New Characterizations in Turnstile Streams with Applications

Yuqing Ai[1], Wei Hu[2], Yi Li[3], and David P. Woodruff[4]

1 **Institute for Theoretical Computer Science (ITCS), Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China**
ayq12@mails.tsinghua.edu.cn

2 **Institute for Theoretical Computer Science (ITCS), Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China**
huw12@mails.tsinghua.edu.cn

3 **Facebook, Inc., Menlo Park, USA**
leeyi@umich.edu

4 **IBM Research – Almaden, San Jose, USA**
dpwoodru@us.ibm.com

Abstract

Recently, [Li, Nguyen, Woodruff, STOC'2014] showed any 1-pass constant probability streaming algorithm for computing a relation f on a vector $x \in \{-m, -(m-1), \ldots, m\}^n$ presented in the turnstile data stream model can be implemented by maintaining a linear sketch $A \cdot x \bmod q$, where A is an $r \times n$ integer matrix and $q = (q_1, \ldots, q_r)$ is a vector of positive integers. The space complexity of maintaining $A \cdot x \bmod q$, not including the random bits used for sampling A and q, matches the space of the optimal algorithm[1].

We give multiple strengthenings of this reduction, together with new applications. In particular, we show how to remove the following shortcomings of their reduction:

1. *The Box Constraint.* Their reduction applies only to algorithms that must be correct even if $\|x\|_\infty = \max_{i\in[n]} |x_i|$ is allowed to be much larger than m at intermediate points in the stream, provided that $x \in \{-m, -(m-1), \ldots, m\}^n$ at the end of the stream. We give a condition under which the optimal algorithm is a linear sketch even if it works only when promised that $x \in \{-m, -(m-1), \ldots, m\}^n$ at all points in the stream. Using this, we show the first super-constant $\Omega(\log m)$ bits lower bound for the problem of maintaining a counter up to an additive ϵm error in a turnstile stream, where ϵ is any constant in $(0, \frac{1}{2})$. Previous lower bounds are based on communication complexity and are only for relative error approximation; interestingly, we do not know how to prove our result using communication complexity. More generally, we show the first super-constant $\Omega(\log m)$ lower bound for additive approximation of ℓ_p-norms; this bound is tight for $1 \leq p \leq 2$.
2. *Negative Coordinates.* Their reduction allows x_i to be negative while processing the stream. We show an equivalence between 1-pass algorithms and linear sketches $A \cdot x \bmod q$ in dynamic graph streams, or more generally, the strict turnstile model, in which for all $i \in [n]$, $x_i \geq 0$ at all points in the stream. Combined with [Assadi, Khanna, Li, Yaroslavtsev, SODA'2016], this resolves the 1-pass space complexity of approximating the maximum matching in a dynamic graph stream, answering a question in that work.
3. 1-*Pass Restriction.* Their reduction only applies to 1-pass data stream algorithms in the turnstile model, while there exist algorithms for heavy hitters and for low rank approximation which provably do better with multiple passes. We extend the reduction to algorithms which make any number of passes, showing the optimal algorithm is to choose a new linear sketch at the beginning of each pass, based on the output of previous passes.

[1] Note the [LNW14] reduction does not lose a $\log m$ factor in space as they claim if it maintains $A \cdot x \bmod q$ rather than $A \cdot x$ over the integers.

31st Conference on Computational Complexity (CCC 2016).
Editor: Ran Raz; Article No. 20; pp. 20:1–20:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases communication complexity, data streams, dynamic graph streams, norm estimation

Digital Object Identifier 10.4230/LIPIcs.CCC.2016.20

1 Introduction

In the turnstile streaming model [6, 10], there is an underlying n-dimensional vector x which is initialized to $\vec{0}$. The data stream consists of updates of the form $x \leftarrow x + e_i$ or $x \leftarrow x - e_i$, where e_i is the i-th standard unit vector in $\mathbb{R}^n$. The goal of a streaming algorithm is to make one or more passes over the stream and use limited memory to approximate a function of x with high probability.

1.1 Linear Sketches and Simultaneous Communication Complexity

All known algorithms for problems in the turnstile model have a similar form: they first choose a (possibly random) integer matrix A, then maintain the "linear sketch" $A \cdot x$ in the stream, and finally output a function of $A \cdot x$. Li et al. [9] showed that any 1-pass constant probability streaming algorithm for approximating an arbitrary function f of x in the turnstile model can be reduced to an algorithm which, before the stream begins, samples a matrix A uniformly from $O(n \log m)$ hardwired integer matrices, then maintains the linear sketch $A \cdot x \bmod q$, where $q = (q_1, \ldots, q_r)$ is a vector of positive integers and r is the number of rows of A. Furthermore, the logarithm of the number of all possibilities for $A \cdot x \bmod q$, as x ranges over $\{-m, -(m-1), \ldots, m\}^n$, plus the number of random bits for sampling A, is larger than the space used by the original algorithm for approximating f by at most an additive $O(\log n + \log \log m)$ bits. Here the extra $O(\log n + \log \log m)$ bits are used only for sampling A and q. We refer to this as the LNW reduction[2].

The LNW reduction is non-uniform, i.e., the space complexity does not count the number of bits to store the $O(n \log m)$ hardwired possible sketching matrices A, nor does it count the space to compute the output given $A \cdot x$. The space counts only the space of storing $A \cdot x$. Thus, the LNW reduction is mostly useful in proving *lower bounds*, which only become stronger by not counting some parts in the space complexity. A widely used technique for proving lower bounds on the space of streaming algorithms is communication complexity [15]. One can get a 1-pass space lower bound for any streaming algorithm $\mathcal{A}$ by constructing a communication problem in which the players create data streams based on their inputs and run $\mathcal{A}$ on these streams sequentially. At the end of each stream, the current player passes the memory contents of $\mathcal{A}$ to the next player, and the next player continues with the received intermediate state. If $\mathcal{A}$ outputs the correct answer for the communication problem with constant probability at the end of the stream of the last player, then the space of $\mathcal{A}$ is at least the one-way communication complexity of the communication problem divided by (the number of players $- 1$).

The LNW reduction makes it possible to instead consider the *simultaneous communication model*. Compared to one-way communication, the simultaneous communication model is a

[2] In [9] the linear sketch of the form $A \cdot x \bmod q$ is further reduced to the form $A \cdot x$, which leads to a multiplicative $O(\log m)$ factor loss in the space complexity, assuming $m = \text{poly}(n)$; we do not consider that further reduction in this paper.

more restrictive model in which each player can only send a single message to an additional player called the referee, who receives no input in the communication problem. The referee then announces its output. By reducing an algorithm in an arbitrary form to a linear sketch and exploiting the linearity of matrix multiplication, the LNW reduction shows that to obtain lower bounds in the turnstile model, it suffices to consider the simultaneous communication model. This technique was applied in the original paper [9] and followup work for estimating frequency moments [13].

1.2 Shortcomings of the LNW Reduction

The LNW reduction has several drawbacks, which we now describe.

The Box Constraint. First, the reduction can only be performed under the assumption that the algorithm works as long as the underlying vector x belongs to $\{-m, -(m-1), \dots, m\}^n$ *at the end* of the stream, while in certain settings a more natural requirement may be that x belongs to $\{-m, -(m-1), \dots, m\}^n$ *at all intermediate points* of the stream. We refer to the restriction that the algorithm must be correct (with constant probability) provided that $x \in \{-m, -(m-1), \dots, m\}^n$ at the end of the stream, even if $\|x\|_\infty > m$ at an intermediate point, as *the box constraint*. It is possible that there are more space-efficient algorithms, not based on linear sketches, which abort if $\|x\|_\infty$ ever becomes larger than m. Due to this reason, the lower bounds obtained via simultaneous communication complexity only apply to the class of streaming algorithms assuming the box constraint.

Negative Coordinates. The second drawback is that the reduction works only in the turnstile model which allows negative frequencies, and does not work in the *strict turnstile model* in which the underlying vector always has no negative entries. For graph problems, a multi-graph with n vertices is defined as a stream in which each update corresponds to the addition or the deletion of an edge between two vertices. The multiplicity of every edge is naturally required to be always non-negative, so the strict turnstile model is standard for graph problems. The input for graphs in this model is called a *dynamic graph stream*. Similar to the turnstile model, linear sketching is the only existing technique for designing streaming algorithms in dynamic graph streams. It is unknown whether there is an equivalence between linear sketches and single-pass algorithms in the strict turnstile model.

1-Pass Restriction. Another shortcoming of the LNW reduction is that it only applies to 1-pass data stream algorithms in the turnstile model, while there exist algorithms for heavy hitters and for low rank approximation which provably do better with multiple passes. It is unknown if there exists a similar characterization for multi-pass algorithms.

1.3 Our Contributions

We make significant progress on removing the above shortcomings of the LNW reduction.

The Box Constraint. We give a condition under which the box constraint on the algorithm can be removed. Under this condition, we show that the streaming algorithm can be reduced to a linear sketch if it is correct with constant probability for streams whose underlying vector x always belongs to $\{-m, -(m-1), \dots, m\}^n$ at any point in the stream. In other words, we do not require algorithms to be correct when $\|x\|_\infty > m$ at intermediate points in the stream. Consequently, when our condition is satisfied, the lower bounds obtained via simultaneous communication complexity are stronger since the box constraint on algorithms is removed. Our condition for removing the box constraint is that the algorithm has space complexity at most $O((\log m)/n)$; so it is most useful when $m \gg n$ or $n = O(1)$. Note that while this does not apply to a number of data stream problems, it does apply to some very fundamental ones, described below, such as maintaining a counter in a stream, for which $n = 1$. We also show our condition that the space be $O((\log m)/n)$ bits is tight in the sense that for larger space algorithms, the LNW reduction fails unless one allows $\|x\|_\infty > m$ at intermediate points. That is, we give an example of an algorithm with $\Omega((\log m)/n)$ bits of space for which if one applies the LNW reduction, to argue correctness one needs $\|x\|_\infty > m$.

Negative Coordinates. We show in the *strict turnstile model*, there is a reduction from a general 1-pass algorithm to a linear sketch, that is, the optimal algorithm is a linear sketch even if promised that $x_i \geq 0$ at all points in the stream. Here we assume the space complexity of the algorithm depends only on n, even if the underlying vector is allowed to have very large entries at intermediate points in the stream. This assumption is suitable for graph problems for which the desired lower bounds are usually in terms of n [2]. Note that for graph and multi-graph problems with edge weight multiplicity bounded by $\mathrm{poly}(n)$, such a condition does not affect known upper bounds. Indeed, known algorithms are linear sketches, for which each coordinate can be maintained modulo $\mathrm{poly}(n)$.

1-Pass Restriction. We extend the reduction to algorithms which make any number of passes, showing the optimal algorithm is to choose a new linear sketch at the beginning of each pass, based on the output of previous passes. We note that in [7], significantly better bounds for finding ℓ_2-heavy hitters were found using multiple passes, while in [3] the 2-round protocol in the arbitrary partition model there can be implemented as a 2-pass streaming algorithm with better space than possible of any 1-pass algorithm [14].

1.4 Applications

Norm Approximation and Maintaining a Counter

A fundamental problem in the turnstile streaming model is norm approximation [1], in which the goal is to output an approximation of the ℓ_p-norm $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$, for given $p > 0$.[3] In particular, we are interested in proving space lower bounds for any 1-pass algorithm that outputs *additive error* approximation of the ℓ_p-norm: for $x \in \{-m, -(m-1), \ldots, m\}^n$, the algorithm outputs a number in $\left[\|x\|_p - \epsilon n^{1/p} m, \|x\|_p + \epsilon n^{1/p} m\right]$ with high probability. Since we have $\|x\|_p \leq n^{1/p} m$ for all $x \in \{-m, -(m-1), \ldots, m\}^n$, a $(1 \pm \epsilon)$-relative error approximation implies an $(\pm \epsilon n^{1/p} m)$-additive error approximation; however, an additive error approximation is much weaker. In some applications, relative error is too restrictive and one may only be interested in the value of a norm if it is sufficiently large. However, all previous lower bounds, e.g., [8], only apply if the norm is allowed to be very small, that is, they do not apply to additive error approximation.

We obtain the first super-constant $\Omega(\log m)$ lower bound for approximating $\|x\|_p$ up to an *additive* $\epsilon n^{1/p} m$ error in the turnstile model, without any assumptions such as the box constraint, where ϵ is any constant in $(0, 1/2)$. Our lower bound of $\Omega(\log m)$ bits is optimal for the important case of $p \in [1, 2]$, which includes the Manhattan and Euclidean norms.

[3] For $0 < p < 1$, $\|x\|_p$ is not a norm, though it is still a well-defined function.

Indeed, for $p \in [1,2]$ one can obtain a relative error approximation using $O(\log m)$ bits of space [1, 5].

Previous lower bound techniques are based on two-player communication complexity in which the players, Alice and Bob, hold inputs $x \in \{-m, -(m-1), \ldots, m\}^n$ and $y \in \{-m, -(m-1), \ldots, m\}^n$ respectively, and should output an approximation to $\|x - y\|_p$. But for an additive error approximation, it suffices for Alice to send the most significant $O(1)$ bits of x to Bob, and thus only an $\Omega(1)$ lower bound can be proved via communication complexity.

For the special case of $n = 1$, the data stream is composed of $+1$'s and -1's and the underlying vector x is an integer, i.e., a "counter". When x is promised to stay in $\{-m, -(m-1), \ldots, m\}$, we are interested in the space complexity of maintaining $|x|$ up to an additive ϵm error, for constant $\epsilon \in (0, 1/2)$. Surprisingly, the space complexity of this problem in the turnstile model with additive error was previously unknown. There is an obvious $O(\log m)$ upper bound for this problem because the algorithm can just maintain x. The question is whether this upper bound is tight. By removing the box constraint of the LNW reduction, we give the first tight $\Omega(\log m)$ bits lower bound for this fundamental problem. As a simple corollary, we show that outputting the most significant bit of $|x|$ in a turnstile stream requires $\Omega(\log m)$ bits of space. Here we let $|x|$ take $(\lfloor \log m \rfloor + 1)$ bits so its most significant bit can be 0. Note that this is in sharp contrast to maintaining the least significant $O(1)$ bits, which can be done with $O(1)$ bits of space. Indeed, if one is interested in the C least significant bits, it suffices to maintain a counter modulo 2^C.

Matching Problems

Matching problems are among the most studied graph problems in the streaming model. In a recent work [2], Assadi et al. give a 1-pass algorithm using $\tilde{O}(n^{2-3\epsilon})$ bits of space to recover an n^ϵ-approximate maximum matching in dynamic graph streams. They also show a lower bound of $n^{2-3\epsilon-o(1)}$ bits for any linear sketch that approximates the maximum matching to within a factor of $O(n^\epsilon)$. Their bounds are essentially tight for linear sketches, but it remains to see whether they are also tight for general 1-pass algorithms.

Our result for non-negative streams implies that the upper and lower bounds in [2] for approximating maximum matching are tight not only for linear sketches, but for all 1-pass algorithms. Thus the space complexity of approximating maximum matching in dynamic graph streams is resolved.

2 Preliminaries

We present some notations and definitions in this section.

Data Streams in the Turnstile Model. Let e_i be the i-th standard unit vector in $\mathbb{R}^n$. In the turnstile streaming model, the input $x \in \mathbb{Z}^n$ is represented as a data stream $\sigma = (\sigma_1, \sigma_2, \ldots)$ in which each element σ_i belongs to $\Sigma = \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$ and $\sum_i \sigma_i = x$.

The frequency of a stream σ is denoted by $\operatorname{freq} \sigma = \sum_i \sigma_i$. For two streams σ and τ, let $\sigma \circ \tau$ be the stream obtained by concatenating τ to the end of σ. The inverse stream of σ, denoted by σ^{-1}, is defined inductively by $e_i^{-1} = -e_i$, $(-e_i)^{-1} = e_i$ and $(\sigma \circ \tau)^{-1} = \tau^{-1} \circ \sigma^{-1}$.

Let $\Lambda_m = \{\sigma \mid \|\operatorname{freq} \sigma\|_\infty \le m\}$ and $\Gamma_m = \{\sigma \mid \text{for any prefix } \sigma' \text{ of } \sigma, \|\operatorname{freq} \sigma'\|_\infty \le m\}$; the former is the set of input streams without the removal of the box constraint, and the latter the set of input streams when the box constraint is removed.

The Strict Turnstile Model. In the strict turnstile model, there is an additional requirement that the underlying vector should not have negative coordinate at any intermediate point. In other words, this model allows input streams in $\Lambda_m^* = \{\sigma \mid \|\operatorname{freq}\sigma\|_\infty \leq m$, and for any prefix σ' of σ, $\sigma' \geq \vec{0}\}$.

Stream Automata. A stream automaton $\mathcal{A}$ is a Turing machine that uses two tapes, a unidirectional read-only input tape and a bidirectional work tape. The input tape contains the input stream σ. After processing its input, the automaton writes an output, denoted by $\phi_{\mathcal{A}}(\sigma)$, on the work-tape. A configuration of $\mathcal{A}$ is determined by its state of the finite control, head position and contents on the work tape. We often use the word "state" to mean a configuration. The computation of $\mathcal{A}$ can be described by a transition function $\oplus : C \times \Sigma \to C$, where C is the set of all possible configurations. For a configuration $c \in C$ and a stream σ, we also denote by $c \oplus \sigma$ the configuration after processing σ on c. The set of configurations of $\mathcal{A}$ that are achievable by some input stream $\sigma \in \Gamma_m$ is denoted by $C(\mathcal{A}, m)$. The space of $\mathcal{A}$ with stream parameter m is then defined to be $S(\mathcal{A}, m) = \log |C(\mathcal{A}, m)|$.

A problem P is characterized by a family of binary relations $P_n \subseteq \mathbb{Z}^{p(n)} \times \mathbb{Z}^n$, where $n \geq 1$ and $p(n)$ is the dimension of the output. We say an automaton $\mathcal{A}$ solves a problem P (with domain size n) on a distribution Π if $(\phi_{\mathcal{A}}(\sigma), \operatorname{freq}\,\sigma) \in P_n$ with probability $1 - \delta$, where the probability is over $\sigma \sim \Pi$ (and where δ is a small positive constant specified when needed).

Path-Reversible Automata and Path-Independent Automata. An automaton is said to be path-reversible if for any configuration c and any input stream σ, $c \oplus (\sigma \circ \sigma^{-1}) = c$. An automaton is said to be path-independent if for any configuration c and any input stream σ, $c \oplus \sigma$ depends only on freq σ and c.

Transition Graph. The transition graph of an automaton $\mathcal{A}$ is a directed graph $G_{\mathcal{A}} = (V, E)$, where the vertex set V is the set of configurations of $\mathcal{A}$, and the arcs in E describe the transition function of $\mathcal{A}$: there is an arc a from vertex c_1 to c_2 if and only if there is an update $u \in \Sigma$ such that $c_1 \oplus u = c_2$, and we denote this update u by f_a. Note that every vertex in V has $2n$ outgoing arcs, each of which corresponds to a possible update in Σ.

Zero-Frequency Path and Zero-Frequency Graph. In a transition graph $G_{\mathcal{A}} = (V, E)$, a path p of length k from v_1 to v_{k+1} is a sequence of k arcs $(v_1, v_2), (v_2, v_3), \ldots, (v_k, v_{k+1})$. Let $f_p = \sum_{i=1}^{k} f_{(v_i, v_{i+1})}$ be the frequency of path p, which is the frequency of the stream along p. A path p is called a zero-frequency path if $f_p = \vec{0}$. The zero-frequency graph $G'_{\mathcal{A}} = (V', E')$ based on $G_{\mathcal{A}} = (V, E)$ is a directed graph with $V' = V$ and $E' = \{(v_1, v_2) \mid$ there exists a zero-frequency path from v_1 to v_2 in $G_{\mathcal{A}}\}$.

Randomized Stream Automata. A randomized stream automaton is a deterministic automaton with one additional tape for the random bits. The random bit string R is initialized on the random bit tape before any input token is read; then the random bit string is used in a bidirectional read-only manner. The rest of the execution proceeds as in a deterministic automaton. A randomized automaton $\mathcal{A}$ is said to be path-independent (reversible) if, for each possible randomness R, the deterministic instance $\mathcal{A}_R$ is path-independent (reversible). The space of a randomized automaton $\mathcal{A}$ with stream parameter m is defined as $S(\mathcal{A}, m) = \max_R (|R| + S(\mathcal{A}_R, m))$.

Multi-Pass Stream Automata. A p-pass ($p \geq 2$) deterministic automaton $\mathcal{A}$ consists of automata of p layers (passes), in which (i) the 1-st pass automaton $\mathcal{A}^1$ which contains a starting state, when reading an input stream σ and arriving at state s, outputs a second-pass deterministic automaton $\mathcal{A}^2_s$; (ii) for $2 \leq q \leq p-1$, a q-th pass automaton $\mathcal{A}^q$, when reading input σ and arriving at a state s, outputs a $(q+1)$-th pass deterministic automaton $\mathcal{A}^{q+1}_s$; (iii) a p-th pass automaton $\mathcal{A}^p$, when reading input σ and arriving at a state s, outputs a final answer for the input σ. When the context is clear, we also use σ to mean the terminating state of the automaton when reading stream σ, e.g., $\mathcal{A}^2_\sigma$ is the same as $\mathcal{A}^2_{o \oplus \sigma}$, where o is the initial state of $\mathcal{A}^2$.

For an input stream σ, a sequence of automata will be generated, $\mathcal{A}^1, \mathcal{A}^2_{s_1}, \ldots, \mathcal{A}^p_{s_{p-1}}$, where s_i $(1 \leq i \leq p-1)$ is the terminating state of $\mathcal{A}^i_{s_{i-1}}$ (where $\mathcal{A}^1_{s_0} = \mathcal{A}^1$) on reading σ. A p-pass deterministic automaton $\mathcal{A}$ solves a problem P on input stream σ if the output of $\mathcal{A}^p_{s_{p-1}}$ on σ is an acceptable answer for σ. We say that the answer is *acceptable* for σ in this case. The space complexity $S(\mathcal{A}, m)$ is defined as $S(\mathcal{A}, m) = \max_{\sigma : \|\operatorname{freq}(\sigma)\|_\infty \leq m} \{S(\mathcal{A}^1, m) + S(\mathcal{A}^2_{s_1}, m) + \ldots + S(\mathcal{A}^p_{s_{p-1}}, m)\}$.

A p-pass automaton is said to be path-independent if all of its constituent automata are path-independent. A p-pass randomized automaton is defined similarly as a 1-pass randomized automaton.

3 Removing the Box Constraint and Application to Additive Error Norm Approximation

In this section, we give a condition under which the box constraint can be removed. As an application, we obtain an $\Omega(\log m)$ bit lower bound for the additive error ℓ_p-norm $\|x\|_p (p > 0)$ estimation in the turnstile streaming model.

First of all, we remark that the LNW reduction in [9] can be simplified. The LNW reduction consists of three main steps: (i) reduction from a general automaton to a path-reversible automaton; (ii) reduction from a path-reversible automaton to a path-independent automaton; (iii) reduction from a path-independent automaton to a linear sketch $A \cdot x$.[4] We notice that step (ii) is not necessary, since the automaton obtained after step (i), which was shown to be path-reversible in [9], is already path-independent. The proof of this fact in given in Appendix A.

3.1 Removing the Box Constraint in the Turnstile Model

In the LNW reduction, given a problem P, if an algorithm $\mathcal{A}$ solves P on Λ_m, then it is reduced to a linear sketch that solves P on Λ_m. (Recall that Λ_m is the set of all streams σ with $\|\operatorname{freq} \sigma\|_\infty \leq m$.) Our goal is to remove the box constraint, i.e., to apply the reduction to algorithms that solve P on Γ_m, which is the set of streams σ such that $\|\operatorname{freq} \sigma'\|_\infty \leq m$ for any prefix of σ' of σ. We will show that if $\mathcal{A}$ uses space $S(\mathcal{A}, m) \leq c \cdot \frac{\log m}{n}$ for some fixed constant $c > 0$ and solves P on Γ_m, then it can be reduced to a linear sketch that solves P on $\Gamma_{m/2}$. The reason why the LNW reduction requires $\mathcal{A}$ to solve the problem on Λ_m instead of Γ_m comes from the step of reducing a general automaton to a path-independent automaton. Given a certain deterministic instance of the general automaton $\mathcal{A}$, the transition graph $G_\mathcal{A} = (V, E)$ and the zero-frequency graph $G'_\mathcal{A} = (V', E')$ are built. The states of

[4] Note that a path-independent automaton is equivalent to maintaining a linear sketch $A \cdot x \bmod q$ [4, 9]. The only goal of step (iii) is to remove the "mod q".

a corresponding deterministic instance of the new automaton $\mathcal{B}$ are defined to be all the *terminal* strongly connected components[5] of $G'_{\mathcal{A}}$. The transition function of $\mathcal{B}$ is then defined based on the original transition function of $\mathcal{A}$. When the final state in $\mathcal{B}$ is a strongly connected component C of $G'_{\mathcal{A}}$, $\mathcal{B}$ chooses a vertex v from C according to the stationary distribution π_C of a random walk in C, and then outputs what $\mathcal{A}$ outputs when its state is v. We note that when a stream σ is executed by $\mathcal{B}$, what essentially happens is that some zero-frequency streams (corresponding to moving along arcs in $G'_{\mathcal{A}}$) are inserted into σ to form a new stream σ' which will be executed by $\mathcal{A}$. We have that $\sigma \in \Lambda_m$ implies $\sigma' \in \Lambda_m$; but when $\sigma \in \Gamma_m$, the frequency of some prefix of σ' could be very large. Thus $\mathcal{A}$ is required to solve the problem on Λ_m to make the reduction work.

Define L to be the maximum length of the shortest zero-frequency path connecting a pair of vertices. Here the maximum is taken over all pairs of vertices that are connected by at least one zero-frequency path. Note that the zero-frequency streams inserted into σ correspond to taking a walk in the zero-frequency graph $G'_{\mathcal{A}}$. Thus we can assume that the lengths of inserted zero-frequency streams are at most L: this can be achieved by always choosing the shortest zero-frequency paths between pairs of vertices. Then it is easy to see that $\sigma \in \Gamma_{m/2}$ implies $\sigma' \in \Gamma_{m/2+L/2}$. Hence, if $L \leq m$, we will be able to perform the reduction from an initial algorithm for input streams in Γ_m to a linear sketch for input streams in $\Gamma_{m/2}$.

Note that the reduction we use here is the same as step (i) in the LNW reduction. We obtain a path-independent automaton *regardless* of what input streams we are considering. What changes is that we have argued, when $L \leq m$ is satisfied, the path-independent automaton we obtain will be correct on $\Gamma_{m/2}$ if the original automaton is correct on Γ_m. Now it remains to find a sufficient condition for $L \leq m$.

We will prove an upper bound on L in terms of n and the number of vertices $s = |V|$ in order to obtain a condition for removing the box constraint. The idea to upper bound L is to build linear equations based on the graph $G_{\mathcal{A}}$ such that the equations have a positive integer solution if and only if there exists a zero-frequency path from a given state to another. Then, Lemma 3.1 below enables us to find a positive integer solution of small magnitude if there exists one. Finally, we convert the bound on the magnitude of the solutions to the bound on the length of the path. In the LNW reduction, one way to obtain a finite bound on L as mentioned in that work is to build a system of linear equations in terms of simple paths and simple cycles, though an exact bound is not given in [9]. We note that an upper bound $s^{O(s+n)}$ on L can be obtained via this approach. (See Appendix B.) Unfortunately the bound $s^{O(s+n)}$ is not strong enough for our applications. Here in Lemma 3.3 we propose a better way to build the linear equations, which gives us a tighter bound of $\mathrm{poly}(sn) \cdot (\frac{s}{n}+1)^n$. Instead of writing the linear equations in terms of simple cycles, we write them in terms of arcs.

▶ **Lemma 3.1.** *Let A be an $m \times n$ integer matrix and $b \in \mathbb{Z}^m$. Suppose that M_1 is an upper bound on the absolute value of any sub-determinant of the matrix $\begin{pmatrix} A & b \end{pmatrix}$. If $Ax = b$ has a positive integer solution, then it has one whose all coordinates are at most $(n+1)^2 M_1$.*

Proof. We make use of a result in [12]. Let C be a $p \times n$ integer matrix and $d \in \mathbb{Z}^p$. Let r be the rank of A. Suppose that M is an upper bound on the absolute value of any sub-determinant of the matrix $\begin{pmatrix} A & b \\ C & d \end{pmatrix}$, which contains at least r rows from $\begin{pmatrix} A & b \end{pmatrix}$. The

[5] A strongly connected component is said to be terminal if there is no arc coming from it to the rest of the graph.

following upper bound is shown on the magnitude of an integer solution to the linear system $\{Ax = b, Cx \geq d\}$:

▶ **Lemma 3.2** ([12]). *If $Ax = b$ and $Cx \geq d$ have a common integer solution, then they have one whose coordinates have absolute values at most $(n+1)M$.*

Now we let $p = n$, $C = I_n$, and $d = \vec{1} = (1, \dots, 1)^\top$ and invoke Lemma 3.2. Then we know that $Ax = b$ has a positive integer solution whose coordinates are at most $(n+1)M$, where M is an upper bound on the absolute value of any sub-determinant of the matrix $\begin{pmatrix} A & b \\ I_n & \vec{1} \end{pmatrix}$, which contains at least r rows from $\begin{pmatrix} A & b \end{pmatrix}$.

Then it suffices to prove $M \leq (n+1)M_1$. Consider an arbitrary submatrix T of $\begin{pmatrix} A & b \\ I_n & \vec{1} \end{pmatrix}$, which contains at least r rows from $\begin{pmatrix} A & b \end{pmatrix}$. Note that all entries of $\begin{pmatrix} I_n & \vec{1} \end{pmatrix}$ are in $\{0, 1\}$ and that there are $n+1$ ways to choose one non-zero entry from each row of $\begin{pmatrix} I_n & \vec{1} \end{pmatrix}$ such that no two entries are in the same column. Thus, after expanding the determinant of T along its rows from $\begin{pmatrix} I_n & \vec{1} \end{pmatrix}$, $\det(T)$ can be written as the sum of no more than $n+1$ sub-determinants (multiplied by ± 1) of $\begin{pmatrix} A & b \end{pmatrix}$. Therefore $|\det(T)|$ is at most $n+1$ times the largest absolute value of any sub-determinant of $\begin{pmatrix} A & b \end{pmatrix}$, which implies $M \leq (n+1)M_1$. ◀

▶ **Lemma 3.3.** *Let $s = |V|$ be the number of vertices in the transition graph $G_{\mathcal{A}} = (V, E)$. Then $L \leq 2ns(2ns+1)^2 \cdot (\frac{s}{n}+1)^n$, where L is the maximum length of the shortest zero-frequency path between any two vertices connected by at least one zero-frequency path.*

Proof. Consider any two vertices $o_1, o_2 \in V$ such that there exists a zero-frequency path from o_1 to o_2. We fix a subset of edges $E' = \{(u_1, v_1), (u_2, v_2), \dots, (u_t, v_t)\}$ satisfying the following condition: it is possible to use and only use $(u_1, v_1), \dots, (u_t, v_t)$ to reach o_2 from o_1. For every possible E' satisfying this condition, we build a linear system as follows.

Let $x \in \mathbb{Z}_+^t$ be the variable whose i-th coordinate x_i represents the number of times the arc (u_i, v_i) occurs in the path from o_1 to o_2. We will need two types of constraints to write the linear equations: (1) the frequency of the path is $\vec{0}$; (2) for each node v, the number of times we go out from v minus the number of times we go into v is 1 if $v = o_1$, is -1 if $v = o_2$, and is 0 otherwise.[6] Then each positive integer solution x to the above constraints corresponds to a zero-frequency path from o_1 to o_2 using the arcs in E'. It is easy to see that the above constraints can be written as linear equations $Ax = b$, where $A = \begin{pmatrix} f_{(u_1,v_1)} & f_{(u_2,v_2)} & \dots & f_{(u_t,v_t)} \\ e_{u_1} - e_{v_1} & e_{u_2} - e_{v_2} & \dots & e_{u_t} - e_{v_t} \end{pmatrix}$ is an $(n+s) \times t$ matrix, and $b = \begin{pmatrix} \vec{0} \\ e_{o_1} - e_{o_2} \end{pmatrix} \in \mathbb{R}^{n+s}$. Here e_v is the standard unit column vector in $\mathbb{R}^s$ with the non-zero coordinate corresponding to node v. (Note that there are s nodes in total so we can map every node to a coordinate.) The upper n rows guarantee the frequency of the path is 0. Here, recall that $f_{(u_i,v_i)} \in \mathbb{R}^n$ is the positive or negative standard unit column vector which is the update corresponding to the arc (u_i, v_i). The lower s rows are the network flow constraints. Note that all entries in $\begin{pmatrix} A & b \end{pmatrix}$ are in $\{-1, 0, 1\}$.

Since there exists a zero-frequency path from o_1 to o_2, at least one such system of the linear equations (i.e., for at least one fixing of E') has a positive integer solution. Next, we consider such a fixing of E' that leads to a positive integer solution to the corresponding linear

[6] These are called the network flow constraints.

system. By Lemma 3.1, if there exists a positive integer solution to $Ax = b$, then there exists such a solution x satisfying $\|x\|_\infty \le (t+1)^2 M_1$, where M_1 is the largest possible absolute value of any sub-determinant of $\begin{pmatrix} A & b \end{pmatrix}$. This solution x corresponds to a zero-frequency path of length $\|x\|_1 \le t\|x\|_\infty \le t(t+1)^2 M_1$.

Let S be an arbitrary square submatrix of $\begin{pmatrix} A & b \end{pmatrix}$. We write S as $S = \begin{pmatrix} X \\ Y \end{pmatrix}$, where X comes from the top n rows and Y comes from the bottom s rows of $\begin{pmatrix} A & b \end{pmatrix}$. Let the size of X be $h \times w$, and the size of Y be $(w-h) \times w$. (Clearly, we have $h \le n$ and $w \le s+h$.) We expand $\det(S)$ along its rows in X. Since each column in X has at most one non-zero entry (± 1), the rows of X have support on disjoint subsets of columns, and thus X has at most w non-zero entries in total. Then, by the AM-GM inequality, the number of ways to choose one non-zero entry from each row of X such that no two of them are in the same column is at most $\left(\frac{w}{h}\right)^h \le \left(\frac{s+h}{h}\right)^h = \left(1+\frac{s}{h}\right)^h \le \left(1+\frac{s}{n}\right)^n$. Then we have that $|\det(S)|$ is at most $\left(1+\frac{s}{n}\right)^n$ times the maximum absolute value of any sub-determinant of Y. Let $C = (c_{ij})_{k\times k}$ be any submatrix of Y, which is also a submatrix of $\begin{pmatrix} e_{u_1}-e_{v_1} & e_{u_2}-e_{v_2} & \dots & e_{u_t}-e_{v_t} & e_{o_1}-e_{o_2} \end{pmatrix}$.

We show that $\det(C) \in \{0,1,-1\}$. Note that C has at most two non-zero entries in each column, and if a column of C has two non-zero entries, they must be 1 and -1. If all columns in C have two non-zero entries, then $(1,1,\dots,1)\cdot C = (0,0,\dots,0)$, which implies $\det(C) = 0$. If there exists a column in C without non-zero entries, then we also have $\det(C) = 0$. Otherwise we can find a column with exactly one non-zero entry c_{ij}, and then we have $\det(C) = (-1)^{i+j} c_{ij} \det(D_{ij})$, where D_{ij} is formed by deleting row i and column j from C. By induction, we have $\det(C) \in \{-1,0,1\}$. Therefore, we know that $|\det(S)| \le \left(1+\frac{s}{n}\right)^n \cdot 1 = \left(1+\frac{s}{n}\right)^n$.

Since S is arbitrary, we have $M_1 \le \left(1+\frac{s}{n}\right)^n$. Note that there are at most $2ns$ arcs in the graph, so we have $t \le 2ns$. Then we can bound the length of a zero-frequency path by $t(t+1)^2 M_1 \le 2ns(2ns+1)^2\left(1+\frac{s}{n}\right)^n$. ◀

Let $r = |C(\mathcal{A}, m)|$. Note that we only care about the correctness of $\mathcal{A}$ on input streams from Γ_m. Thus we can, without loss of generality, combine all states not in $C(\mathcal{A}, m)$ (if there are any) into an "irreversible crash" state: the automaton will stay in this state after leaving $C(\mathcal{A}, m)$. This modification will not affect the correctness of $\mathcal{A}$ on Γ_m. Therefore, we can assume $s \le r+1$. Then Lemma 3.3 implies $L \le 2n(r+1)(2n(r+1)+1)^2 \cdot \left(\frac{r+1}{n}+1\right)^n$.

Assume $r > 1$. Then we have $L \le (4nr)^3 \cdot (r+2)^n \le r^{c_1 n}$ for a sufficiently large constant $c_1 > 0$. Recall that we want $L \le m$. Taking logarithms on both sides of the desired inequality $r^{c_1 n} \le m$, we equivalently want $c_1 n \log r \le \log m$, i.e., $\log r \le \frac{\log m}{c_1 n}$. Therefore, when we have $S(\mathcal{A}, m) = \log r \le \frac{\log m}{c_1 n}$ for some fixed constant $c_1 > 0$, the condition $L \le m$ will be satisfied. In this case, according to our analysis, the box constraint can be removed. Combining this condition and [9, Theorem 10] (with minor correction), we summarize our results on removing the box constraint as the following theorem.

▶ **Theorem 3.4** (Removing the box constraint). *Suppose that a randomized 1-pass streaming algorithm $\mathcal{A}$ solves a problem P on any stream in Γ_m with probability at least $1-\delta$, and that the space used by any deterministic instance of $\mathcal{A}$ is no more than $c \cdot \frac{\log m}{n}$, where $c > 0$ is a universal constant. Then there exists an algorithm $\mathcal{B}$ implemented by maintaining a linear sketch $A \cdot x \bmod q$ in the stream, where A is a random $r \times n$ integer matrix and q is a random positive integer vector of length r, such that $\mathcal{B}$ solves P on any stream in $\Gamma_{m/2}$ with probability $1-6\delta$ and that $S(\mathcal{B}, m/2) \le S(\mathcal{A}, m) + O(\log n + \log\log m + \log\frac{1}{\delta})$.*[7]

[7] The extra $O(\log n + \log\log m + \log\frac{1}{\delta})$ bits are used only for the randomness for sampling A and q.

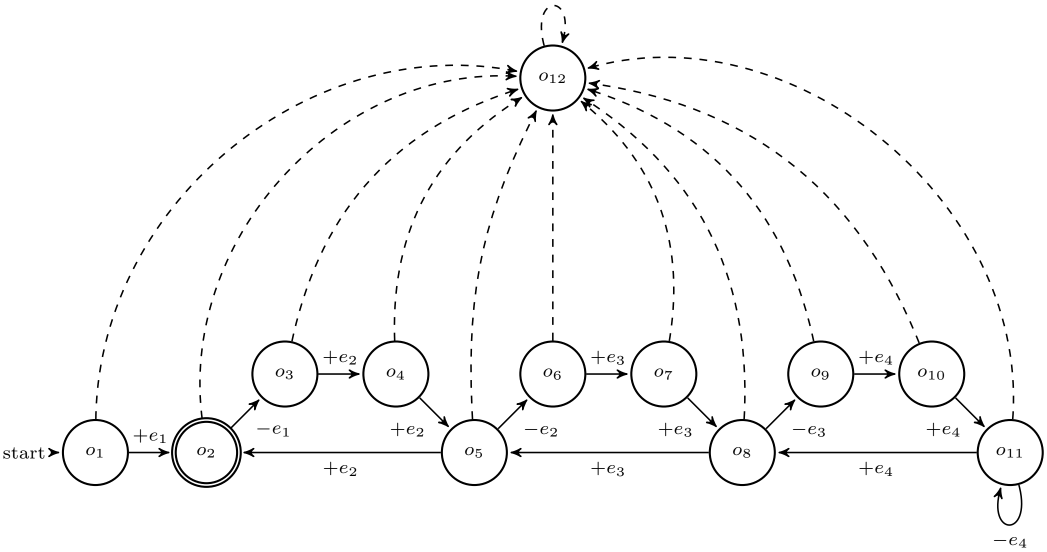

Figure 1 An illustrative example.

3.2 Tightness of Our Condition

In Lemma 3.3, we show that an upper bound of L is $\text{poly}(ns) \cdot (1 + \frac{s}{n})^{O(n)}$. Now we give an example of an automaton to show this upper bound is essentially tight (assuming s is much larger than n). In our example, there exist two vertices in the transition graph of the automaton so that the length of the shortest zero-frequency path between them is at least $(\frac{s}{n})^{\Omega(n)}$. The matching upper bound and lower bound on L eliminate the possibility of getting any further improvement in the condition for removing the box constraint, if the same reduction in [9] is used.

An illustrative example of our construction with $n = 4$ and $s = 12$ is given in Figure 1. Here each dashed arc represents all remaining outgoing arcs from a vertex. Consider the shortest zero-frequency path from o_1 to o_2. (Note that there exists one such path.) After going from o_1 to o_2 by the arc with $+e_1$ update, the path needs to go through the cycle $o_2 \to o_3 \to o_4 \to o_5 \to o_2$ one time in order to make the first coordinate of the frequency vector x equal to 0. However, that causes the second coordinate to be 3 and the path then needs to go through the cycle $o_5 \to o_6 \to o_7 \to o_8 \to o_5$ three times for compensation. That further causes the third coordinate to be 3^2 and the path needs to go through $o_8 \to o_9 \to o_{10} \to o_{11} \to o_8$ a total of 3^2 times. Finally, the path needs to go through the self-loop at o_{11} with $-e_4$ update a total of 3^3 times. Note that the number of times the shortest zero-frequency path from o_1 to o_2 passes through each cycle goes up exponentially.

▶ **Lemma 3.5.** *There exists an automaton $\mathcal{A}$ such that the transition graph $G_{\mathcal{A}} = (V, E)$ satisfies $L = (\frac{s}{n})^{\Omega(n)}$, where $s = |V|$ is the number of vertices, and L is the maximum length of the shortest zero-frequency path between any two vertices connected by at least one zero-frequency path.*

Proof. We assume $n > 3$, $s - 3 > 3(n - 1)$ and $(n - 1) \mid (s - 3)$. For $(n - 1) \nmid (s - 3)$, we can decrease s until $(n - 1) \mid (s - 3)$ is satisfied. We construct a transition graph $G_{\mathcal{A}} = (V, E)$ whose structure is similar to the example in Figure 1. Let $V = \{o_1, o_2, \ldots, o_s\}$. There are $2n$ outgoing arcs from each of the vertices in V. We write $o_i \oplus \pm e_k = o_j$ to stand for an arc

(o_i, o_j) in E where $f_{(o_i,o_j)} = \pm e_k$. Let $len = (s-3)/(n-1)+1$. The arcs in E are defined as follows:

- $o_1 \oplus +e_1 = o_2$.
- $o_{s-1} \oplus -e_n = o_{s-1}$.
- For $i \equiv 2 \bmod (len-1)$ and $i \neq s-1$, $o_i \oplus -e_{\lceil i/(len-1)\rceil} = o_{i+1}$.
- For $i \equiv 2 \bmod (len-1)$ and $i \neq 2$, $o_i \oplus +e_{\lceil i/(len-1)\rceil} = o_{i-(len-1)}$.
- For $i \not\equiv 2 \bmod (len-1)$, $i \neq 1$ and $i \neq s$, $o_i \oplus +e_{\lfloor i/(len-1)\rfloor+1} = o_{i+1}$.
- For a vertex o_i and $e \in \{+e_1, \ldots, +e_n, -e_1, \ldots, -e_n\}$ where $o_i \oplus e$ is undefined from above, $o_i \oplus e = o_s$.

There are $n-1$ cycles each of which has length len. The i-th cycle consists of nodes $o_{(i-1)(len-1)+2}, o_{(i-1)(len-1)+3}, \ldots, o_{i(len-1)+2}$. Among the arcs in the i-th cycle, there is one arc with $-e_i$ update, and all other $len-1$ arcs are with $+e_{i+1}$ update. In this case, any zero-frequency path from o_1 to o_2 passes through the i-th cycle at least $(len-1)^{i-1}$ times, and thus its length is at least $\sum_{i=1}^{n-1} len \cdot (len-1)^{i-1} = len \cdot (len-1) \cdot \frac{(len-1)^{n-1}-1}{len-2} = (\frac{s}{n})^{\Omega(n)}$. ◀

In fact, for our example we can have a stronger statement that along any zero-frequency path from o_1 to o_2, some coordinate of the underlying vector achieves at least $(\frac{s}{n})^{\Omega(n)}$. This is the reason why the underlying vector could escape from $\{-m, -(m-1), \ldots, m\}^n$ at the middle of the stream after inserting zero-frequency streams in the reduction. In order to remove the box constraint, there must be some constants C_1, C_2 such that $(\frac{s}{n})^{C_1 n} \leq C_2 m$, i.e., $\log s \leq \frac{\log m + \log C_2}{C_1 n} + \log n$. When m is sufficiently large and n is fixed, this implies $\log s \leq \frac{C \log m}{n}$ for some constant C. Hence our condition for removing the box constraint is tight.

3.3 Space Lower Bounds for Additive Error Norm Approximation

We consider the problem of estimating the ℓ_p-norm $\|x\|_p (p > 0)$ in the turnstile streaming model, where the underlying vector x is promised to be in $\{-m, -(m-1), \ldots, m\}^n$ at all points in the stream (i.e., without the box constraint). We prove an $\Omega(\log m)$ bit space lower bound for approximating $\|x\|_p$ up to an additive $\epsilon n^{1/p} m$ error, where $\epsilon \in (0, \frac{1}{2})$ is a constant. Our proof makes use of the LNW reduction with the box constraint removed.

A Norm Decision Problem. First we consider the following promise problem: we are given the promise that the input $x \in \{-m, -(m-1), \ldots, m\}^n$ satisfies either $\|x\|_p \leq \alpha n^{1/p} m$ or $\|x\|_p \geq \beta n^{1/p} m$, where $0 < \alpha < \beta < 1$ are constants, and need to decide whether $\|x\|_p \leq \alpha n^{1/p} m$ or $\|x\|_p \geq \beta n^{1/p} m$. We first prove that this problem has an $\Omega(\log m)$ space lower bound.

▶ **Theorem 3.6** (Norm decision problem). *For any constants $p > 0$, $0 < \alpha < \beta < 1$ and $0 \leq \delta < \frac{\min\{\alpha, 1-\beta\}}{6(\alpha+1-\beta)}$, any 1-pass streaming algorithm which, for any input $x \in \{-m, -(m-1), \ldots, m\}^n$, decides whether $\|x\|_p \leq \alpha n^{1/p} m$ or $\|x\|_p \geq \beta n^{1/p} m$ (provided that x satisfies one of them) with probability at least $1-\delta$ in the turnstile model uses $\Omega(\log m)$ bits of space.*

Proof. We assume without loss of generality that $n = 1$, since one can always use an algorithm for larger n to solve the problem for $n = 1$ by assigning all n coordinates the same value. Suppose that the theorem does not hold. Then for *any* sufficiently small constant $\epsilon > 0$, there exists m and an algorithm $\mathcal{A}$ such that $\mathcal{A}$ uses less than $\epsilon \log m$ bits of space and solves the given problem (with parameters m and $n = 1$) with probability $1-\delta$. Since the space used by $\mathcal{A}$ is less than $\frac{\epsilon \log m}{n}$ bits (using $n = 1$), from Theorem 3.4 we know that there is an algorithm $\mathcal{B}$ that maintains a linear sketch $A \cdot x \bmod q$ and solves the same problem[8] with probability $1 - 6\delta$. Furthermore, the space used by any deterministic instance of $\mathcal{B}$ is also less than $\epsilon \log m$ bits.

Let $U_1 = \{0, 1, \dots, \lfloor \alpha m \rfloor\}$ and $U_2 = \{\lceil \beta m \rceil, \lceil \beta m \rceil + 1, \dots, m\}$. Let Π be the uniform distribution on $U = U_1 \cup U_2$. By Yao's minimax principle, there exists a fixing of A and q that solves the problem for x drawn from Π (i.e., decides if x is in U_1 or U_2) with probability at least $1 - 6\delta$. Since $n = 1$, we can write $Ax \bmod q = (a_1 x \bmod q_1, a_2 x \bmod q_2, \dots, a_r x \bmod q_r)$, where $a_1, \dots, a_r \in \mathbb{Z}$ and $q_1, \dots, q_r \in \mathbb{Z}_+$. Without loss of generality, we assume $\gcd(a_i, q_i) = 1$ for $i = 1, \dots, r$. (If $\gcd(a_i, q_i) = d > 1$, we can let $a_i' = a_i/d, q_i' = q_i/d$ and then there is a one-to-one correspondence between $a_i x \bmod q_i$ and $a_i' x \bmod q_i'$. So a_i and q_i can be replaced by a_i' and q_i'.) Let $l = \operatorname{lcm}(q_1, \dots, q_r)$.

We now prove $l = \Omega(m)$. Suppose that $l < \min\{\alpha, 1 - \beta\} \cdot m$ (otherwise we already have $l = \Omega(m)$). The input space U can be partitioned into l groups $G_i = \{j \in U \big| (i - j) \bmod l = 0\} (i = 0, 1, \dots, l - 1)$. Note that the algorithm outputs the same answer for inputs from the same group. Within group G_i, the algorithm outputs the correct answer for at most a $\frac{\max\{|G_i \cap U_1|, |G_i \cap U_2|\}}{|G_i|}$ fraction of inputs. For $i \in \{0, 1, \dots, l - 1\}$ we have $|G_i \cap U_1| = \lfloor \frac{\alpha m - i}{l} \rfloor + 1 \in (\frac{\alpha m}{l} - 1, \frac{\alpha m}{l} + 1]$ and $|G_i \cap U_2| = \lfloor \frac{m - i}{l} \rfloor - \lceil \frac{\beta m - i}{l} \rceil + 1 \in (\frac{(1-\beta)m}{l} - 1, \frac{(1-\beta)m}{l} + 1]$. Thus

$$\frac{\max\{|G_i \cap U_1|, |G_i \cap U_2|\}}{|G_i|} \le \frac{\max\{\frac{\alpha m}{l} + 1, \frac{(1-\beta)m}{l} + 1\}}{(\frac{\alpha m}{l} - 1) + (\frac{(1-\beta)m}{l} - 1)} = \frac{\max\{\alpha, 1 - \beta\}\frac{m}{l} + 1}{(\alpha + 1 - \beta)\frac{m}{l} - 2}.$$

The above is an upper bound of the success probability on Π, so we must have $1 - 6\delta \le \frac{\max\{\alpha, 1-\beta\}\frac{m}{l} + 1}{(\alpha + 1 - \beta)\frac{m}{l} - 2}$, which means $l \ge \frac{(1-6\delta)(\alpha+1-\beta) - \max\{\alpha, 1-\beta\}}{3 - 12\delta} m$. Since $\delta < \frac{\min\{\alpha, 1-\beta\}}{6(\alpha+1-\beta)}$, we have $(1 - 6\delta)(\alpha + 1 - \beta) - \max\{\alpha, 1 - \beta\} > 0$. Therefore $l = \Omega(m)$.

Next we show that as x varies in $\{1, 2, \dots, l\}$, $A \cdot x \bmod q$ takes l distinct values. Suppose that there are $x, y \in \{1, 2, \dots, l\} (x \neq y)$ such that $A \cdot x \bmod q = A \cdot y \bmod q$. Then for all $i \in \{1, \dots, r\}$ we have $a_i(x - y) \bmod q_i = 0$, which means $(x - y) \bmod q_i = 0$ since $\gcd(a_i, q_i) = 1$. Therefore $(x - y) \bmod (\operatorname{lcm}(q_1, \dots, q_r)) = 0$, i.e., $(x - y) \bmod l = 0$, a contradiction. So $A \cdot x \bmod q$ takes l distinct values as x varies in $\{1, 2, \dots, l\}$. This means that as x varies in $\{1, 2, \dots, m\}$, $A \cdot x \bmod q$ takes $\min\{m, l\}$ distinct values, so the space complexity of maintaining $A \cdot x \bmod q$ is at least $\Omega(\log(\min\{m, l\})) = \Omega(\log m)$, which is a contradiction. ◀

The following $\Omega(\log m)$ lower bounds are corollaries of Theorem 3.6.

▶ **Theorem 3.7** (Additive error norm approximation). *For any constants $p > 0$ and $0 \le \epsilon < \frac{1}{2}$, any 1-pass streaming algorithm which, for any input $x \in \{-m, -(m-1), \dots, m\}^n$, outputs an approximation of $\|x\|_p$ in the interval $\left[\|x\|_p - \epsilon n^{1/p} m, \|x\|_p + \epsilon n^{1/p} m\right]$ with probability greater than $\frac{11}{12}$ in the turnstile model uses $\Omega(\log m)$ bits of space.*

Proof. Suppose that the theorem does not hold. Then for *any* sufficiently small constant $\eta > 0$, there exists m and an algorithm $\mathcal{A}$ such that $\mathcal{A}$ uses less than $\eta \log m$ bits of space and estimate $\|x\|_p$ (for any $x \in \{-m, -(m-1), \dots, m\}^n$) up to additive $\epsilon n^{1/p} m$ error with probability $1 - \delta$, where $0 \le \delta < \frac{1}{12}$. Below, we make use of $\mathcal{A}$ to solve the norm decision problem in $\eta \log m$ bits of space and thus reach a contradiction to Theorem 3.6.

[8] According to Theorem 3.4, $\mathcal{B}$ can only solve the problem with parameter $m/2$ instead of m. Since we are proving an $\Omega(\log m)$ lower bound, we can replace $m/2$ by m for simplicity.

We invoke $\mathcal{A}$ to solve the norm decision problem with parameters $\alpha < \frac{1}{2} - \epsilon$, $\beta > \frac{1}{2} + \epsilon$, and δ. We can choose α and β such that $\alpha = 1 - \beta$, then we have $\delta < \frac{1}{12} = \frac{\min\{\alpha, 1-\beta\}}{6(\alpha+1-\beta)}$ as required in Theorem 3.6. When $\|x\|_p \leq \alpha n^{1/p} m$, a successful estimate for $\|x\|_p$ given by $\mathcal{A}$ will be at most $(\alpha + \epsilon) n^{1/p} m$; when $\|x\|_p \geq \beta n^{1/p} m$, a successful estimate for $\|x\|_p$ given by $\mathcal{A}$ will be at least $(\beta - \epsilon) n^{1/p} m$. Since $\alpha + \epsilon < \frac{1}{2} < \beta - \epsilon$, by looking at the most significant $O(1)$ bits of the output of $\mathcal{A}$, we are able to tell whether the output is at most $(\alpha + \epsilon) n^{1/p} m$ or at least $(\beta - \epsilon) n^{1/p} m$, and thus to decide whether $\|x\|_p \leq \alpha n^{1/p} m$ or $\|x\|_p \geq \beta n^{1/p} m$ (with probability at least $1 - \delta$). This solves the norm decision problem using $\eta \log m$ bits of space, contradicting Theorem 3.6. ◀

▶ **Theorem 3.8** (Approximating a counter up to additive error). *For any constant $0 \leq \epsilon < \frac{1}{2}$, any 1-pass algorithm which, for any input $x \in \{-m, -(m-1), \ldots, m\}$, outputs $|x|$ up to additive ϵm error with probability larger than $\frac{11}{12}$ in the turnstile model uses $\Omega(\log m)$ bits of space.*

Proof. This is a special case of Theorem 3.7 with $n = 1$. ◀

▶ **Theorem 3.9** (Maintaining the most significant bit of a counter). *Any 1-pass algorithm which, for any input $x \in \{-m, -(m-1), \ldots, m\}$, outputs the most significant bit of $|x|$ with probability larger than $\frac{11}{12}$ in the turnstile model uses $\Omega(\log m)$ bits of space.*

Proof. Without loss of generality, we assume $m = 2^k - 1 (k \in \mathbb{Z}_+)$. In this case $|x|$ has k bits. If an algorithm can output the most significant bit of $|x|$, it must be able to distinguish whether $|x| \leq \frac{1}{4} m$ or $|x| \geq \frac{3}{4} m$: the most significant bit of $|x|$ is 0 in the former case, and is 1 in the latter case. Then the $\Omega(\log m)$ lower bound follows from Theorem 3.6. ◀

4 Reduction to Linear Sketches in the Strict Turnstile Model

In this section, we show that in the strict turnstile model, there is also an equivalence between general algorithms and linear sketches, similar to the LNW reduction for the turnstile model. We will consider algorithms that allow input streams from Λ_m^*, which is the set of streams such that the underlying vector never has a negative entry and is in $\{0, 1, \ldots, m\}^n$ at the end of the stream. We further assume that the algorithms have space complexity depending only on the dimension n, which is suitable when we want to prove lower bounds as functions of n, such as in graph problems.

The following theorem is an adaptation of [9, Theorem 10] for the strict turnstile model. It implies that to obtain lower bounds depending only on dimension in the strict turnstile model, it suffices to consider linear sketches, or the simultaneous communication model.

▶ **Theorem 4.1** (Reduction in the strict turnstile model). *Suppose that a randomized algorithm $\mathcal{A}$ solves a problem P on any stream in Λ_m^* with probability at least $1 - \delta$, and that the space complexity of $\mathcal{A}$ depends only on n. Then there exists an algorithm $\mathcal{B}$ implemented by maintaining a linear sketch $A \cdot x \bmod q$ in the stream, where A is a random $r \times n$ integer matrix and q is a positive integer vector of length r, such that $\mathcal{B}$ solves P on any stream in Λ_m^* with probability at least $1 - 6\delta$ and that the space used by any deterministic instance of $\mathcal{B}$ is no more than the space used by $\mathcal{A}$.*

Proof. We modify the reduction from general automaton to path-independent automaton (which was only claimed to be path-reversible in [9], as mentioned in the beginning of this section).

We view $\mathcal{A}$ as an automaton. Since the space used by $\mathcal{A}$ depends only on n, there is a function g such that the number of states of every deterministic instance of $\mathcal{A}$ is no more than $g(n)$. As in Section 3.1, let L be the maximum length of the shortest zero-frequency path between any two states in the transition graph of any deterministic instance of $\mathcal{A}$. From Lemma 3.3 we know that $L \leq h(n)$ for some function h.

Let γ be a fixed stream with frequency $(h(n), \dots, h(n))$, which consists of only positive updates (i.e., $+e_i$'s). We construct another automaton $\mathcal{A}'$ as follows: for any randomness, (1) $\mathcal{A}'$ has the same transition graph as $\mathcal{A}$; (2) the starting state of $\mathcal{A}'$ is $o \oplus \gamma$, where o is the starting state of $\mathcal{A}$; (3) for any state u, the output of $\mathcal{A}'$ on u is the output of $\mathcal{A}$ on the state $u \oplus \gamma^{-1}$.

It is easy to see that executing a stream σ on $\mathcal{A}'$ is equivalent to running the stream $\gamma \circ \sigma \circ \gamma^{-1}$ on $\mathcal{A}$. Since $\mathcal{A}$ succeeds in solving P on any stream in Λ_m^* with probability at least $1 - \delta$, we know that $\mathcal{A}'$ solves P on a stream σ with probability $1 - \delta$ as long as $\gamma \circ \sigma \circ \gamma^{-1} \in \Lambda_m^*$, i.e., $\mathcal{A}'$ solves P on any stream in the set $\Phi = \{\|\text{freq } \sigma\|_\infty \leq m, \text{freq } \sigma \geq \vec{0}$, and the frequency of any prefix of σ has all its coordinates at least $-h(n)\}$ with probability $1 - \delta$.

Now we invoke the LNW reduction from $\mathcal{A}'$ to a path-reversible automaton $\mathcal{C}$. According to Appendix A, $\mathcal{C}$ is as well a path-independent automaton. Recall that when a stream $\sigma \in \Lambda_m^*$ is executed by $\mathcal{C}$, equivalently another stream σ' is executed by $\mathcal{A}'$, where σ' is obtained by inserting zero-frequency streams into σ. Note that we can assume that all the inserted zero-frequency streams have length at most $L \leq h(n)$, and then $\sigma \in \Lambda_m^*$ implies $\sigma' \in \Phi$. Therefore we have a path-independent automaton $\mathcal{C}$ solving P on any stream in Λ_m^* (with high probability). ◀

Maximum Matching. In a recent work [2], tight upper and lower bounds are shown for turnstile algorithms that approximate maximum matching in dynamic graph streams, but the lower bound is only proved in the simultaneous communication model. Using Theorem 4.1, we are able to conclude that their results hold for any 1-pass algorithm and thus to resolve the 1-pass space complexity of this problem. Namely, to compute an n^ϵ-approximate maximum matching, $\Theta(n^{2-3\epsilon})$ bits of space is both sufficient and necessary (up to polylogarithmic factors), where n is the number of vertices.

5 Reduction for Multi-Pass Automata

Our main result in this section is that the LNW reduction can be extended to multi-pass automata, i.e., that a randomized p-pass automaton can be reduced to a path-independent one without blowing up the space complexity. Throughout this section p is a constant.

The main difficulty of this reduction is that when we consider an automaton in the i-th pass, for $i > 1$, we have to restrict to a subset of input streams that lead to the same state in the automaton processing the previous pass. Fortunately there is still sufficient randomness remaining even with this restriction so that the padding argument with zero-frequency streams in the LNW reduction still works.

▶ **Theorem 5.1** (Reduction of p-pass automata). *Let $\mathcal{A}$ be a p-pass randomized automaton that solves P with probability $\geq 1 - \delta$. Let $\epsilon > 0$. For any distribution Π over streams, there exists a p-pass path-independent deterministic automaton $\mathcal{B}$ that solves P over input drawn from Π with probability $\geq 1 - \delta - \epsilon$. Furthermore, $S(\mathcal{B}, m) \leq S(\mathcal{A}, m)$.*

Proof for $p = 2$. We first give a detailed proof for the case $p = 2$. The same method can be easily generalized to larger p.

Let S_0 be a set of zero-frequency streams such that whenever o_1 and o_2 are two states of $\mathcal{A}^1$ or of any automaton $\mathcal{A}_s^2$, there exists $\sigma \in S_0$ such that $o_1 \oplus \sigma = o_2$, where $\oplus$ is the transition function of the corresponding automaton.

Define a distribution Π' as follows. For a stream $\sigma \sim \Pi$ and $\sigma_0 = (\sigma_1, \ldots, \sigma_{2W}) \sim \mathrm{Unif}(S_0^{2W})$, we include $\sigma \otimes \sigma_0 := \sigma_1 \circ \ldots \circ \sigma_W \circ \sigma \circ \sigma_{W+1} \ldots \circ \sigma_{2W}$ in Π'. Here for any set S, $\mathrm{Unif}(S)$ is defined to be the uniform distribution over S. We shall choose W to be sufficiently large so that certain conditions are satisfied. The conditions will be described explicitly later in the proof.

By Yao's minimax principle, we can pick a deterministic instance of $\mathcal{A}$, also denoted by $\mathcal{A}$, which is correct on input distribution Π' with probability $\geq 1 - \delta$. Henceforth in the proof $\mathcal{A}$ refers to this deterministic instance. Everything is similar to [9] so far.

For a state s of $\mathcal{A}^1$, denote by $\Pi'(s)$ the marginal distribution of Π' on the event that reading the input stream in $\mathcal{A}^1$ ends at state s. By the correctness assumption of $\mathcal{A}$, it holds that

$$\mathop{\mathbb{E}}_{\sigma' \in \Pi'} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_{\sigma'}}(\sigma') \text{ is acceptable for } \sigma'\right\}} \geq 1 - \delta,$$

or, equivalently,

$$\mathop{\mathbb{E}}_{s \sim \mu} \mathop{\mathbb{E}}_{\sigma' \sim \Pi'(s)} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_{\sigma'}}(\sigma') \text{ is acceptable for } \sigma'\right\}} \geq 1 - \delta,$$

where μ the distribution over the states of $\mathcal{A}^1$ induced by Π'.

For each second-pass automaton $\mathcal{A}_s^2$ (s is a state in $\mathcal{A}^1$), we reduce it to a path-independent automaton $\mathcal{B}_s^2$ with the same transition functions as in the LNW reduction. Since $\mathcal{A}_s^2$ will only be run on the input streams in $\mathrm{supp}(\Pi'(s))$, we may assume that the states of $\mathcal{B}_s^2$ are all the terminal equivalence classes $\langle \tau' \rangle$ of $\mathcal{A}_s^2$, where $\tau' \in \mathrm{supp}(\Pi'(s))$.

To specify the output on the terminal equivalence class, we need the following proposition, whose proof is postponed to the end of this section.

▶ **Proposition 5.2.** *Let $\epsilon > 0$ and W be a sufficiently large integer. Suppose that $\sigma = \sigma_1 \circ \ldots \circ \sigma_W \sim \mathrm{Unif}(S_0^W)$ and C is a terminal equivalence class of some automaton. Let s_0 and s be arbitrary states in C and let event $\mathcal{E} = \{s_0 \oplus \sigma = s\}$. There exist a positive integer $L \leq W$ and a distribution $\mathcal{D}$ over S_0^{W-L} (both L and $\mathcal{D}$ are independent of s_0 and s) such that*

$$d_{TV}(\mathcal{L}(\sigma_{L+1} \circ \ldots \circ \sigma_W | \mathcal{E}), \mathcal{D}) \leq \epsilon,$$

where $\mathcal{L}(\sigma_{L+1} \circ \ldots \circ \sigma_W | \mathcal{E})$ is the conditional distribution of $\sigma_{L+1} \circ \ldots \circ \sigma_W$ on the event $\mathcal{E}$. Furthermore, there exists an integer $R \in [L, W]$ independent of s_0 and s such that $d_{TV}(\mathcal{L}(\sigma_{L+1} \circ \ldots \circ \sigma_R | \mathcal{E}), \mathrm{Unif}(S_0^{R-L})) \leq \epsilon$, and $R - L$ can be made arbitrarily large.

Now we specify the output of $\mathcal{B}_s^2$. Let $\langle t \rangle$ be a terminal class of $\mathcal{A}_s^2$ and $\Pi'(s, \langle t \rangle)$ be the marginal distribution of $\Pi'(s)$ on the streams terminating in $\langle t \rangle$. The random output on $\langle t \rangle$ is defined as $\phi_{\mathcal{A}_s^2}(\tau')$ with $\tau' \sim \Pi'(s, \langle t \rangle)$.

For a stream prefix ρ, we denote by $\Pi'(s, \rho)$ the marginal distribution of $\Pi'(s)$ on the streams with prefix ρ.

We choose W *sufficiently large* such that the following two conditions hold.

(A) With probability $\geq 1 - \epsilon$ over $\sigma' \sim \Pi'$, the second-pass automaton $\mathcal{A}_{\sigma'}^2$ arrives at a state in a terminal equivalence class on input stream σ'.

(B) Conditioned on (A), for any second-pass automaton $\mathcal{A}^2_s$, and for any stream prefixes ρ_1 and ρ_2 of streams in $\operatorname{supp}(\Pi'(s))$ that satisfy (i) ρ_i has the form $\sigma_1 \circ \ldots \circ \sigma_W \circ \sigma \circ \sigma_{W+1} \circ \ldots \circ \sigma_{W+n_i}$ for some $0 \le n_i \le W$ $(i = 1, 2)$ and (ii) ρ_1 and ρ_2 arrive in the same equivalence class of $\mathcal{A}^2_s$, the induced distribution on the terminating states of streams in $\Pi'(s, \rho_1)$ and that on the terminating states of streams in $\Pi'(s, \rho_2)$ are ϵ-close in total variation distance.

Condition (A) is possible by Proposition 5.2, which indicates that there is a sufficiently long random walk in $\mathcal{A}^2_s$, so starting from a node outside any terminating equivalence class, it will arrive at a state in a terminal equivalence class with a high probability; then take a union bound. We shall be conditioned on (A). Condition (B) is possible again because of Proposition 5.2. The streams with prefix ρ_1 and the streams with prefix ρ_2 will be close to $\operatorname{Unif}(S_0^L)$ on a segment of length L (where L can be made arbitrarily large) so they will first mix in the terminal equivalence class and the streams have similar distribution afterwards. Therefore, the induced distribution on the terminating states of streams in $\Pi'(s, \rho)$ is close to that induced by $\Pi'(s, \langle\rho\rangle)$.

Furthermore, when W is large enough, the random zero-padding $\sigma_1 \circ \ldots \circ \sigma_W$ before $\sigma \sim \Pi$ always leads to a state in a (random) equivalence class in $\mathcal{A}^2_s$. Choose the initial state of $\mathcal{B}^2_s$ according to the induced distribution on the terminal equivalence classes by streams drawn from $\Pi'(s)$. It follows that on reading $\sigma' \sim \Pi'$, the distribution on the terminating equivalence classes in $\mathcal{A}^2_{\sigma'}$ is ϵ-close to the distribution on the corresponding states in $\mathcal{B}^2_{\sigma'}$. They are not necessarily the same distribution because we choose a random initial state in $\mathcal{B}^2_s$. This is called the terminal class property.

To show the correctness of $\mathcal{B}$, we need to show (we may rescale ϵ if necessary)

$$\mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\text{randomness of } \mathcal{B}} \mathbf{1}_{\{\phi_{\mathcal{B}}(\sigma) \text{ is acceptable for } \sigma\}} \ge 1 - \delta - O(\epsilon). \tag{1}$$

We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\text{randomness of } \mathcal{B}} \mathbf{1}_{\{\phi_{\mathcal{B}}(\sigma) \text{ is acceptable for } \sigma\}} \\
&= \mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{s\sim\text{Stationary}(\langle\sigma\rangle)}\ \mathop{\mathbb{E}}_{\text{randomness of } \mathcal{B}^2_s} \mathbf{1}_{\{\phi_{\mathcal{B}^2_s}(\sigma) \text{ is acceptable for } \sigma\}} \\
&\ge \mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\sigma_0\sim\operatorname{Unif}(S_0^{2W})}\ \mathop{\mathbb{E}}_{\text{randomness of } \mathcal{B}^2_{\sigma\otimes\sigma_0}} \mathbf{1}_{\left\{\phi_{\mathcal{B}^2_{\sigma\otimes\sigma_0}}(\sigma) \text{ is acceptable for } \sigma\right\}} - \epsilon \\
&\ge \mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\sigma_0\sim\operatorname{Unif}(S_0^{2W})}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(\sigma\otimes\sigma_0, \langle\sigma\otimes\sigma_0\rangle)} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_{\sigma\otimes\sigma_0}}(\tau') \text{ is acceptable for } \sigma\right\}} - 2\epsilon.
\end{aligned} \tag{2}$$

In the above, line 2 follows from the random output of $\mathcal{B}^1$, line 3 from the fact that $\sigma \otimes \sigma_0$ with σ_0 is ϵ-close to the stationary distribution on $\langle\sigma\rangle$, line 4 from the definition of the random output of $\mathcal{B}^2$ and the terminal class property.

The event of the indicator function in (2) has a slight mismatch: the input stream is τ' while we only know the automaton's correctness on σ. To overcome this, we break up the streams in $\operatorname{supp}(\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle))$ according to the frequency vectors v. We say $\operatorname{freq}(\tau')$ is admissible for $\tau' \in \operatorname{supp}(\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle))$. Further conditioned on admissible v, the conditional distribution of $\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle)$ is denoted by $\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle, v)$. By condition (B), when W is sufficiently large, for any admissible v, the distribution of states in $\langle\sigma \otimes \sigma_0\rangle$ induced by $\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle, v)$ is close to that induced by $\Pi'(\sigma \otimes \sigma_0, \langle\sigma \otimes \sigma_0\rangle)$. It therefore holds that

$$\left| \mathop{\mathbb{E}}_{\tau'\sim\Pi'(\sigma\otimes\sigma_0, \langle\sigma\otimes\sigma_0\rangle)} f(\tau') - \mathop{\mathbb{E}}_{\tau'\sim\Pi'(\sigma\otimes\sigma_0, \langle\sigma\otimes\sigma_0\rangle, \operatorname{freq}(\sigma))} f(\tau') \right| \le \epsilon, \tag{3}$$

for all (measurable) f with $\|f\|_\infty \le 1$. Now we can continue from (2):

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\text{randomness of }\mathcal{B}} \mathbf{1}_{\{\phi_{\mathcal{B}}(\sigma)\text{ is acceptable for }\sigma\}}\\
&\ge \mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\sigma_0\sim\mathrm{Unif}(S_0^{2W})}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(\sigma\otimes\sigma_0,\langle\sigma\otimes\sigma_0\rangle,\mathrm{freq}(\sigma))} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_{\sigma\otimes\sigma_0}}(\tau')\text{ is acceptable for }\sigma\right\}} - 3\epsilon \text{ (using (3))}\\
&= \mathop{\mathbb{E}}_{\sigma'\sim\Pi'}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(\sigma',\langle\sigma'\rangle,\mathrm{freq}(\sigma'))} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_{\sigma'}}(\tau')\text{ is acceptable for }\tau'\right\}} - 3\epsilon \text{ (correctness depends only on freq. vec.)}\\
&= \mathop{\mathbb{E}}_{s\sim\mu}\ \mathop{\mathbb{E}}_{\sigma'\sim\Pi'(s)}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(s,\langle\sigma'\rangle,\mathrm{freq}(\sigma'))} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_s}(\tau')\text{ is acceptable for }\tau'\right\}} - 3\epsilon\\
&\ge \mathop{\mathbb{E}}_{s\sim\mu}\ \mathop{\mathbb{E}}_{\sigma'\sim\Pi'(s)}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(s,\langle\sigma'\rangle)} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_s}(\tau')\text{ is acceptable for }\tau'\right\}} - 4\epsilon \text{ (using (3))}\\
&= \mathop{\mathbb{E}}_{s\sim\mu}\ \mathop{\mathbb{E}}_{\tau'\sim\Pi'(s)} \mathbf{1}_{\left\{\phi_{\mathcal{A}^2_s}(\tau')\text{ is acceptable for }\tau'\right\}} - 4\epsilon\\
&\ge 1-\delta-4\epsilon.
\end{aligned}$$

Removing the conditioning on (A) causes a further loss of ϵ in the success probability, that is,

$$\mathop{\mathbb{E}}_{\sigma\sim\Pi}\ \mathop{\mathbb{E}}_{\text{randomness of }\mathcal{B}} \mathbf{1}_{\{\phi_{\mathcal{B}}(\sigma)\text{ is acceptable for }\sigma\}} \ge 1-\delta-5\epsilon.$$

This completes the proof of (1).

Finally, by an averaging argument, there exists a deterministic automaton $\mathcal{B}$ achieving success probability at least as high as that of the randomized $\mathcal{B}$. The claim of space complexity follows from the same argument as in [9]. ◀

Proof of Theorem 5.1 for general p. We only describe the major changes on the proof for the special case $p = 2$. Here we choose W sufficiently large such that:

(A) With probability $\ge 1 - \Theta(\epsilon)$ over $\sigma' \sim \Pi'$, the automaton $\mathcal{A}^q_{\sigma'}$ for all $2 \le q \le p$ arrives at a state in a terminal equivalence class on input stream σ'.

(B) Over $\sigma' \sim \Pi'$, for any q-th pass automaton $\mathcal{A}^q_{\sigma'}$ $(2 \le q \le p)$, the induced distribution on terminating states of $\mathcal{A}^q_{\sigma'}$ is $\Theta(\epsilon)$-close to that on corresponding states of $\mathcal{B}^2_{\sigma'}$ on reading σ'. This is the terminal class condition.

(C) (Conditioned on (A)) For any q-th pass automaton $\mathcal{A}^q_s$, and for any stream prefixes ρ_1 and ρ_2 of streams ending at the same state s of $\mathcal{A}^{q-1}$ such that both ρ_1 and ρ_2 arrive in the same equivalence class of $\mathcal{A}^q_s$, the induced distribution on the terminating states of streams with prefix ρ_1 and that on the terminating states of streams with prefix ρ_2 are $\Theta(\epsilon)$-close in total variation distance.

The random output is drawn from the stationary distribution on the associated equivalence class for the first-pass automaton, and is, for all subsequent passes $i \ge 2$, drawn from the induced distribution on the states of $\mathcal{A}^i_s$ by the conditional distribution of the streams terminating at s in the automaton of pass $i - 1$. A similar argument to that in the case of $p = 2$ gives the result. ◀

After we have Theorem 5.1, similar to [9], we can use Yao's minimax principle to conclude the existence of a randomized p-pass automaton that succeeds with probability $\ge 1 - \delta - \epsilon$ on any input, and can further use Newman's argument [11] to reduce the number of random bits to $O(\log n + \log\log m + \log \frac{1}{\epsilon})$.

Now we give the proof of Proposition 5.2.

Proof of Proposition 5.2. Let π be the stationary distribution on C under the transition probability induced by $\mathrm{Unif}(S_0)$. Note that

$$\begin{aligned}
&\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|s_0\oplus\sigma_1\circ\ldots\circ\sigma_W=s\}\\
=&\sum_{t\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_W=s,s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}\cdot\\
&\quad\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}\\
=&\sum_{t\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_W=s\}\cdot\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}\\
=&\sum_{t\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_W=s\}\cdot\left(\pi(t)\pm\frac{\epsilon}{|C|}\right),
\end{aligned}$$

where line 3 follows from the Markov property of the process, line 4 follows from the fact that L can be chosen large enough so that $s_0\oplus\sigma_1\circ\ldots\circ\sigma_L$ mixes on C. Furthermore, L can be chosen independent of s_0 because there are only finitely many distinct s_0's. Define the probability distribution $\mathcal{D}$ as

$$\begin{aligned}
&\Pr_{\sigma_{L+1}\circ\ldots\circ\sigma_W\sim\mathcal{D}}\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau\}\\
&\qquad=\mathop{\mathbb{E}}_{t\sim\pi}\Pr_{\sigma_{L+1}\circ\ldots\circ\sigma_W\sim\mathrm{Unif}(S_0^{W-L})}\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_W=s\}.
\end{aligned}$$

It is easy to verify that $\mathcal{D}$ is indeed a probability distribution. It follows that

$$\begin{aligned}
&d_{TV}(\mathcal{L}(\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|s_0\oplus\sigma_1\circ\ldots\circ\sigma_W=s),\mathcal{D})\\
\le&\frac{\epsilon}{|C|}\sum_{t\in C}\sum_{\tau}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_W=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_W=s\}\\
=&\frac{\epsilon}{|C|}\cdot|C|=\epsilon.
\end{aligned}$$

For the second part, note that we have (similar to the above)

$$\begin{aligned}
&\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|s_0\oplus\sigma_1\circ\ldots\circ\sigma_W=s\}\\
=&\sum_{t,t'\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|t\oplus\tau=t',s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t,t'\oplus\sigma_{R+1}\circ\ldots\circ\sigma_W=s\}\cdot\\
&\quad\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t,t'\oplus\sigma_{R+1}\circ\ldots\circ\sigma_W=s\}
\end{aligned}$$

Now, t is the last state of a random walk from s_0 and t' is the last state of a random walk from s. The latter random walk is the reverse of $\sigma_{R+1}\circ\ldots\circ\sigma_W$ with all edges reversed in C (denoted the edge-reversed component by C'). It is clear that C' is strongly connected. Let π' be the stationary distribution on C' under the transition induced by $\mathrm{Unif}(S_0^r)$, where S_0^r denotes the reverse streams of S_0. If $L\gg 1$ and $R\ll W$, both walks $\sigma_1\circ\ldots\circ\sigma_W$ and $\sigma_W^r\circ\ldots\circ\sigma_{R+1}^r$ will mix. By Markov property,

$$\begin{aligned}
&\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t,t'\oplus\sigma_{R+1}\circ\ldots\circ\sigma_W=s\}\\
=&\Pr\{t'\oplus\sigma_{R+1}\circ\ldots\circ\sigma_W=s|s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}\cdot\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}\\
=&\Pr\{t'\oplus\sigma_{R+1}\circ\ldots\circ\sigma_W=s\}\Pr\{s_0\oplus\sigma_1\circ\ldots\circ\sigma_L=t\}
\end{aligned}$$

which can be made close to $\pi(t)\pi'(t)$ with an additive error at most $\epsilon/|C|^2$ with choice of L and R uniformly over s_0 and s, as there are only finitely many distinct s_0's and s's. It follows that

$$\begin{aligned}
&\left|\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|s_0\oplus\sigma_1\circ\ldots\circ\sigma_W=s\}-\frac{1}{|S_0|^{R-L}}\right|\\
\le&\left|\sum_{t,t'\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_R=t'\}\pi(t)\pi'(t')-\frac{1}{|S_0|^{R-L}}\right|\\
&\qquad+\frac{\epsilon}{|C|^2}\sum_{t,t'\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_R=t'\}\\
=&\frac{\epsilon}{|C|^2}\sum_{t,t'\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_R=t'\},
\end{aligned}$$

where we use the fact that

$$\sum_{t,t'\in C}\Pr\{\sigma_{L+1}\circ\ldots\circ\sigma_R=\tau|t\oplus\sigma_{L+1}\circ\ldots\circ\sigma_R=t'\}\pi(t)\pi'(t')=\frac{1}{|S_0|^{R-L}}.$$

To see this, imagine that $\sigma_{L+1}\circ\ldots\circ\sigma_R$ is a part of a two-sided infinitely long zero-frequency sequence. Finally, similar to before,

$$d_{TV}(\mathcal{L}(\sigma_{L+1}\circ\ldots\circ\sigma_R|\mathcal{E}),\mathrm{Unif}(S_0^{R-L}))\le\frac{\epsilon}{|C|^2}\cdot|C|^2=\epsilon. \qquad ◀$$

Acknowledgements. Yuqing Ai and Wei Hu were supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, and the National Natural Science Foundation of China Grant 61361136003. Yi Li was supported by ONR grant N00014-15-1-2388 when he was at Harvard University, where his major participation in this work took place. David Woodruff was supported in part by the XDATA program of the Defence Advanced Research Projects Agency (DARPA), administered through Air Force Research Laboratory contract FA8750-12-C-0323.

References

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

2 Sepehr Assadi, Sanjeev Khanna, Yang Li, and Grigory Yaroslavtsev. Maximum matchings in dynamic graph streams and the simultaneous communication model. In *SODA*, pages 1345–1364, 2016.

3 Christos Boutsidis, David P. Woodruff, and Peilin Zhong. Optimal principal component analysis in distributed and streaming models. In *STOC*, 2016.

4 Sumit Ganguly. Lower bounds on frequency estimation of data streams. In *Proceedings of the 3rd International Conference on Computer Science: theory and applications*, CSR'08, pages 204–215, 2008.

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

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

7 Piotr Indyk, Eric Price, and David P. Woodruff. On the power of adaptivity in sparse recovery. In *FOCS*, pages 285–294, 2011.

8 Daniel M. Kane, Jelani Nelson, and David P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *SODA*, pages 1161–1178, 2010.

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

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

11 Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letter*, pages 67–71, 1991.

12 Joachim von zur Gathen and Malte Sieveking. A bound on solutions of linear integer equalities and inequalities. In *Proceedings of the American Mathematical Society*, pages 155–158, 1978.

13 Omri Weinstein and David P. Woodruff. The simultaneous communication of disjointness with applications to data streams. In *ICALP*, pages 1082–1093, 2015.

14 David P. Woodruff. Low rank approximation lower bounds in row-update streams. In *NIPS*, pages 1781–1789, 2014.

15 Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *STOC*, pages 209–213, 1979.

A The LNW Reduction

We show that the reduction from a general automaton to a path-reversible automaton, presented in [9, Theorem 5], actually gives us a path-independent automaton.

The reduction works as follows. Let $\mathcal{A}$ be the original automaton, $G'_{\mathcal{A}}$ be its zero-frequency graph, and $\oplus$ be its transition function. The states of the new automaton $\mathcal{B}$ are defined to be the terminal strongly connected components of $G'_{\mathcal{A}}$. (A strongly connected component is terminal if there is no arc from it to the rest of the graph.) For each strongly connected component v of $G'_{\mathcal{A}}$, let $rep(v)$ be a (fixed) arbitrary vertex in v, and $\alpha(v)$ be a (fixed) arbitrary terminal strongly connected component reachable from v. For each vertex u in $G'_{\mathcal{A}}$, let $com(u)$ be the strongly connected component it belongs to. Then the transition function $\oplus'$ of $\mathcal{B}$ is defined as

$$v \oplus' \pm e_i = \alpha(com(rep(v) \oplus \pm e_i)),$$

where v is a state of $\mathcal{B}$, i.e., a terminal strongly connected component of $G'_{\mathcal{A}}$.

It is shown in [9, Lemma 6] that $\mathcal{B}$ is path-reversible:

▶ **Lemma A.1** (Lemma 6 in [9]). *For any state u of $\mathcal{B}$ and any $i \in [n]$, we have $u \oplus' e_i \circ -e_i = u$.*

We show a stronger result below, which implies $\mathcal{B}$ is path-independent.

▶ **Lemma A.2.** *For any state u of $\mathcal{B}$ and any zero-frequency stream σ, we have $u \oplus' \sigma = u$.*

Proof. Let $\sigma = (\sigma_1, \ldots, \sigma_t)$ ($\sigma_i \in \Sigma$) and $v = u \oplus' \sigma$. Then there exist zero-frequency streams $\gamma^1, \gamma^2, \ldots, \gamma^t$ such that

$$rep(u) \oplus \sigma_1 \circ \gamma^1 \circ \sigma_2 \circ \gamma^2 \circ \ldots \circ \sigma_t \circ \gamma^t = rep(v).$$

Note that $\mathrm{freq}(\sigma_1 \circ \gamma^1 \circ \sigma_2 \circ \gamma^2 \circ \ldots \circ \sigma_t \circ \gamma^t) = \mathrm{freq}(\sigma_1 \circ \sigma_2 \circ \ldots \circ \sigma_t) = \mathrm{freq}\, \sigma = \vec{0}$. Since $rep(u)$ belongs to a terminal strongly connected component u, $rep(v)$ has to be in the same terminal strongly connected component. Hence $u = com(rep(u)) = com(rep(v)) = v$. ◀

B An Alternative Upper Bound on the Length of Shortest Zero-Frequency Paths

▶ **Lemma B.1.** *Let $s = |V|$ be the number of vertices in the transition graph $G_{\mathcal{A}}$. Then $L \leq e^{O((s+n)\log s)}$, where L is the maximum length of the shortest zero-frequency path between any two vertices connected by at least one zero-frequency path.*

Proof. Consider two vertices $o_1, o_2 \in V$ such that there exists a zero-frequency path from o_1 to o_2. We fix a tuple (p, C) satisfying the following condition: every vertex in $c_1, \ldots, c_t$ can be reached from o_1 via arcs in $c_1, \ldots, c_t$ and p. Here p is a simple path from o_1 to o_2 and $C = \{c_1, c_2, \ldots, c_t\}$ is a set of simple cycles in $G_{\mathcal{A}}$. For every possible (p, C) satisfying this condition, we build a linear system $Ax = b$ and want a positive integer solution. Here $A = \begin{pmatrix} f_{c_1} & f_{c_2} & \ldots & f_{c_t} \end{pmatrix}$ is an $n \times t$ matrix and $b = -f_p \in \mathbb{R}^n$. The i-th row in the equations guarantees the i-th component in the frequency of the path is 0. Each positive integer solution x of the equations corresponds to a zero-frequency path from o_1 to o_2 using the simple path p and simple cycles in C. The path p is passed through exactly once and each cycle c_i is passed through x_i times.

By Lemma 3.1, if there exists a positive integer solution, then there exists such a solution x so that $\|x\|_\infty \leq (t+1)^2 M_1$, where M_1 is the largest possible absolute value of any subdeterminant of $\begin{pmatrix} A & b \end{pmatrix}$. Since $c_1, \ldots, c_t$ and p are simple, they all have length at most s. Thus, the sum of the absolute values of entries in any column of $\begin{pmatrix} A & b \end{pmatrix}$ is bounded by s. By the Gershgorin circle theorem, the eigenvalues of any submatrix of $\begin{pmatrix} A & b \end{pmatrix}$ have absolute value at most s. Therefore, M_1 is at most s^n. On the other hand, for the number of simple cycles in a graph with s vertices, we have $t = |C| \leq s^{O(s)}$. Therefore

$$\|x\|_\infty \leq (s^{O(s)} + 1)^2 \cdot s^n = s^{O(n+s)}.$$

Thus, the length of the corresponding path is bounded by

$$s\|x\|_1 \leq st\|x\|_\infty \leq s \cdot s^{O(s)} \cdot s^{O(n+s)} = s^{O(n+s)}.$$

Since there exists a zero-frequency path from o_1 to o_2, we can decompose it into the sum of a simple path from o_1 to o_2 and a linear combination of simple cycles. This leads to the existence of a positive integer solution for the equations $Ax = b$ when we fix this path as p and this set of cycles as C. Therefore, there exists a zero-frequency path from o_1 to o_2 whose length is at most $s^{O(n+s)}$. ◀