Problem Set 4

Sublinear Algorithms

Due Tuesday, September 29

1. Another plausible algorithm for sparse recovery (in the insertion-only model) is *sampling*. If we sample m of the $N = \|x\|_1$ coordinates, with replacement, and let $y \in \mathbb{R}^n$ be the histogram of the sample, we can estimate x by $\widehat{x}_i := y_i \frac{N}{m}$.

   (a) Show that $\mathbb{E}[\widehat{x}_i] = x_i$ for each i.

   (b) Show that $\text{Var}(\widehat{x}_i) \leq N x_i / m$.

   (c) Give an upper bound for $|\widehat{x}_i - x_i|$ as a function of N, m, n, and failure probability δ.

   (d) Give a bound on m such that, with high probability,

   $$\|\widehat{x} - x\|_\infty \leq N/k.$$

   How does this compare to the algorithms we covered in class?

2. Comparison of the CountMinSketch guarantee

   $$\|\widehat{x} - x\|_\infty \leq \frac{1}{k}\|x - H_k(x)\|_1$$

   to the CountSketch guarantee

   $$\|\widehat{x} - x\|_\infty \leq \frac{1}{\sqrt{k}}\|x - H_k(x)\|_2.$$

   (a) For any vector $x \in \mathbb{R}^n$, show that

   $$\|x - H_k(x)\|_2 \leq \frac{1}{\sqrt{k}}\|x\|_1.$$

   (b) Show that if $\widehat{x}$ is the result of CountSketch for $k' = 2k$, then

   $$\|\widehat{x} - x\|_\infty \leq \frac{1}{k}\|x - H_k(x)\|_1.$$

   Compare this to the bound given by CountMinSketch.