CSE 422 Autumn 26 HW 5

Note: For all problems, you will be expected to explain your reasoning. Providing an expression without reasoning will not receive any partial credit.

Problem 1: Optimality of SVD for low-rank approximation

Recall that if a matrix A has SVD given by $A = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$ are the singular values, and u_i and v_i are the corresponding left and right singular vectors, then for any $k \leq r$, the rank-k truncation of the SVD is given by $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$. In class, we mentioned that the rank-k truncation of SVD offers the best rank-k approximation to a given matrix in Frobenius norm. In this problem, you goal is to show a generalization of this bound. Specifically, show the following theorem:

Theorem 0.1. *Let $p \geq 1$, let $A \in \mathbb{R}^{m \times n}$, and let $k \leq \min(m, n)$. Then*

$$\|A - A_k\|_{S_p} \leq \min_{\text{rank}(N) \leq k} \|A - N\|_{S_p} \ .$$

Here $\|M\|_{S_p}$ is the *Schatten-p norm* of a matrix M, and is defined to be the ℓ_p norm of its singular values. That is, if M has singular values $\nu_1, \ldots, \nu_\ell$, then $\|M\|_{S_p} = (\sum_{i=1}^{p} \nu_i^p)^{1/p}$. It turns out that this is a valid norm as long as $p \geq 1$, and you can also check that $\|M\|_{S_2}$ is equal to the Frobenius norm of M. So the $p = 2$ case of this theorem covers what we mentioned in class.

For this problem, you can use the following *variational* characterization of the singular values without proof, which is a generalization of the minimax characterization we showed in class. Namely, matrix $M \in \mathbb{R}^{m \times n}$ with singular values $\nu_1 \geq \nu_1 \geq \ldots \geq \nu_{\min(m,n)} \geq 0$, and for any $i \in \{1, \ldots, \min(m, n)\}$, we have that

$$\nu_i = \max_{\substack{\text{subspaces } V \\ \dim(V) = i}} \min_{\substack{u \in V \\ \|u\|_2 = 1}} \|Mu\|_2 \ .$$

Problem 2: Does low-rank approximation preserve distances?

We argued briefly in class that low-rank approximation is not quite the right tool for preserving distances. In this problem, we'll explore a weak sense in which it still does, but also its limitations as a tool for preserving distances (as opposed to JL).

Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be a set of data vectors satisfying $\sum_{i=1}^n x_i = 0$, and let $X \in \mathbb{R}^{n \times d}$ be the data matrix whose i-th row is $x_i^\top$. For any subspace V, let Π_V denote the projector onto that subspace.

(a) [10 points] For any subspace V, denote the expected distortion of X onto V, denoted $d(X, V)$, to be

$$d(X, V) = \mathbb{E}_{i,j}\left[\left|\|x_i - x_j\|_2^2 - \|\Pi_V x_i - \Pi_V x_j\|_2^2\right|\right] ,$$

where there the i, j are taken to be drawn uniformly at random. What is the subspace of dimension k that achieves the smallest expected distortion?

(b) [5 points] Does this subspace also satisfy the Johnson-Lindenstrauss guarantee? If so, explain why, and if not, give a counter-example.

(c) [5 points] Can you informally explain the difference between this expected distortion objective and the metric distortion (think Johnson-Lindenstrauss) objective we discussed in week 2? Write down what JL is tacitly optimizing for, and compare this objective to the expected distortion objective.

Problem 3: Dimension reduction for tensor spectral norm

Recall that in class, we said that most tensor optimization problems are hard, including tensor injective norm. Let $T \in \mathbb{R}^{n \times n \times n}$, and define the injective norm of T to be

$$\|T\|_{\text{inj}} = \max_{\|u\|=\|v\|=\|w\|=1} |T(u,v,w)| \ .$$

This is one of the many tensor optimization problems which are hard. However, in certain regimes, it does admit non-trivial approximation algorithms. In this problem, your goal is to more or less reconstruct one such algorithm.

Recall that the Frobenius norm of a tensor is the ℓ_2 norm of the tensor, treated as a vector:

$$\|T\|_F^2 = \sum_{i,j,k} T_{ijk}^2 \ .$$

Assume that $\|T\|_F = 1$. Then, we claim that for all $\varepsilon > 0$, there is an algorithm which runs in time $f(1/\varepsilon) \cdot \text{poly}(n)$, which outputs unit vectors u', v', w' so that

$$T(u', v', w') \geq \|T\|_{\text{inj}} - \varepsilon \ .$$

Here f is some potentially very fast-growing function, but crucially, it depends only on ε, and not on n.

(a) [7 points] Show the following, seemingly unrelated fact: if a matrix $M \in \mathbb{R}^{m \times n}$ has Frobenius norm 1, then one can efficiently find a subspace $U \subset \mathbb{R}^n$ of dimension at most $O(1/\varepsilon^2)$ so that for all unit vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ so that

$$|u^T M \Pi_V v| \geq (1-\varepsilon)|u^T M v| \ .$$

Here Π_V means the same thing as in Problem 2.

(b) [7 points] Show that this implies that there exists an efficient algorithm which finds subspaces U, V, W, each of dimension $O(1/\varepsilon^2)$, so that

$$|T(\Pi_U u, \Pi_V v, \Pi_W w)| \geq |T(u,v,w)| - \varepsilon \ ,$$

for all unit vectors u, v, w.

(c) [6 points] Informally describe how this yields the desired algorithm.