# 1 Last time: symmetric matrices

A matrix $A$ is *symmetric* if $A^T = A$.

This happens if and only if $A$ is square and $A_{ij} = A_{ji}$ for all $i, j$.

**Example.** $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ is symmetric but $\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$ is not.

A matrix $U$ is *orthogonal* if $U$ is invertible and $U^{-1} = U^T$.

This happens precisely when $U$ is square with orthonormal columns.

An $n \times n$ matrix $A$ is *orthogonally diagonalisable* if there is an orthogonal matrix $U$ and a diagonal matrix $D$ such that $A = UDU^{-1} = UDU^T$. In this case, the columns of $U$ are an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for $A$, and the eigenvalues of these eigenvectors are the diagonal entries of $D$.

The following summarises the main results from last time:

**Theorem.**

(1) A square matrix is orthogonally diagonalisable if and only if it is symmetric.

(2) Eigenvectors with distinct eigenvalues of a symmetric matrix are orthogonal.

(3) All (complex) eigenvalues of a symmetric matrix $A$ are real, i.e., the characteristic polynomial of $A$ has all real roots and can be expressed as $\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$ for some not-necessarily-distinct real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$.

**Example.** Suppose $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$ for some $a, b \in \mathbb{R}$.

How does the preceding theorem apply to this generic 2-by-2 matrix? Since

$$\det(A - xI) = \det \begin{bmatrix} a - x & b \\ b & a - x \end{bmatrix} = (a - x)^2 - b^2 = (a - b - x)(a + b - x),$$

the eigenvalues of $A$ are $a - b$ and $a + b$.

It's not too hard to guess the eigenvectors corresponding to these eigenvectors just by looking, though the usual method of finding eigenvectors by row reducing $A - \lambda I$ to find a basis for $\text{Nul}(A - \lambda I)$ will also produce the answer.

Namely, the vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector for $A$ with eigenvalue $a - b$.

The vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector for $A$ with eigenvalue $a + b$.

These eigenvectors are orthogonal, as predicted by the theorem. We can convert them to unit vectors by multiplying each vector by the reciprocal of its length. This gives the eigenvectors

$$\begin{bmatrix} 1\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

which form an orthonormal basis for $\mathbb{R}^2$.

It follows that $A = UDU^{-1} = UDU^T$ where $U = \begin{bmatrix} 1\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ and $D = \begin{bmatrix} a - b & 0 \\ 0 & a + b \end{bmatrix}$.

## 2 Singular value decomposition

Today, we'll apply the results from last time to prove the existence of *singular value decompositions*, which will give a sort of approximate orthogonal diagonalisation for any matrix, not just symmetric ones.

Let $A$ be an $m \times n$ matrix.

Then $A^T A$ is a symmetric $n \times n$ matrix, since $(A^T A)^T = A^T (A^T)^T = A^T A$.

It follows from our results last time that $A^T A$ has all real eigenvalues. A stronger statement holds:

**Lemma.** All eigenvalues of $A^T A$ are nonnegative real numbers.

*Proof.* Since $A^T A$ is symmetric, we know that the matrix can be orthogonally diagonalised. In other words, we know there exists an orthonormal basis $v_1, v_2, \dots, v_n$ for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$. Let $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$ be the associated eigenvalues, so that $A^T A v_i = \lambda_i v_i$ for $i = 1, 2, \dots, n$. Then

$$\|Av_i\|^2 = (Av_i) \bullet (Av_i) = (Av_i)^T (Av_i) = v_i^T A^T A v_i = v_i^T (\lambda_i v_i) = \lambda_i \|v_i\|^2 = \lambda_i$$

for each index $i$. Since $\|Av_i\| \geq 0$, it follows that every eigenvalue satisfies $\lambda_i \geq 0$. □

The preceding lemma allows us to make the following definition.

**Definition.** Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of $A^T A$ arranged in decreasing order. Define $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \dots, n$. We call the numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ the *singular values* of $A$.

In other words, the singular values of a matrix $A$ are the squares roots of the eigenvalues of $A^T A$, which are guaranteed to be nonnegative real numbers (and therefore always have well-defined square roots).

**Example.** Suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$. Then $A^T A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$.

This matrix $A^T A$ has characteristic polynomial

$$\det(A^T A - xI) = (360 - x)(90 - x)x$$

so the eigenvalues of $A^T A$ are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$.

The singular values of $A$ are therefore $\sigma_1 = \sqrt{360}$, $\sigma_2 = \sqrt{90}$, and $\sigma_3 = 0$.

As a sequel to the lemma above, we have this nontrivial statement about the eigenvectors of $A^T A$.

**Theorem.** Suppose $v_1, v_2, \dots, v_n$ is an orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of $A^T A$, arranged so that if $\lambda_i \in \mathbb{R}$ is the eigenvalue of $v_i$ then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Assume $A$ has $r$ nonzero singular values.

Then $Av_1, Av_2, \dots, Av_r$ is an orthogonal basis for the column space of $A$ and consequently $\operatorname{rank} A = r$.

*Proof.* Choose indices $i \neq j$. Then $v_i \bullet v_j = 0$ so also $v_i \bullet \lambda_j v_j = 0$.

Therefore $(Av_i)^T (Av_j = v_i^T A^T A v_j = v_i^T (\lambda_j v_j) = v_i \bullet \lambda_j v_j = 0$.

This shows that $Av_1, Av_2, \dots, Av_r$ are orthogonal vectors in $\operatorname{Col} A$.

Since $\|Av_i\| = \sqrt{\lambda_i} > 0$, these vectors are all nonzero and therefore are linearly independent.

To see that these vectors span the column space of $A$, suppose $y \in \operatorname{Col} A$. The $y = Ax$ for some vector $x \in \mathbb{R}^n$, which we can write as $x = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ for some coefficients $c_1, c_2, \ldots, c_n \in \mathbb{R}$. Then

$$y = Ax = c_1Av_1 + c_2Av_2 + \cdots + c_rAv_r + \underbrace{c_{r+1}Av_{r+1} + \cdots + c_nAv_n}_{=0} = c_1Av_1 + c_2Av_2 + \cdots + c_rAv_r$$

since $Av_i = 0$ as $\|Av_i\| = \sqrt{\lambda_i} = 0$ for $i > r$. We conclude that $Av_1, Av_2, \ldots, Av_r$ is a basis for $\operatorname{Col} A$. □

**Corollary.** The rank of a matrix is the same as its number of nonzero singular values.

We arrive at today's main result.

**Theorem** (Existence of SVDs)**.** Let $A$ be an $m \times n$ matrix with rank $r$. Then we can write

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

where

(i) $U$ is some $m \times m$ orthogonal matrix.

(ii) $V$ is some $n \times n$ orthogonal matrix.

(iii) $\Sigma$ is the $m \times n$ matrix

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{where } D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the singular values of $A$.

**Comments.** The three zeros in the matrix defining $\Sigma$ are implicitly blocks of zeros: the upper right 0 stands for an $r \times (n-r)$ zero submatrix, the lower right 0 stands for an $(m-r) \times (n-r)$ zero submatrix, and the lower left 0 stands for an $(m-r) \times r$ zero submatrix.

Another way to think of $\Sigma$: place the diagonal matrix $D$ in the upper left corner of an $m \times n$ matrix, and then fill all of the remaining entries with zeros.

A factorisation $A = U\Sigma V^T$ with $U$, $V$, and $\Sigma$ as in (i)-(iii) is a *singular value decomposition* (*SVD*) of $A$.

The matrices $U$ and $V$ in an SVD $A = U\Sigma V^T$ are not uniquely determined by $A$, but $\Sigma$ is. The columns of $U$ are the *left singular vectors* of $A$ while the columns of $V$ are the *right singular vectors* of $A$.

*Proof that an SVD of A exists.* Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the decreasing list of eigenvalues of $A^TA$.

Let $v_1, v_2, \ldots, v_n$ a list of corresponding orthonormal eigenvectors for $A^TA$.

Then $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$ are $Av_1, Av_2, \ldots, Av_r$ is an orthogonal basis for $\operatorname{Col} A$.

For each $i = 1, 2, \ldots, r$, define

$$u_i = \frac{1}{\|Av_i\|}Av_i = \frac{1}{\sqrt{\lambda_i}}Av = \frac{1}{\sigma_i}Av_i.$$

Then $u_1, u_2, \ldots, u_r$ is an orthonormal basis for $\operatorname{Col} A$.

We can choose vectors $u_{r+1}, u_{r+2}, \ldots, u_m \in \mathbb{R}^m$ such that the extended list of vectors $u_1, u_2, \ldots, u_m$ is an orthonormal basis for $\mathbb{R}^m$. Make any such choice, and define

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \qquad \text{and} \qquad V = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}.$$

These matrices are orthogonal by construction, and

$$\begin{aligned} AV &= \begin{bmatrix} Av_1 & Av_2 & \dots Av_n \end{bmatrix} \\ &= \begin{bmatrix} Av_1 & Av_2 & \dots & Av_r & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}. \end{aligned}$$

If $\Sigma$ is the matrix given in (iii), then we also have

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix} = AV$$

so $U\Sigma V^T = AVV^T = AI = A$, which confirms the theorem statement. □

We conclude this lecture with a small example, continuing from before.

**Example.** Again suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

To find a singular value decomposition for $A$, there are three steps.

1. Find an orthogonal diagonalisation of $A^T A$.

   In this case $A^T A$ is a $3 \times 3$ matrix, and by the usual methods (of row reducing $A - \lambda I$ to find a basis for $\mathrm{Nul}(A - \lambda I)$ for each eigenvalue $\lambda$), you can find that

   $$v_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \qquad \text{and} \qquad v_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

   is an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of $A^T A$, with corresponding eigenvalues $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$.

2. Set up $V$ and $\Sigma$.

   Following the proof of the theorem, we have

   $$V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix} \qquad \text{and} \qquad D = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

   for $\sigma_1 = \sqrt{\lambda_1} = \sqrt{360}$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{90}$.

   Since $\Sigma$ must have the same size as $A$, we get

   $$\Sigma = \begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}.$$

3. Construct $U$.

   We have $U = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$ where $u_i = \frac{1}{\sigma_i} A v_i$.

   In this case you can compute that

   $$u_1 = \frac{1}{\sqrt{360}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} \qquad \text{and} \qquad u_2 = \frac{1}{\sqrt{90}}\begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

   which means that we can write

   $$U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & -1 \\ 1 & -3 \end{bmatrix}.$$

Putting everything together produces the singular value decomposition

$$A = U\Sigma V^T = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}\begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}.$$

Be careful to note that the third matrix factor is the transpose $V^T$ rather than $V$.