# COMPSCI 614: Problem Set 4

**Due: 4/22 by 11:59pm in Gradescope.**

**Instructions:**

- You are allowed to, and highly encouraged to, work on this problem set in a group of up to three members.
- Each group should **submit a single solution set**: one member should upload a pdf to Gradescope, marking the other members as part of their group in Gradescope.
- You may talk to members of other groups at a high level about the problems but **not work through the solutions in detail together**.
- You must show your work/derive any answers as part of the solutions to receive full credit.

**Hint:** The following two inequalities may be helpful throughout the course: for any $x > 0$, $(1 + x)^{1/x} \leq e$ and $(1 - x)^{1/x} \leq 1/e$.

## 1. Tighter Bounds for Trace Estimation (4 points)

Consider any matrix $A \in \mathbb{R}^{n \times n}$. Use the Hanson-Wright inequality to show that if $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \{-1, 1\}^n$ are chosen to have independent and uniformly distributed $\pm 1$ entries, then for $m = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, $\bar{\mathbf{T}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i^T A \mathbf{x}_i$ satisfies,

$$\Pr\left[|\bar{\mathbf{T}} - \operatorname{tr}(A)| > \epsilon \|A\|_F\right] \leq \delta.$$

How does this compare to the bound proven in class using Chebyshev's inequality?

## 2. A Naive Net Bound (4 points)

In class we showed via a volume argument that there is an $\epsilon$-net over the unit ball $\mathcal{S} = \{y \in \mathbb{R}^d : \|y\|_2 = 1\}$ containing $\left(\frac{4}{\epsilon}\right)^d$ points. Consider instead using the following simple net: let $G = [-1, -1 + \delta, -1 + 2\delta, 0, \delta, 2\delta, \ldots, 1 - \delta, 1]$ be a grid of spacing $\delta$ over $[-1, 1]$ and let $\mathcal{N} = G^d$ be the set of all $d$-dimensional vectors on the $d$-dimension grid defined by $G$.

How small must we set $\delta$ such that $\mathcal{N}$ is an $\epsilon$-net for $\mathcal{S}$. How large of a net does this yield? How would using this construction instead of the one shown in class affect our final bounds on the required dimension for subspace embedding?

## 3. Matrix Concentration from Scratch (8 points)

Consider a random symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ where $\mathbf{M}_{ij} = \mathbf{M}_{ji}$ is set independently to 1 with probability 1/2 and $-1$ with probability 1/2. Let $\|\mathbf{M}\|_2 = \max_{x:\|x\|=1} \|\mathbf{M}x\|_2$ be the spectral norm of $\mathbf{M}$. Recall that $\|\mathbf{M}\|_2$ is equal to the largest singular value of $\mathbf{M}$, which equals the largest magnitude of one of its eigenvalues.

1. (2 points) Give upper and lower bounds on $\|\mathbf{M}\|_2$ that hold deterministically – i.e., for any random choice of the entries of $\mathbf{M}$. **Hint:** You may want to use $\|\mathbf{M}\|_F$, and its relation to the singular values to derive your bounds.

2. (2 points) Observe that you can also write $\|\mathbf{M}\|_2 = \max_{x:\|x\|=1} |x^T\mathbf{M}x|$. Show that for any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$, with probability $\geq 1-\delta$, $|x^T\mathbf{M}x| \leq c\sqrt{\log(1/\delta)}$ for some constant $c$.

   **Hint:** Use Hoeffding's inequality, which is a useful variant on the Bernstein inequality. For independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, and scalars $a_1, \ldots, a_n, b_1, \ldots, b_n$ with $\mathbf{X}_i \in [a_i, b_i]$, $\Pr\left[|\sum_{i=1}^n \mathbf{X}_i - \mathbb{E}[\sum_{i=1}^n \mathbf{X}_i]| \geq t\right] \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right)$.

3. (4 points) Prove that with probability $1 - \frac{1}{n^{c_1}}$, $\|\mathbf{M}\|_2 \leq c_2\sqrt{n\log n}$ for some fixed constants $c_1, c_2$. **Hint:** Use an $\epsilon$-net for $\epsilon = 1/n$ and part (1).

## 4. Compressed Sensing From Subspace Embedding (6 points)

Given a vector $x \in \mathbb{R}^n$ and a random matrix $\mathbf{S} \in \mathbb{R}^{m\times n}$, consider computing $\mathbf{y} = \mathbf{S}x$. If $m < n$, you can in general not determine $x \in \mathbb{R}^n$ from $\mathbf{y} \in \mathbb{R}^m$, since $\mathbf{S}$ is not an invertible map. Here, we will argue that you can recover $x$, assuming that it is $k$-sparse for small enough $k$. I.e., that it has at most $k$ nonzero entries. This is known as *compressed sensing* or *sparse recovery*.

1. (2 points) Assume that $\mathbf{S}$ satisfies the distributional JL lemma/subspace embedding theorem proven in class. I.e., for any $A \in \mathbb{R}^{n\times d}$, if $m = O\left(\frac{d+\log(1/\delta)}{\epsilon^2}\right)$, then with probability at least $1-\delta$, $\mathbf{S}$ is an $\epsilon$-subspace embedding for $A$. Prove that if $m = O\left(\frac{k\log(n/k)+\log(1/\delta)}{\epsilon^2}\right)$, with probability $\geq 1-\delta$, for all $z \in \mathbb{R}^n$ such that $z$ is $k$-sparse, $(1-\epsilon)\|z\|_2 \leq \|\mathbf{S}z\|_2 \leq (1+\epsilon)\|z\|_2$.

   **Hint:** Show that with high probability, $\mathbf{S}$ is an $\epsilon$-subspace embedding simultaneously for $\binom{n}{k}$ different matrices.

2. (2 points) Use the above result, applied with $k' = 2k$, to show that if $m = O(k\log(n/k) + \log(1/\delta))$, and $x \in \mathbb{R}^n$ is $k$-sparse, then with probability $\geq 1-\delta$, $x$ can be recovered exactly from $\mathbf{y} = \mathbf{S}x$.

   **Hint:** Consider solving the equation $\mathbf{y} = \mathbf{S}x$, under the restriction that $x$ is $k$-sparse. Show that there is a unique solution.

3. (2 points) Argue that the above result is nearly optimal in terms of how much $x$ is compressed. In particular, prove that for any function $f : \mathbb{R}^n \to \{0,1\}^{o(k\log(n/k))}$, given $f(x)$ for some $k$-sparse $x \in \mathbb{R}^n$, one cannot recover $x$ uniquely, even under the assumption that all entries of $x$ are either 0 or 1.

## 5. Sparse Subspace Embedding (14 points)[1]

In this problem we will show how to construct very efficient subspace embeddings via Count Sketch random matries. In particular, let $\mathbf{S} \in \mathbb{R}^{m\times n}$ be a Count Sketch matrix where for each column we

[1]Credit to Prof. Hung Le for giving me the idea for this problem.

independently pick a single entry uniformly at random and set it to 1 or $-1$, each with probability 1/2. All other entries are set to 0.

You may use the following fact, proven on the midterm exam: for $m = O(\frac{1}{\epsilon^2 \delta})$, for any fixed $y \in \mathbb{R}^n$, $(1-\epsilon)\|y\|_2^2 \leq \|\mathbf{S}y\|_2^2 \leq (1+\epsilon)\|y\|_2^2$ with probability at least $1-\delta$.

1. (2 points) What is the runtime required to multiply $\mathbf{S}$ by a matrix $A \in \mathbb{R}^{n \times d}$? How does this compare to the dense random sign sketching matrices studied in class?

2. (2 points) Using the $\epsilon$-net + union bound proof approach from class, how large would we have to set $m$ to ensure that $\mathbf{S}$ is a subspace embedding for any $A \in \mathbb{R}^{n \times d}$ with probability at least $1-\delta$. How does this compare to what was shown in class for dense random sign matrices? **Hint:** The dependence on the failure probability for norm preservation for Count-Sketch is $1/\delta$, not $\log(1/\delta)$, and this cannot be improved significantly.

3. (2 points) Let $\mathbf{V} \in \mathbb{R}^{n \times d}$ be an orthonormal basis for the column span of $A$. Consider the $d \times d$ matrix $\mathbf{M} = \mathbf{I} - V^T \mathbf{S}^T \mathbf{S} V$. Argue that if $\|\mathbf{M}\|_2 = \max_{x \in \mathbb{R}^k} \frac{|x^T \mathbf{M} x|}{\|x\|_2^2} \leq \epsilon$ then $\mathbf{S}$ is an $\epsilon$-subspace embedding for $A$. **Hint:** Rewrite any $y \in \mathbb{R}^n$ in the column span of $A$ as $Vc$ for some coefficient vector $c$ and expand out $|\|y\|_2^2 - \|\mathbf{S}y\|_2^2|$.

4. (2 points) Prove that for $m = O\left(\frac{d^4}{\delta \epsilon^2}\right)$, with probability at least $1-\delta$, for every pair of columns $v_i, v_j$ of $V$, we have $2 - \frac{\epsilon}{d} \leq \|\mathbf{S}v_i - \mathbf{S}v_j\|_2^2 \leq 2 + \frac{\epsilon}{d}$, and further for every $v_i$, $1 - \frac{\epsilon}{2d} \leq \|\mathbf{S}v_i\|_2^2 \leq 1 + \frac{\epsilon}{2d}$.

   **Hint:** Apply the result form the midterm.

5. (2 points) Prove that if the bounds for part (4) hold, then for all pairs $v_i, v_j$ with $i \neq j$, we have $|v_i^T \mathbf{S}^T \mathbf{S} v_j| \leq \frac{\epsilon}{d}$. **Hint:** Expand out $\|\mathbf{S}v_i - \mathbf{S}v_j\|_2^2$ as an inner product.

6. (2 points) Use part (5) to prove that if the bounds from part (4) hold, then $\|\mathbf{M}\|_F \leq \epsilon$ and in turn that $\|\mathbf{M}\|_2 \leq \epsilon$.

7. (2 points) Conclude that a random Count Sketch matrix $\mathbf{S}$ with $m = O(\frac{d^4}{\epsilon^2 \delta})$ is a subspace embedding for any $A \in \mathbb{R}^{n \times d}$ with probability at least $1-\delta$. How does this compare to the result on dense sketching matrices shown in class?