Lecture 18: Eckart–Young Theorem

1 Singular Values, Matrix Norms, and Truncated SVD

Suppose

$$A \in \mathbb{C}^{m \times n}$$

has rank r, with compact singular value decomposition

$$A = U_r \Sigma_r V_r^* = \sum_{j=1}^{r} \sigma_j u_j v_j^*,$$

where

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

The singular values tell us not only the rank of A, but also how large A is in different directions. This will allow us to approximate A using a matrix of smaller rank.

1.1 Two Matrix Norms

We will use two ways to measure the size of a matrix.

Definition (Frobenius Norm)**.** *For*

$$A = (A_{ij}) \in \mathbb{C}^{m \times n},$$

the Frobenius norm is

$$\boxed{\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2 \right)^{1/2}.}$$

Equivalently,

$$\|A\|_F^2$$

is the sum of the squared Euclidean norms of the columns of A.

Definition (ℓ_2 Operator Norm)**.** *For*

$$A \in \mathbb{C}^{m \times n},$$

define

$$\boxed{\|A\|_{2 \to 2} = \max_{\|x\|_2 = 1} \|Ax\|_2 .}$$

Thus $\|A\|_{2 \to 2}$ measures the largest factor by which A can stretch a vector.

1.2 Unitary Matrices Preserve These Norms

If Q is unitary, then

$$\|Qx\|_2 = \|x\|_2 .$$

Therefore, if U and V are unitary,

$$\boxed{\|UA\|_F = \|A\|_F = \|AV\|_F}$$

and

$$\boxed{\|UA\|_{2\to 2} = \|A\|_{2\to 2} = \|AV\|_{2\to 2} .}$$

For the Frobenius norm, left multiplication by U preserves the norm of every column. Right multiplication follows by applying the same argument to the conjugate transpose.

1.3 Matrix Norms from the Singular Values

Let

$$A = U\Sigma V^*$$

be a full SVD, and let

$$q = \min\{m, n\}.$$

Since multiplication by unitary matrices does not change either norm, we may measure A by measuring Σ.

For the Frobenius norm,

$$\|A\|_F = \|\Sigma\|_F .$$

Since the only nonzero entries of Σ are the singular values,

$$\boxed{\|A\|_F^2 = \sum_{j=1}^{q} \sigma_j^2.}$$

For the operator norm,

$$\|A\|_{2\to 2} = \|\Sigma\|_{2\to 2} .$$

For any unit vector z,

$$\|\Sigma z\|_2^2 = \sum_{j=1}^{q} \sigma_j^2 |z_j|^2 \le \sigma_1^2 \sum_{j=1}^{q} |z_j|^2 \le \sigma_1^2.$$

Equality is achieved by taking $z = e_1$. Therefore,

$$\boxed{\|A\|_{2\to 2} = \sigma_1.}$$

Thus:

$$\boxed{\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_q^2, \qquad \|A\|_{2\to 2} = \sigma_1.}$$

The Frobenius norm measures the total size of all singular directions, while the operator norm measures only the largest one.

1.4 Truncating the SVD

Suppose

$$1 \leq s < r.$$

The rank-one expansion gives

$$A = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^*.$$

If we keep only the first s terms, define

$$\boxed{A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^*.}$$

Equivalently,

$$\boxed{A_s = U_s \Sigma_s V_s^*,}$$

where

$$U_s = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_s \\ | & & | \end{pmatrix}, \qquad V_s = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_s \\ | & & | \end{pmatrix},$$

and

$$\Sigma_s = \operatorname{diag}(\sigma_1, \ldots, \sigma_s).$$

Since A_s is a sum of s rank-one matrices,

$$\operatorname{rank}(A_s) \leq s.$$

In fact, because $\sigma_1, \ldots, \sigma_s > 0$,

$$\boxed{\operatorname{rank}(A_s) = s.}$$

This matrix is called the *truncated SVD* of A.

1.5 The Approximation Error

Subtracting A_s from A gives

$$A - A_s = \sum_{j=s+1}^{r} \sigma_j u_j v_j^*.$$

Thus the nonzero singular values of $A - A_s$ are

$$\sigma_{s+1}, \ldots, \sigma_r.$$

Therefore,

$$\boxed{\|A - A_s\|_{2\to 2} = \sigma_{s+1},}$$

and

$$\boxed{\|A - A_s\|_F^2 = \sum_{j=s+1}^{r} \sigma_j^2.}$$

So the singular values tell us exactly how much information is lost by discarding the smaller singular directions.

1.6 Why Low Rank Helps with Storage

A general matrix

$$A \in \mathbb{C}^{m \times n}$$

requires mn numbers to store.

A rank-s truncated SVD

$$A_s = U_s \Sigma_s V_s^*$$

can be stored using approximately

$$ms + ns + s$$

numbers.

When

$$s \ll \min\{m, n\},$$

this can be much smaller than mn.

The natural question is now:

$$\boxed{\text{Is } A_s \text{ actually the best rank-}s \text{ approximation to } A?}$$

The answer is yes.

2 Optimal Low-Rank Approximation

Let

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*, \qquad \sigma_1 \geq \cdots \geq \sigma_r > 0,$$

and define

$$A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^*.$$

We have already seen that A_s has rank s.

We now show that no other rank-s matrix can approximate A better.

2.1 Best Approximation in Operator Norm

Theorem 2.1 (Best Rank-s Approximation in Operator Norm)**.** *Let*

$$B \in \mathbb{C}^{m \times n}$$

have rank at most s. Then

$$\boxed{\|A - B\|_{2\to 2} \geq \sigma_{s+1}.}$$

Since

$$\|A - A_s\|_{2\to 2} = \sigma_{s+1},$$

the truncated SVD A_s is a best rank-s approximation to A in operator norm.

Proof. Since

$$\operatorname{rank}(B) \leq s,$$

the rank-nullity theorem gives

$$\dim \mathcal{N}(B) \geq n - s.$$

Let

$$L = \operatorname{span}\{v_1, \ldots, v_{s+1}\}.$$

Then

$$\dim L = s + 1.$$

Therefore,

$$\dim L + \dim \mathcal{N}(B) > n.$$

Hence

$$L \cap \mathcal{N}(B)$$

contains a nonzero vector.

Choose

$$x \in L \cap \mathcal{N}(B)$$

with

$$\|x\|_2 = 1.$$

Since $Bx = 0$,

$$(A - B)x = Ax.$$

Also, since $x \in L$, we may write

$$x = \sum_{j=1}^{s+1} \alpha_j v_j, \qquad \sum_{j=1}^{s+1} |\alpha_j|^2 = 1.$$

Therefore,

$$Ax = \sum_{j=1}^{s+1} \sigma_j \alpha_j u_j,$$

and hence

$$\|Ax\|_2^2 = \sum_{j=1}^{s+1} \sigma_j^2 |\alpha_j|^2.$$

Since

$$\sigma_j \geq \sigma_{s+1} \qquad \text{for } j \leq s + 1,$$

we obtain

$$\|Ax\|_2^2 \geq \sigma_{s+1}^2 \sum_{j=1}^{s+1} |\alpha_j|^2 = \sigma_{s+1}^2.$$

Thus

$$\|A - B\|_{2\to 2} \geq \|(A - B)x\|_2 = \|Ax\|_2 \geq \sigma_{s+1}.$$

□

The proof has a simple interpretation: a rank-s matrix B must kill at least $n-s$ directions. Among the first $s+1$ right singular directions of A, there is always a nonzero direction that B kills. On that direction, the error cannot be smaller than σ_{s+1}.

2.2 Best Approximation in Frobenius Norm

We now prove the analogous result for the Frobenius norm.

Theorem 2.2 (Best Rank-s Approximation in Frobenius Norm)**.** *Let*

$$B \in \mathbb{C}^{m\times n}$$

have rank at most s. Then

$$\boxed{\|A-B\|_F^2 \geq \sum_{j=s+1}^{r} \sigma_j^2.}$$

Since

$$\|A-A_s\|_F^2 = \sum_{j=s+1}^{r} \sigma_j^2,$$

the truncated SVD A_s is also a best rank-s approximation to A in Frobenius norm.

Proof. Let

$$S = \mathcal{C}(B),$$

and let

$$P$$

be the orthogonal projection onto S.

Since

$$\dim S = \operatorname{rank}(B) \leq s,$$

the subspace S has dimension at most s.

Write the columns of A and B as

$$A = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix}, \qquad B = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{pmatrix}.$$

Each b_j lies in S. Since Pa_j is the closest vector in S to a_j,

$$\|a_j - b_j\|_2 \geq \|a_j - Pa_j\|_2 .$$

Summing over all columns gives

$$\|A-B\|_F^2 \geq \|(I-P)A\|_F^2 .$$

Now use the SVD

$$A = U\Sigma V^*.$$

Since right multiplication by the unitary matrix V preserves the Frobenius norm,

$$\|(I-P)A\|_F^2 = \|(I-P)U\Sigma\|_F^2 .$$

Therefore,

$$\|(I-P)A\|_F^2 = \sum_{j=1}^{r} \sigma_j^2 \, \|(I-P)u_j\|_2^2 .$$

Set

$$\alpha_j = \|Pu_j\|_2^2 .$$

Since

$$\|u_j\|_2 = 1,$$

we have

$$\|(I-P)u_j\|_2^2 = 1 - \alpha_j.$$

Thus,

$$\|(I-P)A\|_F^2 = \sum_{j=1}^{r} \sigma_j^2 - \sum_{j=1}^{r} \sigma_j^2 \alpha_j.$$

Clearly,

$$0 \le \alpha_j \le 1.$$

To control their sum, let

$$p_1, \ldots, p_t$$

be an orthonormal basis for S, where $t \le s$. Since

$$Pu_j = \sum_{k=1}^{t} \langle p_k, u_j \rangle p_k,$$

we have

$$\|Pu_j\|_2^2 = \sum_{k=1}^{t} |\langle p_k, u_j \rangle|^2.$$

Because

$$u_1, \ldots, u_m$$

form an orthonormal basis of $\mathbb{C}^m$,

$$\sum_{j=1}^{m} \|Pu_j\|_2^2 = \sum_{k=1}^{t} \sum_{j=1}^{m} |\langle p_k, u_j \rangle|^2 = \sum_{k=1}^{t} \|p_k\|_2^2 = t.$$

Therefore,

$$\sum_{j=1}^{r} \alpha_j \le t \le s.$$

Since

$$\sigma_1^2 \ge \cdots \ge \sigma_r^2,$$

the largest possible value of

$$\sum_{j=1}^{r} \sigma_j^2 \alpha_j$$

is obtained by placing all available weight on the first s terms. Hence,

$$\sum_{j=1}^{r} \sigma_j^2 \alpha_j \le \sum_{j=1}^{s} \sigma_j^2.$$

Therefore,

$$\|(I - P)A\|_F^2 \ge \sum_{j=s+1}^{r} \sigma_j^2.$$

Combining the inequalities,

$$\|A - B\|_F^2 \ge \sum_{j=s+1}^{r} \sigma_j^2.$$

□

2.3 The Eckart–Young Theorem

Combining the two results gives the central low-rank approximation theorem.

Theorem 2.3 (Eckart–Young)**.** *Let*

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*, \qquad \sigma_1 \ge \cdots \ge \sigma_r > 0.$$

For $s < r$, define

$$A_s = \sum_{j=1}^{s} \sigma_j u_j v_j^*.$$

Then A_s is a best rank-s approximation to A in both the operator norm and the Frobenius norm. Moreover,

$$\boxed{\|A - A_s\|_{2\to 2} = \sigma_{s+1},}$$

and

$$\boxed{\|A - A_s\|_F^2 = \sum_{j=s+1}^{r} \sigma_j^2.}$$

2.4 Interpretation

The SVD orders the rank-one pieces of A from largest to smallest:

$$A = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots .$$

The truncated SVD keeps the s most important pieces:

$$A_s = \sigma_1 u_1 v_1^* + \cdots + \sigma_s u_s v_s^*.$$

The Eckart–Young theorem says that no other rank-s matrix can do better.