New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 6763: Homework 4.
Due Monday, May 2nd, 2022, 11:59pm.

Collaboration is allowed on this problem set, but solutions must be written-up individually. Please list collaborators for each problem separately, or write "No Collaborators" if you worked alone.

Problem 1: Optimal Low-Rank Approximation

(10 pts) In class we saw the Eckart–Young–Mirsky theorem, which claimed that the best low-rank approximation to any matrix $X \in \mathbb{R}^{n\times d}$ is given by $XV_kV_k^T$, where $V_k \in \mathbb{R}^{d\times k}$ contains the top k right singular vectors of $X = U\Sigma V^T$ – i.e., the top k eigenvectors of the positive semidefinite matrix X^TX. Here you will prove this from scratch, using just basic properties of projection matrices and eigenvectors.

1. Let $X \in \mathbb{R}^{n\times d}$ be as above, and let $M \in \mathbb{R}^{n\times d}$ be a candidate k rank approximation that has singular value decomposition $M = QDZ^T$ for orthonormal $Q \in \mathbb{R}^{n\times k}$, $Z \in \mathbb{R}^{d\times k}$, and diagonal $D \in \mathbb{R}^{k\times k}$. Prove that $\|X-M\|_F^2 = \|XZZ^T - M\|_F^2 + \|X - XZZ^T\|_F^2$ and conclude that, if $M = \arg\min_{\text{rank } kB} \|X-B\|_F^2$, then $M = XZZ^T$.

2. Using a similar argument as above, one can show that, if $M = \arg\min_{\text{rank } kB} \|X-B\|_F^2$, then $M = QQ^TX$. Use this and part (1) to prove that $X^TXZ = ZD^2$ for any optimal rank k approximation $M = QDZ^T$. Conclude that each column of Z is an eigenvector of X^TX. **Hint: It may be helpful to prove as an intermediate step that $XZ = QD$ and $Q^TX = DZ^T$.**

3. Complete the proof, showing that the best low-rank approximation of X is given by $XV_kV_k^T$ where V_k contains the top k eigenvectors of X^TX.

Problem 2: Matrix Concentration from Scalar Concentration

(15 pts) In this problem, we will show that random matrices have small spectral norms with high probability – this is a form of a *matrix concentration inequality*. Specifically, construct a random matrix $R \in \mathbb{R}^{n\times n}$ by setting R_{ij} to $+1$ or -1, uniformly at random. Prove that, with high probability, we have

$$\|R\|_2 \leq c\sqrt{n\log n},$$

For some constant $c > 0$. This is much better than the naive bound of $\|R\|_2 \leq \|R\|_F = n$ and it's nearly tight: we always have that $\|R\|_2^2 \geq \|R\|_F^2/n$ (do you see why?) so $\|R\|_2 \geq \sqrt{n}$ no matter what.

Here are a few hints that might help you along:

- You can use the following fact (which you may like to prove for yourself as an exercise): for any matrix $R \in \mathbb{R}^{n\times n}$, we have

$$\|R\|_2 = \max_{x,y\in\mathbb{R}^n} \frac{x^TRy}{\|x\|_2\|y\|_2}$$

- To bound $\|R\|_2$, first try to first bound

$$x^TRy = \sum_{i=1}^n \sum_{j=1}^n R_{i,j}x_iy_j$$

  for one particular pair of unit vectors $x, y \in \mathbb{R}^m$ (notice that it suffices to consider the max only over unit vectors). You might want to use a Chernoff-Hoeffding bound,[1] or the Khintshine inequality that we saw in Lecture 10.

[1] See https://en.wikipedia.org/wiki/Hoeffding%27s_inequality

- Then try to extend the result to hold for all pairs $x, y \in \mathbb{R}^n$ simultaneously, using an ϵ-net argument.

For the next part: You will want to generalize your above proof to show the following statement: there exists a constant C such that, for any $\lambda > 1$, we have

$$\Pr[\|R\|_2 \geq \lambda\sqrt{n \log n}] \leq \exp(-C\lambda n \log n)$$

Problem 3: Random Subspaces do not Contain Sparse Vectors

(15 pts) In this problem, we will use the matrix concentration inequality you developed above to prove an important fact about random subspaces: that they do not (even approximately) contain sparse vectors.

We first formalize what it means for a subspace to approximately contain a vector. Recall that, given a k-dimensional linear subspace $\mathcal{U} \subset \mathbb{R}^n$, the orthogonal projection of any vector $x \in \mathbb{R}^n$ onto $\mathcal{U}$ is defined as $\mathbf{V}^T\mathbf{V}x$, where $\mathbf{V} \in \mathbb{R}^{k\times n}$ is a matrix with k *orthonormal* rows which span $\mathcal{U}$. In other words, the rows $v_1 \ldots, v_k \in \mathbb{R}^n$ of $\mathbf{V}$ satisfy $\|v_i\|_2^2 = 1$ for all $i \in [k]$, and $\langle v_i, v_j\rangle = 0$ for all $i \neq j$. Note that if x is contained in the subspace $\mathcal{U}$, the projection of x onto $\mathcal{U}$ is just itself, i.e. $\mathbf{V}^T\mathbf{V}x = x$. Also recall that, for any $x \in \mathbb{R}^n$, by the Pythagorean theorem we have

$$\|x\|_2^2 = \|\mathbf{V}^T\mathbf{V}x\|_2^2 + \|x - \mathbf{V}^T\mathbf{V}x\|_2^2$$

Here, $\mathbf{V}^T\mathbf{V}x$ is the "part" of x living in $\mathcal{U}$, and $x - \mathbf{V}^T\mathbf{V}x$ is the "part" of x orthogonal to $\mathcal{U}$. By the above equation, it is clear that $\|x\|_2^2 \geq \|\mathbf{V}^T\mathbf{V}x\|_2^2$ for all $x \in \mathbb{R}^n$. Moreover, $\|x\|_2^2 = \|\mathbf{V}^T\mathbf{V}x\|_2^2$ if and only if $x \in \mathcal{U}$. Given this, we say that a vector x is *ϵ-approximately contained in $\mathcal{U}$*, for some $\epsilon \in (0, 1)$, if we have:

$$\|\mathbf{V}^T\mathbf{V}x\|_2^2 > \epsilon\|x\|_2^2$$

In other words, at least an ϵ fraction of x lies in the subspace $\mathcal{U}$. Since $\mathbf{V}^T$ has *orthonormal columns*, we have $\|\mathbf{V}^T\mathbf{V}x\|_2 = \|\mathbf{V}x\|_2$, and so it suffices to just consider the norm of $\mathbf{V}x$. Your goal is to show that, if $\mathbf{V}$ spans a *random* k-dimensional subspace (i.e., the rows of $\mathbf{V}$ are random orthonormal vectors), then $\mathbf{V}$ does not ϵ-approximately contain any k-sparse vector, for every $k \leq c\epsilon\frac{n}{\log n}$ (where c is a constant that you can choose). Recall that $x \in \mathbb{R}^n$ is called k-sparse if $\|x\|_0 \leq k$.

Specifically, Let $\mathbf{V} \in \mathbb{R}^{k\times n}$ be a matrix with independent entries $\mathbf{V}_{i,j}$ set to either $\frac{1}{\sqrt{n}}$ or $-\frac{1}{\sqrt{n}}$ uniformly at random.[2] Prove that for every $k \leq c\epsilon\frac{n}{\log n}$, with high probability, we have:

$$\max_{\substack{x\in\mathbb{R}^n\\ \|x\|_0\leq k}} \frac{\|\mathbf{V}x\|_2^2}{\|x\|_2^2} \leq \epsilon$$

Hint 1: Start by considering a specific set $S \subset [n]$ of $|S| = k$ coordinates. Apply the concentration inequality you developed in the last problem to show that, with large probability, $\mathbf{V}$ does not ϵ-approximately contain any k-sparse vector $x \in \mathbb{R}^n$ whose non-zero coordinates are contained in S. Once you have shown this for a *fixed* $S \subset [n]$ with $|S| = k$, proceed to show it for *all* such subsets to complete the proof.

Hint 2: You may want to remember the useful inequality $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.

Bonus: Communicating in the Dark is Easier with Shared Random Coins

(5 pts extra credit) Suppose Alice holds a subset of elements $A \subseteq \{1, \ldots, n\}$. Bob holds another subset $B \subseteq \{1, \ldots, n\}$. Alice and Bob do not know what elements the other holds. Using as little communication as possible, the two of them want to determine if they hold any unique elements – i.e. if there is any $j \in A \cup B - A \cap B$.

Show that, for some constant c, Alice can send Bob a single message of $O(\log^c n)$ bits that allows Bob to find such a j if one exists, with probability at least 2/3.

You can assume that Alice and Bob decide on a strategy in advance, and that they have access to an unlimited source of shared random bits (e.g. that are published by some third party).

[2] You can check that each row of $\mathbf{V}$ has unit norm. However, the rows of $\mathbf{V}$ are not *exactly* orthogonal, but they are pretty close, namely one can show with a Chernoff bound that $\langle \mathbf{V}_i, \mathbf{V}_j\rangle \leq O(\frac{1}{\sqrt{n}})$. So while $\mathbf{V}$ is not exactly orthonormal, it is close enough for the purpose of this problem.