LOW RANK APPROXIMATION IN THE FROBENIUS NORM AND PERTURBATION OF SINGULAR VALUES

MRINAL RAGHUPATHI

1. Introduction

In these notes we will prove that the rank-p approximation obtained from the SVD is the best approximation in the Frobenius norm. The ideas involved are a elementary but subtle. The proof is due to Weyl and his ideas on eigenvalue perturbation. This beautiful piece of pure mathematics has applications to graph theory, networks, computer science, random matrix theory and statistics, to name just a few.

Definition 1. Let $A \in \mathbb{C}^{m\times n}$ and let $A = U\Sigma V^*$ be a SVD for A. We call $A_p = \sum_{i=1}^p \sigma_i u_i v_i^*$ the rank-p SVD approximation to A.

2. Weyl's ideas

Let us recall some results about the connection between the Frobenius norm and the singular values of a matrix. Suppose that $A \in \mathbb{C}^{m\times n}$ and that $\sigma_1, \ldots, \sigma_\mu$ are the singular values of A, where $\mu = \min\{n, m\}$. Then we have seen that $\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_\mu^2$ and $\|A\|_2 = \sigma_1$. Further since $A - A_p = \sum_{i=p+1}^r \sigma_i u_i v_i^*$ we see that $\|A - A_p\|_2 = \sigma_{p+1}$ and $\|A - A_p\|_F^2 = \sigma_{p+1}^2 + \cdots + \sigma_\mu^2$.

The proof is divided into four steps. Since there will be several matrices and singular values involved we will denote by $\sigma_i(A)$ the ith largest singular value with the convention that $\sigma_i(A) = 0$ for $i > \operatorname{rank}(A)$.

Lemma 1. *Let $A \in \mathbb{C}^{m\times n}$ and let $B \in \mathbb{C}^{m\times n}$ with $\operatorname{rank}(B) \le p$. Then*

$$\sigma_1(A - B) \ge \sigma_{p+1}(A).$$

Proof. Recall that $\sigma_1(A) = \|A\|$ and so for any unit vector v we have that $\sigma_1(A-B)^2 = \|A - B\|_2^2 \ge v^*(A-B)^*(A-B)v$. We will now choose $v \in \operatorname{span}\{v_1, \ldots, v_{p+1}\}$ carefully so that the right hand side reduces to $\|Av\| \ge \sigma_{p+1}$.

Since $\operatorname{rank}(B) \le p$ it follows from the rank-nullity theorem that $\dim(N(B)) \ge n-p$. Hence, there exists a unit vector in the intersection of $\operatorname{span}\{v_1, \ldots, v_{p+1}\}$ and $N(B)$. It follows that for this choice of vector that $v^*(A-B)^*(A-B)v = v^*A^*Av = \|Av\|^2$.

Now using the SVD we see that $Av = \sum_{i=1}^{\mu} \sigma_i u_i v_i^* v = \sum_{i=1}^{p+1} \sigma_i u_i \langle v, v_i \rangle$. Computing the norm of this we get that $\|Av\|^2 = \sum_{i=1}^{p+1} \sigma_i^2 |\langle v, v_i \rangle|^2 \geq \sigma_{p+1}^2$. Hence, $\sigma_1(A-B) \geq \sigma_{p+1}(A)$ □

Lemma 2. *Let $A = A' + A''$. Then $\sigma_{i+j-1}(A) \leq \sigma_i(A') + \sigma_j(A'')$.*

Proof. We first prove this for the case where $i = j = 1$. We have,

$$\sigma_1(A) = u_1^* A v_1 = u_1^*(A' + A'')v_1 = u_1^* A' v_1 + u_1^* A'' v_1 \leq \sigma_1(A') + \sigma_1(A'').$$

For any matrix A, we have $\sigma_1(A - A_p) = \sigma_{p+1}(A)$. It follows from the previous calculation we get

$$\sigma_i(A') + \sigma_j(A'') = \sigma_1(A' - A'_{i-1}) + \sigma_1(A'' - A''_{j-1}) \geq \sigma_1(A' + A'' - (A'_{i-1} + A''_{j-1})).$$

Note that $\operatorname{rank}(A'_{i-1} + A''_{j-1}) \leq i + j - 2$. Now we apply Lemma 1 to get

$$\sigma_1(A' + A'' - (A'_{i-1} + A''_{j-1})) \geq \sigma_{(i+j-2)+1}(A' + A'') = \sigma_{i+j-1}(A).$$

□

Lemma 3. *Let $A \in \mathbb{C}^{m \times n}$ and let $B \in \mathbb{C}^{m \times n}$ with $\operatorname{rank}(B) \leq p$. Then $\sigma_i(A - B) \geq \sigma_{i+p}(A)$.*

Proof. Since $\operatorname{rank}(B) \leq p$ we see that $\sigma_{p+1}(B) = 0$. Apply Lemma 2 with $A' = A - B$ and $A'' = B$. ence,

$$\sigma_i(A - B) = \sigma_i(A - B) + \sigma_{p+1}(B) \geq \sigma_{i+p+1-1}((A - B) + B) = \sigma_{i+p}(A).$$

□

We can now prove our theorem.

Theorem 4. *Let $B \in \mathbb{C}^{m \times n}$ with $\operatorname{rank}(B) \leq p$ and consider the matrix $E = A - B$. Then,*

$$\|A - B\|_F^2 = \sigma_1(A-B)^2 + \cdots + \sigma_\mu(A-B)^2 \geq \sigma_{p+1}(A)^2 + \cdots + \sigma_\mu(A)^2 = \|A - A_p\|_F^2.$$

Proof. We have,

$$\|A - B\|_F^2 = \sum_{i=1}^{\mu} \sigma_i(A - B)^2 \geq \sum_{i=1}^{\mu - p} \sigma_{i+p}(A)^2 = \sigma_{p+1}(A)^2 + \cdots + \sigma_\mu(A)^2.$$

□

USNA

E-mail address: raghupat@usna.edu