Problem Set 6

Randomized Algorithms

Due Wednesday, December 2

1. You are given a set of vectors $V = v_1, \ldots, v_m \in \mathbb{R}^n$, with all coordinates nonnegative. The goal of this problem is to "sparsify" these vectors: construct $w_1, \ldots, w_k \in \mathbb{R}$ and $u_1, \ldots, u_k \in V$ such that

$$\overline{u} := \sum_{i=1}^{k} w_i u_i$$

and

$$\overline{v} := \sum_{i=1}^{m} v_i$$

satisfy

$$\overline{u}_j = (1 \pm \epsilon)\overline{v}_j$$

for all j.

We would like to solve this using the random sampling techniques developed in class: choose a probability distribution $p_1, \ldots, p_m \in \mathbb{R}$, and draw k independent samples $(w, u) = (\frac{1}{kp_j}, v_j)$ for $j \sim p$.

(a) Show, for some setting of p, that $k = O(n \log n)$ suffices for constant ϵ and all V. What is the dependence on ϵ?

(b) For many inputs V, your result should work with many fewer samples (e.g., $k \ll n$). When does this happen? Give a data-dependent bound for the necessary k as a function of V.

(c) (Optional) Show a different, deterministic, algorithm that uses $k = n$ for all V.