Sublinear Time and Space Algorithms 2026A – Problem Set 2

Robert Krauthgamer

Due: December 3, 2025

General instructions: Please keep your answers short and easy to read. You can use results, calculations or notation seen in class without repeating them, unless asked explicitly to redo them.

1. Consider the frequency-vector model, where the input stream contains additive updates to a vector $x \in \mathbb{R}^n$ whose coordinates are integers bounded by $\mathrm{poly}(n)$.

   Explain how to $(1+\epsilon)$-approximate $\sum_{i<j}(x_i + x_j)^2$ by a streaming algorithm with storage requirement $(\epsilon^{-1} \log n)^{O(1)}$ bits.

2. Prove the norm-comparison inequality: $\|x\|_2 \le n^{1/2-1/p}\|x\|_p$ for all $p \ge 2$. Then design a streaming algorithm for ℓ_p point query with space $\tilde{O}(\alpha^{-2} n^{1-2/p})$.

   Hint: Use Holder's inequality which generalizes the Cauchy-Schwartz inequality and asserts that for all $p, q \in [1, \infty]$ satisfying $1/p + 1/q = 1$,

   $$\forall a, b \in \mathbb{R}^n, \qquad \langle a, b \rangle \le \|a\|_p \|b\|_q.$$

   Use $a_i = x_i^2$ and $b_i = 1$.

3. Fix parameters $\phi \in (0, 1)$ and $p \in [1, \infty)$. The set of *heavy hitters* of a vector $x \in \mathbb{R}^n$ is defined as

   $$HH^p_\phi(x) = \{i \in [n] : \ |x_i| \ge \phi \|x\|_p\}.$$

   Observe that $\left|HH^p_\phi(x)\right| \le 1/\phi^p$.

   Consider the frequency-vector model, where the input stream contains additive updates to a vector $x \in \mathbb{R}^n$ whose coordinates are integers bounded by $\mathrm{poly}(n)$. Design a streaming algorithm (for p and ϕ known in advance) that outputs a set S that approximates the heavy hitters set, in the sense that

   $$HH^p_\phi \subseteq S \subseteq HH^p_{\phi/2}.$$

4. Consider an insertion-only frequency-vector model, i.e., the input is a stream of items from $[n]$ and $x \in \mathbb{R}^n$ is its frequency vector. Let $1 \le T \le n$ be a threshold given in advance (before the stream).

   Design a streaming algorithm that $O(1)$-approximates $f(x) = \sum_{i=1}^n \min\{x_i, T\}$ (in words, $f(x)$ sums the coordinates after truncating them at T).

   Hint: First solve the easy case $T = 1$, then consider subsampling the insertions at rate $1/T$.