Lower Bounds for Streaming Algorithms

Barna Saha
Communication Complexity

• In this class, we have seen many single pass streaming algorithms that require sublinear amount of memory and return approximate answers.

• Are there space requirements optimal?

• Do they have the best approximation possible?

• Communication complexity is a tool to prove such lower bounds.
One-way Communication Complexity

- Alice has \( x \) and Bob has \( y \)—together they want to compute \( f(x,y) \)
- Only one way communication from Alice to Bob is allowed

One-way communication complexity of a Boolean function \( f \) is the minimum worst-case number of bits used by any 1-way protocol that correctly decides the function or decides with probability > 1/2
Connection to Streaming Algorithms

- Small space streaming algorithm implies low communication complexity (CC)
- Consider a problem that can be solved using a streaming algorithm $S$ that uses space $s$
- Treat $(x,y)$ as stream
- Alice feeds $x$ to $S$ → summary of size $s$ → sends to Bob
- Bob feeds the summary to $S$ and then $y$
- One way communication: $s$ bits
Streaming Lower Bound for CC

To prove lower bound on space usage of a streaming algorithm, we need to come up with a Boolean function that

(i) can be reduced to a streaming problem that we want to study, and

(ii) does not admit a low one-way communication complexity.
The Disjointness Problem

• Alice and Bob both hold n bit vectors $x$ and $y$ respectively
• $\text{DISJ}(x,y)=1$ if there is no index $i$ such that $x_i=y_i=1$

• **Theorem:** Every deterministic one-way communication protocol that computes the DISJ function uses at least $n$ bits in CC in the worst case.
• Similar result holds for randomized protocol as well.
Lower Bound for $F_\infty$

**Theorem 3.** Every randomized streaming algorithm that, for every data stream of length $m$, computes $F_\infty$ to within $(1 \pm .2)$ factor with probability at least $2/3$ uses space $\Omega(\min\{m, n\})$.

- Proof. In the board