Exercise 10.1

Saturday, 14 October 2023 12:18

Let b be the number of bits of storage available for the algorithm

Let S be the set of all subsets of $m-1$ distinct items from $[n]$

Let X be a set of streams of size $m-1$, one for each subset in S.

i.e. the number of memory states available is less than the number of streams

Now, assume that $2^b < |X|$. Then, there must be two streams in X, say σ_1 and σ_1', such that the deterministic streaming algorithm is in the same memory state after processing one of them. Note that there must be an element $j \in [n]$ such that $j \in \sigma_1$ and $j \notin \sigma_1'$.

Now, consider what happens when the algorithm processes $\sigma_1 \circ \sigma_2$ and $\sigma_1' \circ \sigma_2$, for $\sigma_2 = \langle j \rangle$. Then, the correct algorithm will return

- the same answer for both inputs, given that the state prior to processing σ_2 is equal.
- the answer $m-1$ for the input $\sigma_1 \circ \sigma_2$, as σ_1 consists of $m-1$ distinct items, and j is not distinct from these items
- the answer m for the input $\sigma_1' \circ \sigma_2$, as σ_1 consists of $m-1$ distinct items, and j is distinct from these items.

But this gives a contradiction; hence, the assumption that $2^b < |X|$ must be wrong; thus, we need to have that the number of bits b complies with $2^b \geq |X|$. The minimum value of b for which this holds is given by

$$2^b = |X| = |S| = \binom{n}{m-1} \geq \left(\frac{n}{m-1}\right)^{m-1} = 2^{\log_2\left(\left(\frac{n}{m-1}\right)^{m-1}\right)}$$

$$= 2^{(m-1)\log_2\left(\frac{n}{m-1}\right)}$$

$$b = (m-1)\log_2\left(\frac{n}{m-1}\right) = \Omega\left(m \log_2\left(\frac{n}{m}\right)\right)$$