Mostafa Sabri

# Lecture 26

7.4 Singular Value Decomposition

جامعة نيويورك أبوظبي
NYU ABU DHABI

# Beyond Diagonalization

We saw that any real symmetric square matrix can be diagonalized. Unfortunately this cannot be done for matrices which are not symmetric, e.g. $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, or not even square, e.g. $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

Diagonalization can still be performed for complex matrices that are Hermitian, meaning that $A^* = A$, where $A^* = \overline{A^T}$ is the conjugate transpose.

A general square matrix can still be put in *Jordan form*. Instead of being similar to a diagonal matrix, $A$ is shown to be similar to an upper triangular matrix that has a nice shape.

# Beyond Diagonalization

In this section we discuss a different decomposition. Its advantage is that it is applicable to any $m \times n$ matrix. Its drawback is that we no longer have a similarity relation. Instead of having $B = P^{-1}AP$ for a nice matrix $B$ (diagonal or triangular), and thus $A = PBP^{-1}$, we shall have $A = U\Sigma V^T$ for two orthogonal matrices $U, V$ which are in general distinct. In particular, the “nice” matrix $\Sigma$ no longer captures the eigenvalues of $A$, it captures its *singular values* instead.

The singular values $\{\sigma_1, \ldots, \sigma_n\}$ of an $m \times n$ matrix $A$ are the square roots of the eigenvalues $\{\lambda_i\}$ of the $n \times n$ matrix $A^T A$, i.e. $\sigma_i = \sqrt{\lambda_i}$.

# Singular Values

A good thing about $A^T A$ is that it is a square matrix and is symmetric. This motivates considering it. Another good thing is that its eigenvalues are not only real, but actually nonnegative. To see this, note that if $\lambda_i$ is such an eigenvalue with normalized eigenvector $\mathbf{v}_i$, then

$$\|A\mathbf{v}_i\|^2 = \langle A\mathbf{v}_i, A\mathbf{v}_i\rangle = \langle A^T A\mathbf{v}_i, \mathbf{v}_i\rangle = \lambda_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle = \lambda_i$$

so $\lambda_i = \|A\mathbf{v}_i\|^2 \geq 0$. This shows that $\sigma_i = \sqrt{\lambda_i}$ are also nonnegative numbers.

This calculation also gives a geometric meaning to $\sigma_i$. It is equal to $\|A\mathbf{v}_i\|$, where $\mathbf{v}_i$ is a unit eigenvector of $A^T A$. If we arrange the eigenvalues of $A^T A$ as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, then we can say more.

# Singular Values and Stretching

Thm: The largest singular value $\sigma_1 = \sqrt{\lambda_1}$ satisfies

$$\sigma_1 = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|\,.$$

In other words, $\sigma_1$ is the maximum amount by which $A$ stretches a unit vector. This maximal stretching occurs in the direction $\mathbf{v}_1$, the unit eigenvector of $A^TA$ for $\lambda_1$. Similarly,

$$\sigma_n = \min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$$

is the minimum by which $A$ stretches a unit vector.

Proof. Let $\mathbf{x}$ be a unit vector, $\|\mathbf{x}\| = 1$. We have

$$\|A\mathbf{x}\|^2 = \langle A\mathbf{x}, A\mathbf{x}\rangle = \langle A^TA\mathbf{x}, \mathbf{x}\rangle = \sum_{i=1}^{n} \lambda_i \langle P_i\mathbf{x}, \mathbf{x}\rangle\,, \qquad (1)$$

# Singular Values and Stretching

where $A^TA = \sum_{i=1}^n \lambda_i P_i$ is the spectral expansion of $A^TA$. Now $\lambda_n \leq \lambda_i \leq \lambda_1$ for each $i$. So recalling that $\sum_{i=1}^n P_i = I_n$, we deduce from (1) that

$$\lambda_n = \lambda_n \sum_{i=1}^n \langle P_i\mathbf{x}, \mathbf{x}\rangle \leq \|A\mathbf{x}\|^2 \leq \lambda_1 \sum_{i=1}^n \langle P_i\mathbf{x}, \mathbf{x}\rangle = \lambda_1 \,.$$

As this holds for any unit vector $\mathbf{x}$, this implies

$$\lambda_n \leq \min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2 \qquad \text{and} \qquad \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2 \leq \lambda_1 \,.$$

On the other hand, $\lambda_i = \|A\mathbf{v}_i\|^2$ as we showed before. So

$$\lambda_n = \|A\mathbf{v}_n\|^2 \geq \min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2 \quad \text{and} \quad \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2 \geq \|A\mathbf{v}_1\|^2 = \lambda_1 \,.$$

Conclusion: $\lambda_n = \min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2$ and $\lambda_1 = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2$.

Taking the square-root, we deduce the theorem. $\square$

# Singular Values and Stretching

If $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$, then $\mathbf{x} \mapsto A\mathbf{x}$ is a map from $\mathbb{R}^2 \to \mathbb{R}^3$. Find a unit vector $\mathbf{x}$ at which $\|A\mathbf{x}\|$ is maximized, and compute this maximum length.

The intermediate singular values can be similarly characterized as stretching in appropriate subspaces. This is known as the *min-max theorem*.

Thm: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$, with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Suppose $A$ has $r$ nonzero singular values. Then $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$, and $\mathrm{rank}(A) = r$.

## $A^TA$ and $\mathrm{Col}(A)$

Proof. We have

$$\langle A\mathbf{v}_i, A\mathbf{v}_j\rangle = \langle A^TA\mathbf{v}_i, \mathbf{v}_j\rangle = \lambda_i\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0\,,$$

so $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ is an orthogonal set. Moreover, $\sigma_i = \|A\mathbf{v}_i\|$, so by hypothesis $A\mathbf{v}_i \neq \mathbf{0}$ for $i = 1, \ldots, r$ and $A\mathbf{v}_i = \mathbf{0}$ for $i > r$. So $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is linearly independent and is a subset of $\mathrm{Col}(A) = \mathrm{Range}(A)$.

Given $\mathbf{y} \in \mathrm{Col}(A)$, say $\mathbf{y} = A\mathbf{x}$, we have $\mathbf{x} = \sum_{i=1}^n c_i\mathbf{v}_i$ for some $c_i$, so $\mathbf{y} = A\mathbf{x} = A\sum_{i=1}^n c_i\mathbf{v}_i = \sum_{i=1}^n c_iA\mathbf{v}_i = \sum_{i=1}^r c_iA\mathbf{v}_i$. This shows that $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ also spans $\mathrm{Col}(A)$. It is thus an orthogonal basis, and we deduce that $\mathrm{rank}(A) = \dim \mathrm{Col}(A) = r$. □

# Singular Value Decomposition

In practice, the most reliable methods for computers to estimate the rank is to count the number of nonzero singular values (extremely small ones can be considered to be zero).

SVD Thm: For any $m \times n$ matrix $A$ of rank $r$, there exists an $m \times n$ matrix $\Sigma$ , an $m \times m$ orthogonal matrix $U$ and an $n \times n$ orthogonal matrix $V$ such that

(i) $A = U\Sigma V^T$.

(ii) $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$, where $D$ is an $r \times r$ diagonal matrix whose diagonal entries are the first $r$ singular values of $A$, $\sigma_1 \geq \cdots \geq \sigma_r > 0$.

# Singular Value Decomposition

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be the orthonormal basis of the previous theorem. We know that $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis of $\mathrm{Col}(A)$. We normalize it,

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

to obtain an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ of $\mathrm{Col}(A)$ satisfying

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i\,.$$

Now extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathbb{R}^m$ and define

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Then $U, V$ are orthogonal and

# Examples

$$AV = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

Now define $\Sigma$ and $D$ as in (ii). Then

$$\begin{aligned} U\Sigma &= \begin{bmatrix} U(\sigma_1\mathbf{e}_1) & \ldots & U(\sigma_r\mathbf{e}_r) & U\mathbf{0} & \cdots & U\mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = AV\,. \end{aligned}$$

Since $V$ is orthogonal, we deduce that $U\Sigma V^T = A$. □

Construct a singular value decomposition for

$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$.

We call the columns $\mathbf{u}_i$ of $U$ the *left singular vectors*, and the columns $\mathbf{v}_i$ of $V$ the *right singular vectors* of $A$.

## Bases for the column, row and null spaces

If $A$ is an $n \times n$ invertible matrix, we define its *condition number* to be the ration $\frac{\sigma_1}{\sigma_n}$. It is used to understand the sensitivity of a solution of $A\mathbf{x} = \mathbf{b}$ to small changes (errors) in the entries of $A$ in computer analysis.

Thm: Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be the right and left singular vectors of the $m \times n$ matrix $A$ of rank $r$. Then

(i) $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal basis of $\mathrm{Col}(A)$.

(ii) $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis of $\mathrm{Nul}(A^T)$.

(iii) $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathrm{Nul}(A)$.

(iv) $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis of $\mathrm{Row}(A)$.

# Bases for the column, row and null spaces

Proof. We proved (i) before. The vectors $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ all lie in $\mathrm{Col}(A)^\perp = \mathrm{Nul}(A^T)$. By the rank theorem, $\dim \mathrm{Nul}(A^T) = m - \mathrm{rank}(A^T) = m - r$, so they must form a basis. Here used that $\mathrm{rank}(A^T) = \dim \mathrm{Row}(A^T) = \dim \mathrm{Col}(A) = r$.

Next, since $A\mathbf{v}_i = \mathbf{0}$ for all $i > r$, the vectors $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ are all in $\mathrm{Nul}(A)$. By the rank theorem, $\dim \mathrm{Nul}(A) = n - \mathrm{rank}(A) = n - r$, so they form a basis.

Finally, the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ all lie in $\mathrm{Nul}(A)^\perp = \mathrm{Col}(A^T) = \mathrm{Row}(A)$, which has dimension $r$, so they form a basis. $\square$

# Invertibility criteria

Thm: Let $A$ be an $n \times n$ matrix. Then

$$A \text{ is invertible} \iff (\mathrm{Col}(A))^{\perp} = \mathbf{0}$$
$$\iff (\mathrm{Nul}(A))^{\perp} = \mathbb{R}^n \iff \mathrm{Row}(A) = \mathbb{R}^n \iff \sigma_i \neq 0 \ \forall i$$

Proof. By Lec 7, A in invertible $\iff \mathrm{Col}(A) = \mathbb{R}^n \iff \mathrm{Nul}(A) = \{\mathbf{0}\}$. Taking orthogonal complements, noting that $(\mathbb{R}^n)^{\perp} = \{\mathbf{0}\}$ and $\{\mathbf{0}\}^{\perp} = \mathbb{R}^n$, we get the first two equivalences. Also, $A$ is invertible iff $\mathrm{rank}(A) = n$, which occurs iff $\mathrm{Row}(A) = \mathbb{R}^n$ as we get a subspace of full dimension. Finally, $\sigma_i \neq 0 \ \forall i$ implies $\mathrm{Rank}(A) = n$ by the thm on p.8., and the converse holds by the SVD thm. $\square$

# Reduced SVD and pseudoinverses

Back to the SVD theorem, we may write $A = U\Sigma V^T$ more compactly in terms of $D$, avoiding the zero columns/rows. Namely,

$$A = U\Sigma V^T = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T\,,$$

where $U_r$ has size $m \times r$ and $V_r$ has size $n \times r$. We call $A = U_r D V_r^T$ the *reduced singular value decomposition* of $A$. Since the diagonal entries of $D$ are nonzero, $D$ is invertible. We call

$$A^+ = V_r D^{-1} U_r^T$$

the *pseudoinverse* or *Moore-Penrose inverse* of $A$.

# Reduced SVD and pseudoinverses

Note that if $A\mathbf{x} = \mathbf{b}$ and we define $\widehat{\mathbf{x}} = A^+\mathbf{b}$, we get

$$A\widehat{\mathbf{x}} = (U_r D V_r^T)(V_r D^{-1} U_r^T \mathbf{b}) = U_r U_r^T \mathbf{b} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) = \widehat{\mathbf{b}} \tag{2}$$

where we used that if $M$ has orthonormal columns then $M_r^T M_r = I$ and $M_r M_r^T = \mathrm{proj}_W$, the projection onto the space spanned by the columns of $M$. This is $\mathrm{Col}(A)$ for $U_r$ by the theorem on bases (i). Equation (2) shows that $\widehat{\mathbf{x}} = A^+\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

1. Given an SVD $A = U\Sigma V^T$, find an SVD for $A^T$. How are the singular values of $A$ and $A^T$ related ?

2. If $A$ is an $n \times n$ matrix, prove that $A^T A$ and $AA^T$ are orthogonally similar. That is, there exists an orthogonal $Q$ such that $AA^T = Q^T(A^T A)Q$.