Problem set I

Collaboration on the homework problems is allowed and encouraged, but every student must write up their own solutions, and list collaborators on the first page of the assignment.

1 (25 pts) In exercise session I we showed that any deterministic algorithm that achieves a factor 1.1 approximation to the number of distinct elements in a data stream of n items must use $\Omega(n)$ space. In this problem you will prove a stronger space lower bound for *exact deterministic* algorithms and a comparable space lower bound for *exact randomized* algorithms.

1a (10 pts) Prove that a deterministic algorithm that computes the number of distinct elements in any stream of $n+1$ elements exactly must use at least n bits of space.

Solution. Since the algorithm is deterministic, we can think of it as a deterministic automata with states being different memory configurations. Suppose the algorithm uses m bits. We will show how we can represent any string $x \in \{0,1\}^n$ using just m bits. This will prove that $m \geq n$ information theoretically.

For any given x, create the following stream $S = \{i | x_i = 1\}$ of size n. The memory of the algorithm is now used as a representation. To recover the actual string again, finish running the algorithm with all possible values for the $n+1$st input $\in [n]$. If $i \in S$, then the answer of the algorithm would be n, else it would be $n+1$. Thus we can infer if $i \in S \forall i$ and hence we can reconstruct our original string x. □

1b (15 pts) Prove that a randomized algorithm that for every stream of n numbers outputs the exact number of distinct elements with probability at least $99/100$ must use $\Omega(n)$ space. *Hint: combine ideas from* **1a** *and the lower bound shown in exercise session II.*

Solution. Consider the family of sets F used in class such that $|F| \geq 2^{cn}$ and for all $S_i \in F$, $|S_i| = n/4$ and $|S_i \cap S_j| \leq n/8 \forall i \neq j$. For each set $S \in \mathcal{F}$ let $X_S = 0$ if ALG returns the correct answer on S and 1 otherwise. For every S and every $i \in [n]$ let $X_{S,i}$ equal 0 if ALG is correct on $S \cup \{i\}$ and 1 otherwise. For each S one has

$$\mathbf{E}[X_S] \geq 1/100$$

and

$$\mathbf{E}[\sum_{i \in [n]} X_{S,i}] \leq n/100.$$

We thus have by Markov's inequality for every $S \in \mathcal{F}$ that

$$\mathbf{E}[X_S + \mathbf{I}[\sum_{i \in [n]} X_{S,i}] > n/50]] \leq 1/100 + 1/2.$$

Summing over all $S \in \mathcal{F}$ we get

$$\sum_{S \in \mathcal{F}} \mathbf{E}[X_S + \mathbf{I}[\sum_{i \in [n]} X_{S,i}] > n/50]] \leq (1/100 + 1/2)|\mathcal{F}|.$$

Thus, by Markov's inequality there exists a setting of the random bits R of ALG and a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at least $0.01|\mathcal{F}|$ such that ALG(R), on every $S \in \mathcal{F}'$ decodes S with at most $1/50$ fraction of mistakes. Since elements of $\mathcal{F}'$ are at least $n/4$ apart in Hamming metric, unique decoding is possible. □

2 (30 pts) In this problem you will design and analyze a sketching matrix Z based on the Hadamard transform that can be applied to a vector $x \in \mathbb{R}^n$ very fast, namely in $O(1/\epsilon^2 + n \log n)$ time, and approximately preserves the Euclidean norm of x with high probability.

Let D be an $n \times n$ diagonal matrix such that $D_{ij} = 0$ for $i \neq j$ and $D_{ii} \in \{-1, +1\}$ are independent uniformly random signs. Let $H \in \mathbb{R}^{n \times n}$ be the Hadamard transform matrix, i.e. a symmetric matrix with ± 1 entries that satisfies $H^T H = n \cdot I_n$. Finally, let P be an $m \times n$ matrix with each row containing a single 1 in a uniformly random position chosen independently of other rows, and all other entries equal to zero. Finally, the sketching matrix Z is defined as $Z := \frac{1}{\sqrt{m}} \cdot P \cdot H \cdot D$. Note that P and D are random, and H is a deterministic.

2a (2 pts) Show that Zx can be computed in time $O(m + n \log n)$ for every x (you can assume that you can generate a uniformly random number between 1 and n at unit cost).

Solution. For any vector x, the Hadamard transform can be coputed very quickly in time $O(n \log n)$. To compute $PHDx$, we have to randomly change the sign of each x_i, then compute its Hadamard transform, and then sample a $O(\frac{1}{\epsilon^2})$ number of co-ordinates. This process takes $O(n \log n + \frac{1}{\epsilon^2})$ time. □

2b (6 pts) For a fixed diagonal sign matrix D let $Y \in \mathbb{R}$ denote the value of a uniformly random coordinate of HDx. Show that $\mathbf{E}[Y^2] = ||x||_2^2$.

Solution.

$$\mathbf{E}[Y^2] = \frac{1}{n}\sum_{i=1}^{n}[HDx]_i^2 = \frac{1}{n}||HDx||_2^2 = \frac{1}{n}(HDx)^\top HDx = \frac{1}{n}x^\top DH^2Dx = ||x||_2^2$$

□

2c (17 pts) Show that for fixed $x \in \mathbb{R}^n$ the random variable Y defined in **2b** satisfies

$$\mathbf{Var}(Y^2) \leq C_1||x||_2^4.$$

for an absolute constant $C_1 > 0$ with probability at least 9/10 over the choice of the diagonal sign matrix D.

Solution. Lets use d_i to represent the ith diagonal value in D and $h_{i,j}$ for the i,j value of H. $\mathbf{Var}(Y^2) = \mathbf{E}[Y^4] - \mathbf{E}[Y^2]^2 = \mathbf{E}[Y^4] - ||x||_2^4$.

$$\begin{aligned}
\mathbf{E}[Y^4] &= \mathbf{E}\frac{1}{n}\sum_{i=1}^{n}[HDx]_i^4 \\
&= \mathbf{E}\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{i=1}^{n}h_{i,j}d_jx_j\right)^4 \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}h_{i,j}^4d_j^4x_j^4 + 3\sum_{i=1}^{n}\sum_{j\neq k\in[n]}^{n}h_{i,j}^2h_{i,j}^2d_j^2d_k^2x_j^2x_k^2\right) \\
&= \sum_{j=1}^{n}x_j^4 + 3\sum_{j\neq k\in[n]}^{n}x_j^2x_k^2 \\
&\leq 3\left(\sum_{j=1}^{n}x_j^2\right)^2 = 3||x||_2^4
\end{aligned}$$

□

2d (5 pts) Prove that if $m \geq C_2/\epsilon^2$ for an absolute constant $C_2 > 0$, then

$$\mathbf{Prob}_{P,D}\left[\left|||Zx||_2^2 - ||x||_2^2\right| > \epsilon||x||_2^2\right] < 1/5.$$

Solution. For the sake of simplicity, we will assume that $||x||_2 = 1$. Everything can be appropriately rescaled as needed. Also let $y = HDx$. Zx is simply a sampling of the vector y. An unbiased estimator of $\frac{1}{n}||y||_2^2$ is as follows - take a random co-ordinate of y, square it. Thus $||Zx||_2^2$ is the mean of m samples of Y^2 whose expectaion is 1 and variance is bounded by c_1. Thus we can use Chebyschev's inequality to say that the estimate is less than $1+\epsilon$ with probability $9/10$ by taking $O(\frac{c_1}{\epsilon^2})$ samples. However we only bounded the variance with probabilty $9/10$. This can be fixed as below. Let an indicator variable G represent the good event that $Var[Y^2] \leq c_1$ occuring.

$$\begin{aligned}
\Pr_{P,D}[||Zx||_2^2 > 1+\epsilon] &= \Pr_{P,D}[||Zx||_2^2 > 1+\epsilon | G=1]\mathbf{E}[G] + \Pr_{P,D}[||Zx||_2^2 > 1+\epsilon | G=0]\Pr[G=0] \\
&\leq \frac{9}{10}\Pr_{P,D}[||Zx||_2^2 > 1+\epsilon | G] + \frac{1}{10} \\
&= \frac{9}{10}\frac{1}{10} + \frac{1}{10} < 1/5
\end{aligned}$$

□

3 (20 pts) In this problem you will design and analyze a very fast sketching method for matrices, namely one that can be applied without forming the sketched matrix explicitly.

Let $n, B \geq 1, B \leq n$, be integers. Let $h_1, h_2 : [n] \to [B]$ be pairwise independent hash functions, let $s_1, s_2 : [n] \to \{-1, +1\}$ be pairwise independent sign functions. Here we let $[n] = \{0, 1, \ldots, n-1\}$ and $[B] = \{0, 1, \ldots, B-1\}$. For $(i, j) \in [n] \times [n]$ let $h(i, j) = h_1(i) + h_2(j) \pmod B$ and $s(i, j) = s_1(i)s_2(j)$.

For a matrix $A \in \mathbb{R}^{n \times n}$ define $\textsc{Sketch}(A)$ as the vector $y \in \mathbb{R}^B$ such that for every $b \in [B]$

$$y_b = \sum_{\substack{(i',j') \in [n]\times[n] \\ h(i',j')=b}} s(i', j') \cdot A_{i',j'}.$$

3a (10 pts) For every $(i, j) \in [n] \times [n]$ let

$$\textsc{Estimate}(i, j) = y_{h(i,j)} \cdot s(i, j).$$

Prove that

$$\Pr[|\textsc{Estimate}(i, j) - A_{i,j}| > C\|A\|_F/\sqrt{B}] < 1/10,$$

where $\|A\|_F^2 = \sum_{(i',j') \in [n]\times[n]} A_{i',j'}^2$ and $C > 0$ is an absolute constant.

Solution. We have

$$\begin{aligned}
\mathbf{E}[(A_{i,j} - s(i,j)y_{h(i,j)})^2] &= \mathbf{E}\left[\left(\sum_{(i',j')} \mathbf{I}[h(i,j) = h(i',j')] \cdot s(i,j)s(i',j')A_{i',j'}\right)^2\right] \\
&= \sum_{(i',j')} \sum_{(i'',j'')} \Pr[h(i,j) = h(i',j') = h(i'',j'')] \cdot \mathbf{E}\left[s(i',j')s(i'',j'')\right] A_{i',j'}A_{i'',j''} \\
&= \frac{1}{B} \sum_{(i',j') \neq (i,j)} A_{i',j'}^2 \\
&\leq \frac{1}{B}\|A_F\|_2^2.
\end{aligned}$$

In the derivation above we used the fact $\mathbf{E}\left[s(i',j')s(i'',j'')\right] = 0$ unless $(i',j') = (i'',j'')$ and $\mathbf{E}\left[s(i',j')^2\right] = 1$ for all i',j', as well as the fact that $\Pr[h(i,j) = h(i',j')] = 1/B$ for every $(i',j') \neq (i,j)$. The required bound now follows by Markov's inequality. □

3b *(0 pts; do not hand in)* Convince yourself that this sketch can be used to obtain an ℓ_2 heavy-hitters primitive using the standard approach involving independent repetitions and a median estimator.

3c *(*10 pts) Let $x_1, \ldots, x_r \in \mathbb{R}^n$ and let

$$A = \sum_{k=1}^{r} x_k x_k^T.$$

Show that, given $x_1, \ldots, x_r$, the vector y can be computed in time $O(r(n + B\log B))$ assuming that the hash functions h_1, h_2 and sign functions can be evaluated in constant time. Note that when $r = o(n/\log n)$, this is faster than forming the matrix A explicitly.

Solution. **Proof**. Let $A = xx^T$ for some $x \in \mathbb{R}^n$, and let $y = \textsc{Sketch}(A)$, so that for $b \in [B]$

$$y_b = \sum_{\substack{(i',j') \in [n] \times [n] \\ h(i',j')=b}} s(i',j') \cdot A_{i',j'} = \sum_{\substack{(i',j') \in [n] \times [n] \\ h_1(i')+h_2(j')=b}} s_1(i')s_2(j') \cdot x_{i'}x_{j'}$$

is the convolution of $y^1 \in \mathbb{R}^B$ defined by

$$y^1_b = \sum_{\substack{i' \in [n] \\ h_1(i')=b}} s_1(i') \cdot x_{i'}$$

and $y^2 \in \mathbb{R}^B$ defined by

$$y^2_b = \sum_{\substack{j' \in [n] \\ h_2(j')=b}} s_2(j') \cdot x_{j'}.$$

This convolution can be computed in time $n + B \log B$ using the Fast Fourier Transform. Then the sketches $\textsc{Sketch}(x_k x_k^T), k = 1, \ldots, r$, can be added up in time $O(rB) = O(rn)$, giving the result. □

□

4 *(25 pts)* In this problem we will show that a matrix with independent Gaussian entries and an appropriately large number of rows is a subspace embedding.

Given any $\epsilon, \delta, k > 0$ suppose $S \in \mathbb{R}^{m \times n}$ is a random matrix where each entry is independently sampled from a Gaussian distribution with 0 mean and $1/m$ variance. Then when $m \geq C\epsilon^{-2} \log(k/\delta)$ for a sufficiently large constant $C > 0$, the following property holds (you do not need to prove it): for any finite set of vectors $V \subset \mathbb{R}^d$ of size $|V| = k$, $|\langle Sv, Sv' \rangle - \langle v, v' \rangle| \leq \epsilon \|v\|_2 \|v'\|_2$ simultaneously for all $v, v' \in V$ with probability $1 - \delta$ over the randomness in S.

You will show that if $m \geq C\epsilon^{-2}(d + \log(1/\delta))$ for a sufficiently large constant $C > 0$, the following holds for any $A \in \mathbb{R}^{n \times d}$ with probability $1 - \delta$ over the randomness in S: for any $x \in \mathbb{R}^d$, $\|SAx\|_2^2 \in [(1-\epsilon)\|Ax\|_2^2, (1+\epsilon)\|Ax\|_2^2]$. This establishes that S is an (ϵ, d, δ)-subspace embedding. You will show this in a few steps.

4a *(7 pts)* Let $\mathcal{S} = \{y \in \mathbb{R}^n : y = Ax \text{ for some } x \in \mathbb{R}^d \text{ and } \|y\|_2 = 1\}$. For any $\gamma > 0$, we say that $\mathcal{N} \subset \mathcal{S}$ is a γ-net of $\mathcal{S}$ if for any $y \in \mathcal{S}$ there exists an $y' \in \mathcal{N}$ such that $\|y - y'\|_2 \leq \gamma$. Show that there exists a γ net $\mathcal{N}$ of $\mathcal{S}$ of size $|\mathcal{N}| \leq (1 + 1/\gamma)^{O(d)}$.

Solution. For $t = \text{rank}(A) \leq d$, we can equivalently express S as

$$\mathcal{S} = \left\{ y \in \mathbb{R}^n \mid y = Ux \text{ for some } x \in \mathbb{R}^t \text{ and } \|y\|_2 = 1 \right\},$$

where U has orthonormal columns and the same column space as A.

We first choose a $\gamma/2$-net $\mathcal{N}'$ of the unit sphere S^{t-1}, where the $\gamma/2$-net has size $\left(1 + \frac{4}{\gamma}\right)^t$.

This can be done by choosing a maximal set $\mathcal{N}'$ of points on S^{t-1} so that no two points are within distance $\gamma/2$ from each other. It follows that the balls of radius $\gamma/4$ centered at these points are disjoint, but on the other hand, they are all contained in the ball of radius $1+\gamma/4$ centered at the origin. The volume of the latter ball is a factor $\left(\frac{1+\gamma/4}{\gamma/4}\right)^t$ larger than the smaller balls, which implies that $|\mathcal{N}'| \leq (1 + 4/\gamma)^t = (1+1/\gamma)^{O(d)}$. Define $\mathcal{N} = \{y \in \mathbb{R}^n \mid y = Ux \text{ for some } x \in \mathcal{N}'\}$. Since the columns of U are orthonormal, if there were a point $Ux \in \mathcal{S}$ for which there were no point $y \in \mathcal{N}$ with $\|y - Ux\|_2 \leq \gamma$, then x would be a point in S^{t-1} for which there is no point $z \in \mathcal{N}'$ with $\|x - z\|_2 \leq \gamma$, a contradiction. □

4b *(8 pts)* Now let $\mathcal{N}$ be a 1/2-net of $\mathcal{S}$. Using this show that for any $y \in \mathcal{S}$, there exists an infinite sequence $y_0, y_1, \ldots$ such that each y_i is a scalar multiple of some point in $\mathcal{N}$, $\|y_i\|_2 \leq 1/2^i$ for all i and

$$y = \sum_{i=0}^{\infty} y_i.$$

Solution. The proof of this is as follows. We can write $y = y_0 + (y - y_0)$, where $y_0 \in \mathcal{N}$ and $\|y - y_0\| \leq \frac{1}{2}$ by the definition of $\mathcal{N}$. Now note that $y - y_0 = U(x - x_0)$ where $y = Ux$ and $y = Ux_0$, thus $\|x - x_0\|_2 = \|y - y_0\|_2 \leq 1/2$. This implies $y - y0$ lies in a scaled version of $\mathcal{S}$ where each element of $\mathcal{S}$ is scaled by $\|x - x_0\|_2 = \|y - y_0\|_2$. Thus we can say there exists a y_1 in a scaled version of $\mathcal{N}$ where each element in $\mathcal{N}$ is scaled by $\|y - y_0\|_2$, such that$y - y_0 = y_1 + ((y - y_0) - y_1)$, and $\|y - y_0 - y_1\|_2 \leq \|y - y_0\|_2/2 \leq 1/4$. Note that $\|y_1\|_2 \leq 1/2$, and repeating this argument gives the proof of the claim. □

4c *(10 pts)* Instantiate random matrix S consider it for the finite set V to be the 1/2-net $V = \mathcal{N}$, thus $S \in \mathbb{R}^{m \times n}$ for $m = O(\epsilon^{-2}(\log(|\mathcal{N}|/\delta)))$. Use the result of **4b** to show that $\|Sy\|_2^2 \in [1 - O(\epsilon), 1 + O(\epsilon)]$ for all $y \in \mathcal{S}$ with probability $1 - \delta$.

Solution. We have the following,

$$\begin{aligned}
\|Sy\|_2^2 &= \|S\left(y_0 + y_1 + y_2 + \cdots\right)\|_2^2 \\
&= \sum_{0 \le i < j < \infty} \|Sy_i\|_2^2 + 2\langle Sy_i, Sy_j\rangle \\
&= \left(\sum_{0 \le i < j < \infty} \|y_i\|_2^2 + 2\langle y_i, y_j\rangle\right) \pm 2\varepsilon \left(\sum_{0 \le i \le j < \infty} \|y_i\|_2 \|y_j\|_2\right) \\
&= \|y_0 + y_1 + \ldots\|_2^2 \pm 2\epsilon \left(\left(\sum_{0 \le i \le \infty} \|y_i\|_2\right)\left(\sum_{0 \le j \le \infty} \|y_i\|_2\right)\right) \\
&= \|y\|_2^2 \pm 2\epsilon \left(\sum_{0 \le i \le \infty} 1/2^i\right)^2 \\
&= 1 \pm O(\varepsilon),
\end{aligned}$$

where the first equality follows from 4b, the second equality follows by expanding the square, the third equality follows from the guarantee of S for the set $\mathcal{N}$. □