{1/q}$. In cases (2) and (3) we have that $c = 1$. If indeed $x_j = 2\varepsilon n^{1/p}d^{1/q}$, then when adding it to $2n^{1/p}d^{1/q}$, one can verify as in Section 4 that $\ell_p(\ell_q)(A)$ will increase by a $(1+\varepsilon)$-factor, that is, we will be in case (3). Otherwise, we will be in case (2). If the streaming algorithm provides a $(1+\Theta(\varepsilon))$-approximation, it can distinguish these two cases. Since the state of the streaming algorithm is passed t times, its state must be of size at least $\Omega(N/t^2) = \Omega(n^{1-2/k}d^{1-2/p}\varepsilon^{-2})$, as desired.

Finally, we briefly sketch our improvement to the lower bound for the $\ell_2(\ell_0)(A)$ problem, see, Section 3 of [JW09] for an $\Omega(n^{1/2})$ lower bound. The authors use a 2-player lower communication problem to achieve this lower bound: Alice gets as input n strings $x_1, \ldots, x_n \in \{0,1\}^d$, while Bob gets as input n strings $y_1, \ldots, y_n \in \{0,1\}^n$. The players are promised that either for all $i \in [n]$, $\|x_i - y_i\|_1 \leq 1$, where $\|\cdot\|_1$ denotes the 1-norm, or there is a unique $i \in [n]$ for which $\|x_i - y_i\|_1 = d$ and for all $j \neq i$, $\|x_j - y_j\|_1 \leq 1$. The authors use the direct sum theorem with a collapsing distribution μ^n, and show a lower bound of $\Omega(n/d)$. For $d = n^{1/2}$, they show that a streaming algorithm obtaining a constant factor approximation to $\ell_2(\ell_0)(A)$ can decide which case the players are in, thereby establishing an $\Omega(n^{1/2})$ lower bound for the streaming algorithm. By instead setting $d = (\varepsilon n)^{1/2}$, the same analysis shows that a streaming algorithm providing a $(1+\varepsilon)$-approximation to $\ell_2(\ell_0)(A)$ can decide which case the players are in, establishing the stronger $\Omega(n^{1/2}/\varepsilon^{1/2})$ lower bound.

We can instead set $d = \varepsilon n^{1/2}$, and introduce a third player Charlie. Charlie holds a bit $c \in \{0,1\}$ together with an identity of a row of A, that is, an index $j \in [n]$. We append each row of the matrix A with $n^{1/2}$ additional zeros, so A is now $n \times (1+\varepsilon)n^{1/2}$. If Charlie's input bit $c = 0$, then the output is $\ell_2(\ell_0)(A)$, where A is the matrix determined by Alice and Bob's input padded by zeros. If the bit $c = 1$, Charlie inserts $n^{1/2}$ ones on the last $n^{1/2}$ entries in the j-th row. One can verify that a streaming algorithm providing a $(1+\varepsilon)$-approximation to $\ell_2(\ell_0)(A)$ applied before Charlie inserts his ones (in the case that his input bit $c = 1$) together with a $(1+\varepsilon)$-approximation to $\ell_2(\ell_0)(A)$ applied after Charlie inserts his ones, can solve the 2-player communication problem between Alice and Bob with $d = \varepsilon n^{1/2}$. Using similar arguments to those in [JW09] for the 2-player game in conjunction with our arguments above for modifying the information-theoretic arguments to account for the new player Charlie, this 3-player game results in the stronger $\Omega(n/d) = \Omega(n^{1/2}/\varepsilon)$ lower bound. As this proof is quite similar to the proofs already given for F_p and $\ell_p(\ell_q)$, we omit further details.