Homework 5 Sample Solutions

Problem 1: Optimality of SVD for low-rank approximation

Let $r = \min(m, n)$ and write $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$ for the singular values of A, including zeros. Fix any matrix N of rank at most k. We will show that the singular values of the error $A - N$ satisfy

$$\sigma_i(A - N) \geq \sigma_{k+i}(A), \qquad 1 \leq i \leq r - k.$$

Let $L = \operatorname{span}(v_1, \ldots, v_{k+i})$, where the v_j are right singular vectors of A. Since $\ker N$ has codimension at most k, the restriction of N to L has a kernel of dimension at least $(k + i) - k = i$. Choose an i-dimensional subspace V of $L \cap \ker N$. For any unit vector $z = \sum_{j=1}^{k+i} a_j v_j$ in V we have

$$\|(A - N)z\|_2^2 = \|Az\|_2^2 = \sum_{j=1}^{k+i} \sigma_j^2 a_j^2 \geq \sigma_{k+i}^2.$$

Using this V in the variational characterization of $\sigma_i(A - N)$ proves the claimed singular value bound.

The nonzero singular values of $A - A_k$ are exactly the positive discarded singular values of A. Thus for every finite $p \geq 1$,

$$\|A - N\|_{S_p}^p = \sum_{i=1}^{r} \sigma_i(A - N)^p \geq \sum_{i=1}^{r-k} \sigma_{k+i}(A)^p = \|A - A_k\|_{S_p}^p.$$

Taking pth roots, and observing that A_k itself has rank at most k, gives

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

Problem 2: Does low-rank approximation preserve distances?

(a) Let $P = \Pi_V$, and put $r_i = (I - P)x_i$. The Pythagorean identity gives

$$\|x_i - x_j\|_2^2 - \|Px_i - Px_j\|_2^2 = \|r_i - r_j\|_2^2 \geq 0.$$

The residuals are centered because $\sum_i x_i = 0$. For independently uniform i, j, the cross term satisfies $\mathbb{E}_{i,j}\langle r_i, r_j\rangle = \|\frac{1}{n}\sum_i r_i\|_2^2 = 0$. Thus

$$d(X, V) = \mathbb{E}_{i,j}\|r_i - r_j\|_2^2 = \frac{2}{n}\sum_i \|r_i\|_2^2 = \frac{2}{n}\|X - XP\|_F^2.$$

Since $\operatorname{rank}(XP) \leq k$ when $\dim V = k$, Problem 1 with $p = 2$ gives a lower bound of $\frac{2}{n}\|X - X_k\|_F^2$. The subspace spanned by the top k right singular vectors of X attains that bound, because its projection satisfies $XP = X_k$. So

$$\boxed{V = \text{the top } k \text{ principal component subspace of } X.}$$

(b) Take the four centered points $(2, 0), (-2, 0), (0, 1), (0, -1)$ and $k = 1$. Here $X^\top X = \operatorname{diag}(8, 2)$, so the unique top principal component line is the horizontal axis. The two vertical points have distance 2 but both project to $(0, 0)$. Their distance cannot satisfy a multiplicative JL guarantee.

(c) The are two main distinctions: first, the JL preserves *all* distances, whereas this objective is preserving distances "on average". Second, JL preserves distances up to multiplicative factors (i.e. $\|x_i - x_j\|_2 \approx (1 + \epsilon)\|F(x_i) - F(x_j)\|_2$ vs $|\|x_i - x_j\|_2 - \|F(x_i) - F(x_j)\|_2| \leq \epsilon$), whereas this objective only preserves them up to additive factors.

In other words, what JL is tacitly optimizing is the embedding F that minimizes the maximum relative error

$$\max_{i,j:\ x_i \neq x_j} \left| \frac{\|F(x_i) - F(x_j)\|_2^2}{\|x_i - x_j\|_2^2} - 1 \right|.$$

By contrast, $d(X, V)$ averages absolute errors in squared distance.

Problem 3: Dimension reduction for tensor spectral norm

(a) *Erratum: the bound is supposed to be* additive*: namely, that there are low-dimensional subspaces so that*

$$\boxed{\left|u^\top M v - u^\top M \Pi_V v\right| \le \epsilon \quad \textit{for } \|u\|_2, \|v\|_2 \le 1.}$$

Given $\epsilon > 0$, let V span the right singular vectors of M whose singular values exceed τ. There are at most $1/\epsilon^2$ of them, since $\sum_i \sigma_i(M)^2 = 1$. The remaining singular values are at most τ, so $\|M(I - \Pi_V)\|_{\text{op}} \le \epsilon$. Consequently

$$\boxed{\left|u^\top M v - u^\top M \Pi_V v\right| \le \epsilon \quad \text{for } \|u\|_2, \|v\|_2 \le 1.}$$

The SVD computes this subspace in polynomial time.

(b) Unfold T along each of its three modes. The first unfolding is the $n^2 \times n$ matrix M_1 with entries $(M_1)_{(j,k),i} = T_{ijk}$, so $T(u, v, w) = (v \otimes w)^\top M_1 u$. The other two unfoldings play the same role for v and w. Each unfolding rearranges entries, so its Frobenius norm is 1.

Apply (a) to the three unfoldings with $\tau = \epsilon/3$ to obtain U, V, W of dimension $O(1/\epsilon^2)$. Write $u_P = \Pi_U u$, $v_P = \Pi_V v$, and $w_P = \Pi_W w$. Project one argument at a time:

$$\begin{aligned} T(u, v, w) - T(u_P, v_P, w_P) = &\big[T(u, v, w) - T(u_P, v, w)\big] \\ &+ \big[T(u_P, v, w) - T(u_P, v_P, w)\big] \\ &+ \big[T(u_P, v_P, w) - T(u_P, v_P, w_P)\big]. \end{aligned}$$

Each bracket has magnitude at most τ by the corresponding unfolding and (a), since the other arguments have norm at most 1 and $\|v \otimes w\|_2 = \|v\|_2 \|w\|_2$. The triangle inequality yields

$$\boxed{|T(u_P, v_P, w_P)| \ge |T(u, v, w)| - \epsilon.}$$

(c) The subspaces U, V, W from part (b) are all formed by SVD on $(n^2 \times n)$-sized matrices, so they can all be computed in polynomial time. By part (b), we know that to find a solution, it suffices to restrict our search within such subspaces. But each subspace is of dimension $O(1/\epsilon^2)$; in particular, its dimensionality is independent of n. So we can do some extremely slow brute force search over these subspaces to solve the problem exactly restricted to these subspaces, yielding a runtime of $f(1/\epsilon) \cdot \text{poly}(n, 1/\epsilon)$.

More formally, but you did not have to write this out: Search over η-nets on the three unit spheres for $\eta = \epsilon/6$. By triangle inequality, moving each input to the tensor by at most η changes T by at most $3\eta = \epsilon/2$ in total. In dimension r a net can be built by normalizing nonzero grid points of mesh $\eta/\sqrt{r}$ near the sphere; rounding and normalizing give error at most η. Each search dimension is $O(1/\epsilon^2)$, so the number of triples depends only on ϵ. Evaluate each triple in polynomial time and choose the largest absolute value, flipping one sign if needed. The returned unit vectors satisfy $\boxed{T(u', v', w') \ge \|T\|_{\text{inj}} - \epsilon}$, in time $f(1/\epsilon)\,\text{poly}(n)$. For $\epsilon \ge 1$, any unit triple oriented to give a nonnegative value suffices, since $\|T\|_{\text{inj}} \le 1$.