

Intelligent Learning and Analysis Systems SS19, Exercise Sheet 1

Andreas Hene, Niklas Mertens, Richard Palme

April 13, 2019

1. Finding Two Missing Items

2. Identifying the Majority

Let $\sigma = \langle a_1, \dots, a_m \rangle$ and $a_i \in [n]$.

Algorithm 1 Algorithm for finding the majority of σ , if it exists

```
1:  $c \leftarrow 0$ 
2: for all  $i \in [m]$  do
3:   if  $c = 0$  then
4:      $s \leftarrow a_i$ 
5:      $c \leftarrow 1$ 
6:   end if
7:   if  $s = a_i$  then
8:      $c += 1$ 
9:   else
10:     $c -= 1$ 
11:  end if
12: end for
13: return  $s$ 
```

Let $A(\sigma)$ be the output of the algorithm.

Claim: $A(\sigma)$ is the majority of σ , if it exists.

Proof: Let k be the number of times c is set to 0 in the for loop. We prove the claim by induction over k .

Base case:

Let $k = 0$. Since c is never set to 0 during the loop, we know that s never changes after iteration 1, i.e. $A(\sigma) = a_1$. Also, c increases at least one time more often than it decreases, so $A(\sigma) = a_1$ has to be the majority of σ .

Step case:

Let $k > 0$. Suppose the claim is true for all streams for which c is set to 0 at most $k - 1$ times in the for loop. Let d be the majority of σ . Let j be the first iteration where c is set to 0. Side note: When c is set to 0 in an iteration j' , j' has to be even. d can occur at most $\frac{j}{2}$ times in $\tau := \langle a_1, \dots, a_j \rangle$, otherwise d would be the majority of τ and c couldn't be set to 0 in iteration j . Consequently d will occur at least

$$\left\lfloor \frac{m}{2} \right\rfloor + 1 - \frac{j}{2} = \left\lfloor \frac{m-j}{2} \right\rfloor + 1$$

times in $\sigma' = \langle a_{j+1}, \dots, a_m \rangle$. Because σ' is a stream of length $m - j$, d is the majority of σ' . Since the number of zeros in σ' is smaller than k , we can apply the inductive hypothesis to see that $A(\sigma') = d$. Since in iteration j of computing $A(\sigma)$, c is set to 0, applying the algorithm to σ is the same as applying it to τ and then to σ' . Hence, $A(\sigma) = d$. ■

3. Reservoir Sampling I.
4. Reservoir Sampling II.
5. The ϕ -HH Problem: Lower Bounds