1 More on Streaming Lower Bounds

1.1 Distributional Communication Complexity

Recall the INDEX model,

1. Alice receives a binary string $x \in \{0, 1\}^n$, Bob receives an integer $j \in \{1, 2, \ldots, n\}$;
2. (1-way communication) Alice sends a single randomized message $M$ to Bob;
3. Bob outputs $b$, which is his guess of $x_j$.

And recall that the INDEX model has the lower bound for deterministic communication complexity:

$$CC^\delta(\text{INDEX}) \geq I(M; X|R) \geq n(1 - H(\delta)),$$

where $R$ is the shared common random string. We need a lower bound when conditioning on $R$ for our earlier Gap-Hamming lower bound, which was a reduction from INDEX using the shared common random string $R$.

Definition. Given $(X,Y) \sim \mu$, the $\mu$-distributional communication complexity of a function $f(X,Y)$ over the distribution $\mu$, denoted by $D_\mu(f)$, is the minimum cost of a protocol that gives the correct answer with probability at least 2/3.

Theorem 1. (Yao’s Minimax Principle) $R(f) = \max_\mu D_\mu(f)$.

Proof. It is easy to see that for all distributions $\mu$, we have

$$R(f) \geq D_\mu(f),$$

hence

$$R(f) \geq \max_\mu D_\mu(f).$$

For the other direction, choose $c$ such that $\max_\mu D_\mu(f) \leq c$. Consider the following 2 player zero-sum game. Player 1 chooses a deterministic protocol $P$ for $f$ of cost $c$ (and whatever error), and Player 2 chooses an input $(x,y)$. Both players make their choices in parallel, so that neither is aware of the other’s choice.

The payoff for Player 1 is $1_{[P(x,y)=f(x,y)]}$.

Now, the fact that $D_\mu(f) \leq c$ for every distribution $\mu$ implies that for every randomized strategy of Player 2 (i.e., for every probability distribution $\mu$), Player 1 can obtain expected payoff 2/3 using
the protocol $P$ of cost $D_\mu(f) \leq c$. By the min-max theorem for zero-sum games, Player 1 has a randomized strategy, with an expected payoff of $2/3$ for every choice of inputs of Player 2. Now note that a randomized strategy for Player 1 is a distribution over cost $c$ deterministic protocols, i.e., a public coins protocol of cost at most $c$. Thus, there exists a public coin protocol of cost at most $c$ that is correct on every input with probability at least $2/3$. Together with the definition of $\max_\mu(\cdot)$ we can conclude that

$$R(f) \leq \max_\mu D_\mu(f).$$

### 1.2 INDEX Problem with Product Distribution

**Definition.** The communication matrix $A_f$ of a Boolean function $f : X \times Y \to \{0,1\}$ is defined such that the $(x,y)$-th entry equal to $f(x,y)$.

**Theorem 2.** [7] If Alice and Bob are independent, i.e. $\mu$ is a product distribution, then

$$\max_{\text{product } \mu} D_\mu(f) = \Theta(\text{VC-dim of } A_f)$$

**Remark 1.** The reduction from INDEX is optimal for product distributions. Since for $x$ and $i$ are independent and uniformly distributed, jointly as $\mu^*$, we have

$$D_{\mu^*}(f) = \Omega(n) \leq \max_{\text{product } \mu} D_\mu(f) \leq \max_{\mu} D_\mu(f) = R(f).$$

### 1.3 Indexing with Low Error

The INDEX problem with $1/3$ error probability and 0 error probability both have $\Omega(n)$ communication complexity. But sometimes we expect to have a lower bound in terms of the error probability. So we considering the **INDEX on Large Alphabets** problem:

1. Alice receives a binary string $x \in \{0,1\}^{n/\delta}$ and $\text{wt}(x) = n$, Bob receives an integer $j \in [n/\delta]$;
2. Bob wants to decide if $x_j = 1$ with error probability $\delta$.

The 1-way communication complexity is

$$\log\left(\frac{n/\delta}{n}\right) = \log(\frac{n/\delta}{n})^n = \Omega(n \log(1/\delta)).$$

Then consider the case where $n = 1$,

1. Alice has a string $x$ in $\{0,1\}^{1/\delta}$ with exactly one coordinate $j$ equal to 1
2. Bob has an integer $i$ in $\{1,2,\ldots,1/\delta\}$.

Let $y$ be the underlying vector that the stream is run on, which is initialized to all 0s. Suppose the dimension of the vector $y$ is at least $1/\delta$. Then Alice creates the stream: $y_j \leftarrow y_j + 1$, and Bob creates the stream: $y_i \leftarrow y_i - 1$. At the end of the stream, we have $y = e_j - e_i$. If $j = i$,
then any norm of $y$ is 0. Otherwise, any norm of $y$ is non-zero. So a norm estimation data stream algorithm which succeeds with probability $1 - \delta$, can solve the **Indexing with low error** problem with probability $1 - \delta$. By the $\log(1/\delta)$ communication lower bound for indexing with low error shown above, we obtain a $\log(1/\delta)$ space lower bound for the data stream algorithm. It worth noticing that this reduction only works if the dimension of the vector $y$ is at least $1/\delta$. For the case when the dimension of $y$ is smaller than $1/\delta$, [6] shows how to get a lower bound of $\log(1/\delta)$.

In the last lecture, we saw an $\Omega(\log n)$ bit lower bound for norm estimation from the **Augmented Indexing communication** problem, and in the last lecture we saw an $\Omega(\epsilon^{-2})$ lower bound from the **Gap-Hamming** communication problem. Since **Indexing with Low Error** gives an $\Omega(\log(1/\delta))$ lower bound, in total we have an $\Omega(\log n + \epsilon^{-2} + \log(1/\delta))$ lower bound, since the lower bounds add. In fact it is known how to get a tighter lower bound of $\Omega(\epsilon^{-2} \log(1/\delta) \log n)$, that is, the three lower bounds we showed in class actually multiply [6].

Sometimes reduction to product distribution may not necessarily be optimal, since

$$\max_{\mu} D_{\mu}(f) \gg \max_{\text{product}\mu} D_{\mu}(f).$$

For example, consider the **Set disjointness** problem

1. Alice chooses a set $S \subset \{1, \ldots, n\}$
2. Bob chooses a set $T \subset \{1, \ldots, n\}$
3. Output 1 if $S \cap T = \emptyset$

It is known that for any deterministic protocol for solving the above problem has lower bound $\Omega(n)$, but for product distribution, $\max_{\text{product}\mu} D_{\mu} = \Omega(\sqrt{n} \log n)$ [2] [4].

### 1.4 Gap$_\infty(x, y)$ Problem and Direct Sums

The Gap$_\infty(x, y)$ problem is described as:

1. Alice has $x \in \{0, \ldots, B\}^n$, Bob has $y \in \{0, \ldots, B\}^n$
2. We are sure that $|x - y|_\infty \leq 1$ or $|x - y|_\infty \geq B$
3. Output 1 if $|x - y|_\infty \leq 1$ and 0 otherwise

It is shown that the Gap$_\infty(x, y)$ problem does not have a hard product distribution, but has a hard distribution $\mu = \lambda^n$ where the coordinate pairs $(x_1, y_1), \ldots, (x_n, y_n)$ are independent, and the distribution $\lambda$ is

1. with probability $1 - 1/n$, $(x, y)$ random subject to $|x - y|_\infty \leq 1$
2. with probability $1/n$, $(x, y)$ random subject to $|x - y|_\infty \geq B$
Hence
\[ \mu(x, y) = \prod_{i=1}^{n} \lambda(x_i, y_i). \]

Therefore, in order to solve $\text{Gap}_\infty(x, y)$ problem, we need to solve the single coordinate sub-problem $g$ for $n$ times, where $g$ is

1. Alice has $J \in \{0, \ldots, B\}$, Bob has $K \in \{0, \ldots, B\}$
2. We are sure that $|J - K|_\infty \leq 1$ or $|J - K|_\infty \geq B$
3. Output 1 if $|J - K|_\infty \leq 1$ and 0 otherwise

Define $IC(g) = \inf_\psi I(\psi; J, K)$, where $\psi$ ranges over all 2/3-correct 1-way protocols for $g$. This is usually referred as the **Direct Sum** method.

Let $\Pi$ be the message from Alice to Bob, concatenated with Bob’s output. For $(X, Y) \sim \mu$, the information cost of the protocol is

\[
I(\Pi; X, Y) = \sum_i I(\Pi; (X_i, Y_i)|X_{<i}, Y_{<i})
\]

\[
= \sum_i H(X_i, Y_i) - H(X_i, Y_i|X_{<i}, Y_{<i}, \Pi)
\]

\[
\geq \sum_i H(X_i, Y_i) - H(X_i, Y_i|\Pi)
\]

\[
= \sum_i I(\Pi|X_i, Y_i).
\]

So we only need to show that $I(\Pi; X_i, Y_i) \geq IC(g)$ for each $i = 1, \ldots, n$.

Now we choose a specific joint distribution of $\lambda$, such that we always have $|X - Y|_\infty \leq 1$. It may be counterintuitive that $\lambda$ always has $|X - Y|_\infty \leq 1$, but since $\Pi$ must be correct on all inputs, the information measured with respect to $\lambda$ will still turn out to be large. Define $D = ((P_1, V_1), \ldots, (P_n, V_n)) = (P, V)^n$,

1. $P_j$ uniform on \{Alice, Bob\}
2. $V_j$ uniform on \{1, \ldots, B\} if $P_j$ is Alice, $V_j$ uniform on \{0, \ldots, B - 1\} if $P_j$ is Bob
3. If $P_j$ is Alice, then $Y_j = V_j$ and $X_j$ is uniform on \{\( V_j - 1, V_j \)\}; If $P_j$ is Bob, then $X_j = V_j$ and $Y_j$ is uniform on \{\( V_j, V_j + 1 \)\}

It is worth noticing that $X$ and $Y$ are independent conditioned on $D$. In this case, we have

\[
I(\Pi; X, Y|D) = \Omega(n)IC(g|(P, V)),
\]

where $IC(g|(P, V)) = \inf_\psi I(\psi; J, K|(P, V))$, $\psi$ ranges over all 2/3-correct protocols for $g$. Notice that for fixed $P = \text{Alice}$ and $V = v$, this is $I(\psi; K)$ where $K$ is uniform on $\{v - 1, v\}$, and

\[
I(\psi; K) \geq D_{JS}(\psi_{v-1,v}, \psi_{v,v}).
\]
Remark 2. Recall the properties for Hellinger Distance

1. $D_{JS}(\psi_{v-1,v}, \psi_{v,v}) \geq h(\psi_{v-1,v}, \psi_{v,v})$;
2. $h^2(\psi_{0,0}, \psi_{0,B}) = \Omega(1)$;
3. For 1-way protocol: $\psi_{a,b}(m, \text{out}) = p_a(m)q_{b,m}(\text{out})$;
4. $h^2(\psi_{a,b}, \psi_{c,d}) \geq 1/2 [h^2(\psi_{a,b}, \psi_{a,d}) + h^2(\psi_{c,b}, \psi_{c,d})]$.

Since

$$\frac{1}{2} \left[(1 - h^2(\psi_{a,b}, \psi_{a,d})) + (1 - h^2(\psi_{c,b}, \psi_{c,d}))\right]$$

$$= \frac{1}{2} \sum_{m, \text{out}} \left[\sqrt{p_a(m)q_{b,m}(\text{out})} \sqrt{p_a(m)q_{d,m}(\text{out})} + \sqrt{p_c(m)q_{b,m}(\text{out})} \sqrt{p_c(m)q_{d,m}(\text{out})}\right]$$

$$= \sum_{m, \text{out}} \frac{p_a(m) + p_c(m)}{2} \sqrt{q_{b,m}(\text{out})q_{d,m}(\text{out})}$$

$$\geq \sum_{m, \text{out}} \sqrt{p_a(m)q_{b,m}(\text{out})p_c(m)q_{d,m}(\text{out})}$$

$$= 1 - h^2(\psi_{a,b}, \psi_{c,d}).$$

Based on the above properties, we have

$$IC(g((P,V)) \geq \frac{1}{2} \sum_{v \in \{0,\ldots,B-1\}} E_v \left[D_{JS}(\psi_{v-1,v}, \psi_{v,v})\right] + \frac{1}{2} \sum_{v \in \{0,\ldots,B-1\}} E_v \left[D_{JS}(\psi_{v,v}, \psi_{v,v+1})\right]$$

$$\geq \frac{1}{2} \sum_{v \in \{0,\ldots,B-1\}} E_v \left[h^2(\psi_{v-1,v}, \psi_{v,v})\right] + \frac{1}{2} \sum_{v \in \{0,\ldots,B-1\}} E_v \left[h^2(\psi_{v,v}, \psi_{v,v+1})\right]$$

$$= \frac{1}{2B} \left[\sum_{v \in \{1,\ldots,B\}} |\sqrt{\psi_{v-1,v}} - \sqrt{\psi_{v,v}}|^2 + \sum_{v \in \{0,\ldots,B-1\}} |\sqrt{\psi_{v,v}} - \sqrt{\psi_{v,v+1}}|^2\right]$$

$$\geq \frac{1}{4B^2} \left[\sum_{v \in \{1,\ldots,B\}} |\sqrt{\psi_{v-1,v}} - \sqrt{\psi_{v,v}}|^2 + \sum_{v \in \{0,\ldots,B-1\}} |\sqrt{\psi_{v,v}} - \sqrt{\psi_{v,v+1}}|^2\right]^2$$

(by Cauchy-Schwartz)

$$\geq \frac{1}{4B^2} \left[\sum_{v \in \{1,\ldots,B\}} |\sqrt{\psi_{v,v}} - \sqrt{\psi_{v+1,v}}|^2\right]^2$$

$$\geq \frac{1}{4B^2} \left[\sqrt{\psi_{0,0}} - \sqrt{\psi_{B,B}}\right]^2$$

$$\geq \frac{1}{8B^2} \left[|\sqrt{\psi_{0,0}} - \sqrt{\psi_{B,B}}|^2 + |\sqrt{\psi_{B,B}} - \sqrt{\psi_{B,B}}|^2\right]$$

$$= \Omega(\frac{1}{B^2}).$$

In summary, we get a $\Omega(n/B^2)$ lower bound for the Gap$_{\infty}(x,y)$ problem. Moreover, we can get a $\Omega(n)$ lower bound for Set disjointness problem [3].

Remark 3. The Direct Sums are nice, but usually a problem cannot be split into simpler subproblems. For example, there is no known embedding step in Gap-Hamming problem.
2 Nonnegative Matrix Factorization

2.1 Problem Setup

Main Question:
Given \( A \in \mathbb{R}^{n \times n} \) and integer \( k \geq 1 \), is there an algorithm that can determine if there exist two matrices \( U, V^T \in \mathbb{R}^{n \times k} \) such that

\[
A = UV, \quad U \geq 0, \quad V \geq 0.
\]

Or are there any hardness results?

Remark 4. The main question is equivalent to computing the nonnegative rank of \( A \)

Definition. For a matrix \( A \in \mathbb{R}^{m \times n} \), the nonnegative rank of \( A \) is defined as

\[
\text{rank}_+(A) = \min\{q : \sum_{i=1}^{q} R_i = A, \text{rank}(R_i) = 1, i = 1, \ldots, q\}.
\]

Remark 5. By definition, it is easy to conclude that

\[
\text{rank}(A) \leq \text{rank}_+(A) \leq \min\{m, n\}.
\]

Remark 6. Determining whether \( \text{rank}(A) = \text{rank}_+(A) \) is NP-hard. [10]

2.2 Main Idea

Polynomial System Verifier:
Given a polynomial system \( P(x) \) over the real numbers, with

- \( v \): # of variables, \( x = (x_1, \ldots, x_v) \),
- \( m \): # polynomial constraints \( f_i(x) \geq 0, i = 1, \ldots, m \),
- \( d \): maximum degree of all polynomial constraints,
- \( H \): the bitsizes of the coefficients of the polynomials,

then in \( (md)^{O(v)} \text{poly}(H) \) time, we can decide if there exists a solution to polynomial system \( P \).

Therefore we can

1. Write \( \min_{U, V^T \in \mathbb{R}^{n \times k}, U \geq 0, V \geq 0} \|UV - A\|_F^2 \) as a polynomial system that has \( \text{poly}(k) \) variables and \( \text{poly}(n) \) constraints and degree;

2. Use polynomial system verifier to solve it
2.3 Algorithms and Bounds

We can formulate the problem as follows

Given: \( A \in \mathbb{R}^{n \times n}, k \in \mathbb{N}^* \)

Question: Are there matrices \( U, V^T \in \mathbb{R}^{n \times k} \) such that

\[ A = UV, \quad U \geq 0, \quad V \geq 0 \]  

Output: Yes or No

(NMF)

There is a way of reducing (NMF) to the following k-Sum problem [1], which is defined as

Given: a set of \( n \) values \( \{s_1, s_2, \ldots, s_n\} \) each in the range \([0, 1]\)

Question: If there is a set of \( k \) numbers that sum to exactly \( k/2 \)

Output: Yes or No

(k-SUM)

2.4 Upper Bounds

1. In [1], we can solve (NMF) in \( n^{2O(k)} \) time.
2. In [8], we can solve (NMF) in \( 2^{O(k^3)} n^{O(k^2)} \) time.

2.5 Lower Bounds

Exponential Time Hypothesis: states that 3-SAT (or any of several related NP-complete problems) cannot be solved in subexponential time in the worst case.

By [9], we can conclude the following claim:

Claim. Assume that 3-SAT on \( n \) variables cannot be solved in \( 2^{O(n)} \) time, then (k-SUM) cannot be solved in \( n^{O(k)} \) time.

So under the Exponential Time Hypothesis [5], (NMF) requires at least \( n^{\Omega(k)} \).

2.6 Open Problem

For (NMF) the upper bound is \( n^{O(k^2)} \) while the lower bound is \( n^{\Omega(k)} \). Can we find a tight bound?

References


