## Warm-up

**Problem 1.** Discussion: what are the parallels between Bloom filters and COUNT-MINSKETCH?

**Problem 2.** Prove the following fact about "monotonicity of $\ell_p$ norms": if $x \in \mathbb{R}^d$, then $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$. Show, in addition, that $\|x\|_2 \geq \|x\|_1/\sqrt{d}$. When are these inequalities tight?
(($\star$) More generally: if $1 \leq p \leq q \leq \infty$, then $\|x\|_q \leq \|x\|_p$.)

**Problem 3.** Discuss the advantages and disadvantages of MISRA-GRIES versus COUNTMINSKETCH when used in the cash register model: speed, memory, approximation. Can you think of a situation where having an overestimate (COUNTMINSKETCH) is better than an underestimate (MISRA-GRIES)?

**Problem 4.** Check your understanding: why are we using a hash function $g$ in the BKJST algorithm? What would happen if we were to store $j$ in the bucket $B$, instead of $g(j)$?

## Problem solving

**Problem 5.** For the same space budget $s$ (ignoring the constants in the $O(\cdot)$'s), are the theoretical guarantees provided by COUNTMINSKETCH better, worse, or incomparable to those of COUNTSKETCH?

**Problem 6.** Generalise the analysis of the COUNTMINSKETCH algorithm to show it works in the *strict* turnstile model, where updates of the stream are of the form $(j, c) \in [n] \times \{-B, \dots, B\}$ (can be negative) but one must have $f_j \geq 0$ at every time. Check the guarantees you can provide on the output $\widehat{f}$. Does the analysis extend to the general turnstile model, where $f_j$ can become negative?

**Problem 7.** Show that the MISRA-GRIES algorithm is a sketching algorithm: namely, suppose we run MISRA-GRIES (with the same parameter $k = \lceil 1/\varepsilon \rceil$) on two streams $\sigma_1, \sigma_2$, getting output vectors $\widehat{f}^{(1)}, \widehat{f}^{(2)}$. Combine then as follows:

1. Set $\widehat{f} \leftarrow \widehat{f}^{(1)} + \widehat{f}^{(2)}$

2. If $\widehat{f}$ has more than $k$ non-zero entries, let $v > 0$ be the value of the $(k+1)$-th, in non-increasing order.

3. Set $\widehat{f}_j \leftarrow \max(\widehat{f}_j - v, 0)$ for all $j$

a) Argue that $\widehat{f}$ has at most $k$ non-zero entries.

b) Show that the sketch $\widehat{f}$ provides the original MISRA-GRIES estimation guarantees, for the combined stream $\sigma_1 \circ \sigma_2$.

c) Is this sketch a linear sketch?

---

## Advanced

---

**Problem 8.** Modify the COUNTMINSKETCH algorithm so that it outputs a *list* of the $\ell_1$ Heavy Hitters in the strict turnstile model: that is (similarly to an exercise in Tutorial 8), given parameter $\varepsilon \in (0, 1]$, it should output a set $H \subseteq [n]$ such that $H_\varepsilon(\sigma) \subseteq H \subseteq H_{\varepsilon/2}(\sigma)$, where

$$H_\varepsilon(\sigma) = \{j \in [n] : f_j \geq \varepsilon \cdot \|f\|_1\}$$