The SVD and low-rank approximation

MATH 6610 Lecture 10

September 25, 2020

Trefethen & Bau: Lectures 4, 5

The SVD

Recall: a(ny) rectangular matrix $A \in \mathbb{C}^{m \times n}$ has the decomposition,

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

where $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary.
The matrix $\Sigma \in C^{m \times n}$ is diagonal with non-negative entries.

The SVD and eigenvalues

The singular values of $A \in \mathbb{C}^{m \times n}$,

$$\{\sigma_1(A), \ldots, \sigma_p(A)\}, \quad p = \min(m, n),$$

coincide with the eigenvalues λ_j of both AA^* and A^*A:

$$\{\lambda_1(AA^*), \ldots \lambda_p(AA^*)\} = \{\lambda_1(A^*A), \ldots \lambda_p(A^*A)\}.$$

The SVD and fundamental spaces

The SVD gives explicit, orthonormal bases for the fundamental subspaces of A:

$$\mathrm{range}(A),\ \ \ker(A),\ \ \mathrm{range}(A^*),\ \ \ker(A^*).$$

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^* \quad (\text{"reduced" SVD})$$

$\mathrm{range}(A) = \mathrm{span}\{u_1, \ldots, u_r\}$, $\ker(A^*) = \mathrm{span}\{u_{r+1}, \ldots, u_m\}$

$\ker(A) = \mathrm{span}\{v_{r+1}, \ldots, v_n\}$, $\mathrm{range}(A^*) = \mathrm{span}\{v_1, \ldots, v_r\}$.

Are the eigenvalues of a (square) matrix A related to singular values?

<u>Nope</u> (Recall: $\rho(A) \leq \|A\|_2$)

<u>But</u> if A is normal: if $\lambda_1(A) \dots \lambda_n(A)$ are ordered by magnitude (decreasing),

then: $\sigma_j(A) = |\lambda_j(A)|$.

Low-rank approximation

For a matrix $A \in \mathbb{C}^{m \times n}$, a common task is to form a rank-r approximation to A:

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

$$A \approx B, \qquad \operatorname{rank}(B) \leqslant r.$$

not the rank of A. Some user-defined rank.

(Of course this is only interesting if $r < \operatorname{rank}(A)$.)

If $r \geq \operatorname{rank}(A)$: set $B = A$

Low-rank approximation

For a matrix $A \in \mathbb{C}^{m \times n}$, a common task is to form a rank-r approximation to A:

$$A \approx B, \qquad \operatorname{rank}(B) \leqslant r.$$

(Of course this is only interesting if $r < \operatorname{rank}(A)$.)

Theorem ((Schmidt)-Eckart-Young-Mirsky)

Let $A \in \mathbb{C}^{m \times n}$ have SVD $A = U\Sigma V^*$. Then

$$\sum_{j=1}^{r} \sigma_j \left(u_j v_j^*\right) = \boxed{\underset{\substack{B \in \mathbb{C}^{m \times n} \\ \operatorname{rank}(B) \leqslant r}}{\arg\min} \|A - B\|_*},$$

where $\|\cdot\|_*$ is either the induced 2-norm or Frobenius norm of a matrix.

Low-rank approximation

For a matrix $A \in \mathbb{C}^{m \times n}$, a common task is to form a rank-r approximation to A:

$$A \approx B, \qquad \mathrm{rank}(B) \leqslant r.$$

(Of course this is only interesting if $r < \mathrm{rank}(A)$.)

Theorem ((Schmidt)-Eckart-Young-Mirsky)

Let $A \in \mathbb{C}^{m \times n}$ have SVD $A = U\Sigma V^*$. Then

$$\sum_{j=1}^{r} \sigma_j \left(u_j v_j^*\right) = \underset{\substack{B \in \mathbb{C}^{m \times n} \\ \mathrm{rank}(B) \leqslant r}}{\arg\min} \; \|A - B\|_*,$$

where $\|\cdot\|_*$ is either the induced 2-norm or Frobenius norm of a matrix.

This theorem is the basis for innumerable applications in matrix approximation, data compression and summarization, and model acceleration and reduction.