

General Instructions: Please write concisely, but rigorously, and show your calculations explicitly. Each problem is worth 5 points, and only “nearly flawless” solutions will earn full credit.

Honor Principle (Different from earlier homeworks, please read carefully): For this homework, the only sources you are allowed to refer to are the class notes and the website for this course, including material (notes, papers) linked to from the course website. You may not refer to *any other* books, research papers, notes, or online material.

As before, you are allowed to discuss the problems and exchange solution ideas with your classmates. But when you write up any solutions for submission, you must work alone. *If in doubt, ask the professor for clarification!*

For this homework, the only streaming model you need to think of is the vanilla (i.e., plain insertions) model. All of the lower bounds you are asked to prove apply even in this simple model. Obviously, as a result, they also apply to the more complicated turnstile models.

Improving a Matching using Additional Passes

14. In class, we gave the following algorithm for the maximum weight matching (MWM) problem. The stream is a *weighted graph stream*, so the tokens are triples (u, v, w) , where $u, v \in [n]$ are vertices and $w \geq 0$ is the weight of the edge $\{u, v\}$.

```

Initialize      :  $M \leftarrow \emptyset$  ;
Process  $(u, v, w)$ :
1  $C \leftarrow \{e \in M : e \cap \{u, v\} \neq \emptyset\}$  ;
2 if  $w > (1 + \alpha) \text{wt}(C)$  then
3    $M \leftarrow M \cup \{u, v\} \setminus C$  ;
Output          :  $M$  ;
    
```

We shall refer to this as an α -improving algorithm. We showed, in class, that by setting $\alpha = 1/\sqrt{2}$, we obtain a one-pass $(3 + 2\sqrt{2})$ -approximation to the MWM.

Now suppose we are given $\varepsilon > 0$, and we run additional passes of an α -improving algorithm with the setting $\alpha = \varepsilon/3$ (the first pass still uses $\alpha = 1/\sqrt{2}$), using the output of the i th pass as the initial M for the $(i + 1)$ th pass. Let M_i denote the output of the i th pass. We stop after the $(i + 1)$ th pass when

$$\frac{\text{wt}(M_{i+1})}{\text{wt}(M_i)} \leq 1 + \frac{\alpha^3}{1 + 2\alpha + \alpha^2 - \alpha^3}, \quad \text{where } \alpha = \frac{\varepsilon}{3}.$$

Prove that this multi-pass algorithm makes $O(\varepsilon^{-3})$ passes and outputs a $(2 + \varepsilon)$ -approximation to the MWM.

Hint: Use the two lemmas that we used in class when we analyzed the first pass of this algorithm. Observe that in subsequent passes, edges in $M_{i+1} \cap M_i$ do not kill any other edges, and prove that this implies

$$\text{wt}(M^*) \leq (1/\alpha + 3 + 2\alpha) \text{wt}(M_{i+1} \setminus M_i) + 2(1 + \alpha) \text{wt}(M_{i+1} \cap M_i),$$

where M^* is an MWM. Separately, show that $(1 + \alpha) \text{wt}(M_i \setminus M_{i+1}) \leq \text{wt}(M_{i+1} \setminus M_i)$ and use this, plus the stopping criterion, to obtain a lower bound on $\text{wt}(M_{i+1} \cap M_i)$. Put all this together to show that $\text{wt}(M^*) \leq (2 + 3\alpha) \text{wt}(M_p)$, where p is the number of passes made.

Space lower bounds on data stream algorithms

For each streaming problem, first think of the appropriate communication problem to reduce from. You have only a small catalogue of problems to try, so it should not be hard to find the right one.

Whenever randomization is allowed in a streaming algorithm, we tacitly assume that an error probability of $1/3$ is allowed.

15. Consider a graph stream describing an unweighted, undirected n -vertex graph G . Prove that $\Omega(n^2)$ space is required to determine, in one pass, whether or not G contains a triangle, even with randomization allowed.
16. Prove that computing F_2 exactly, in one pass with randomization allowed, requires $\Omega(\min\{m, n\})$ space. [Construct an appropriate “hard stream” of length m , with universe size n , where $m = \Theta(n)$ and show that $\Omega(n)$ space is required on this stream.]
Then extend the result to multiple passes, with randomization allowed. The lower bound for p passes should be $\Omega(\min\{m, n\}/p)$.
17. In the Δ -approximate median problem, we are asked to find an element x in a stream S of length m with the property that $|\text{rank}(x, S) - \lfloor m/2 \rfloor| \leq \Delta$, where we define

$$\text{rank}(x, S) = |\{y \in S : y \leq x\}|.$$

Prove that a one-pass randomized streaming algorithm for this problem requires $\Omega(\min\{m, n\}/(1 + \Delta))$ space. [For partial credit, prove this for the special case when $\Delta = 0$.]