Sublinear Time and Space Algorithms 2020B – Problem Set 2

Robert Krauthgamer

Due: June 1, 2020

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. We saw in class how to construct a pairwise independent family H of hash functions $h : [n] \to [M]$, when $M \geq n$ is a prime. Specifically, each hash function was of the form $h_{p,q}(i) = pi + q \pmod{M}$, and can be stored using $\log_2 |H| = O(\log M)$ bits.

   Extend this approach to construct a k-wise independent family H_k for any $3 \leq k \leq n$. How many bits are needed to store a hash function?

   Hint: Use higher-degree polynomials, and rely on the determinant of a Vandermonde matrix.

2. A matrix $A \in \mathbb{R}^{m \times n}$ is called ε-coherent if its columns $A^1, \ldots, A^n \in \mathbb{R}^m$ satisfy (1) all $i \in [n]$, $\|A^i\|_2 = 1$; and (2) for all $i \neq j \in [n]$, $|\langle A^i, A^j \rangle| \leq \varepsilon$.

   (a) Show that for every n and $\varepsilon \in (0, 1/2)$ there is an ε-coherent matrix $A \in \mathbb{R}^{m \times n}$ with $m = O(\varepsilon^{-2} \log n)$.
   Hint: Show that a random matrix whose entries are $\pm 1/\sqrt{m}$ independently at random satisfies the above with positive probability.

   (b) Prove that Algorithm IncoherentSketch below solves the ℓ_1-point query problem, i.e., given a frequency vector $x \in \mathbb{R}^n$ it outputs $(A^T y)_i \in x_i \pm \varepsilon \|x\|_1$.
   Algorithm IncoherentSketch
   1. Init: Fix a matrix A that is ε-coherent
   2. Update: Maintain a linear sketch $y = Ax$
   3. Output: to estimate x_i report $(A^T y)_i$.
   Notice that its storage requirement is $O(m)$ not including the matrix A.

Extra credit:

3. (This question is an attempt to implement algorithm FM seen in class using pairwise independence, however the bound obtained below is too weak.)

   Let $X_1, \ldots, X_n$ be pairwise independent random variables, each distributed uniformly in $[0, 1]$. Fix $S \subset [n]$ and let $Y_S = \min\{X_i : i \in S\}$ be the minimum of the respective variables. Show that its expectation satisfies

   $$\Omega(\tfrac{1}{|S|}) \leq \mathbb{E}[Y_S] \leq O(\tfrac{1}{\sqrt{|S|}}),$$

   and provide an upper bound on its variance.