Exercise VIII, Sublinear Algorithms for Big Data Analysis 2024-2025

These exercises are for your own benefit. Feel free to collaborate and share your answers with other students, and solve as many problems as you can. Problems marked (*) are more difficult, but also more rewarding. These problems have been taken from various sources on the Internet, too numerous to cite individually.

1 (Lower bounds for ℓ_0 samplers) Recall from class that an ℓ_0 sampler is a randomized linear sketch $A \in \mathbb{R}^{m \times n}$ together with a decoding algorithm $D : \mathbb{R}^m \to [n] \cup \{\bot\}$ such that for every fixed $x \in \mathbb{R}^n$ one has

$$\textbf{Prob}[\text{decoding } D(A(x)) \text{ fails or } D(A(x)) = \bot] \leq \delta$$

for some $\delta \in (0, 1)$ and for all $j \in \text{supp}(x)$

$$\textbf{Prob}[D(A(x)) = j| \text{ decoding } D(A(x)) \text{ succeeds and } D(A(x)) \neq \bot] = \frac{1}{|\text{supp}(x)|}.$$

In both cases above the probability is over the random string R used to generate A (we write A_R to denote the matrix A generated using random string R). *Note that it is crucial that A is independent of x.*

Also note that the ℓ_0 sampler sketch construction that we saw in class uses a matrix A with integer entries that are polynomially bounded in n. Thus, if $x \in \mathbb{R}^n$ has integer entries bounded in absolute value by $n^{O(1)}$, the product Ax can be stored using $O(m \log n)$ bits. Show that any ℓ_0 sampler that uses a matrix A with polynomially bounded entries and succeeds with probability $1 - \delta$ as above must use $m = \Omega(\log(1/\delta))$ rows for every $\delta > 1/n^{10}$.

1a Show that there exists a family $\mathcal{F}$ of subsets $S \subseteq [n]$ of size $\Omega(\log(1/\delta))$ and a random string R such that **(a)** $\log_2 |\mathcal{F}| = \Omega(\log n \cdot \log(1/\delta))$ and **(b)** every $S \in \mathcal{F}$ can be uniquely recovered from $A_R \mathbf{1}_S$, where $\mathbf{1}_S \in \mathbb{R}^n$ stands for the indicator vector of S:

$$(\mathbf{1}_S)_j = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{o.w.} \end{cases}$$

Hint: use linearity of the sketch to perform iterative recovery of S, and be very careful when computing success probability

1b Conclude that an ℓ_0 sampler with error probability bounded by $\delta > 1/n^{10}$ for every fixed input must use $m = \Omega(\log(1/\delta))$ rows. You may use the result of **1a** even if you did not prove it.