SPECTRAL THEOREM AND SVD

1. Preface: Induction

In the proof that follows we need mathematical induction. I won't bother to talk about this much, but I will state the axiom and explain—in my mind—what it means.

Axiom of Mathematical Induction. Suppose that for every positive integer n there is a corresponding statement $S(n)$. Then if

- $S(1)$ is true, and
- the truth of $S(k-1)$ implies the truth of $S(k)$

then for all positive integers the statement $S(n)$ is true.

Example. Show that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Proof. We'll prove this by induction. In order to do so, we need to show that the statement is true for $n = 1$, and also show that if the statement is true for $k-1$ it is also true for k. The first step is called the base case, and the second step is called the induction step.

Now the base case is simple: $\sum_{i=1}^{1} i = 1 = \frac{1(1+1)}{2}$.

For the induction step, assume that the statement holds for $k-1$: $\sum_{i=1}^{k-1} i = \frac{(k-1)(k-1+1)}{2}$. Then we have

$$\sum_{i=1}^{k} i = k + \sum_{i=1}^{k-1} i = k + \frac{(k-1)(k)}{2} = \frac{2k + k^2 - k}{2} = \frac{k^2 + k}{2} = \frac{k(k+1)}{2}.$$

□

2. The Spectral Theorem

Recently we have discussed what it means for an operator to be diagonalizable, and more generally have discussed the notion of similarity and change of basis. Today we restrict our attention to a special of similarity.

Definition 2.1. A matrix A is said to be orthogonally diagonalizable if there exists an orthogonal matrix S and diagonal matrix D so that

$$A = SDS^{-1} = SDS^{T}.$$

Notice that the last equality in our definition follows from the fact that S is orthogonal.

One of the questions we will be investigating today is: when is a matrix orthogonally diagonalizable? We know how to answer this question without the 'orthogonally' qualifier: a matrix is diagonalizable if, for instance, it has a basis of eigenvectors. The extra qualification, however, makes this question much more delicate to answer. We begin by noticing that being orthogonally diagonalizable puts some severe restrictions on a matrix.

Theorem 2.1. *A matrix which is orthogonally diagonalizable is symmetric.*

Proof. Suppose $A = SDS^{-1} = SDS^T$, where S is some orthogonal matrix. Then
$$A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T = A.$$
□

Are there further restrictions on being orthogonally diagonalizable, or is it enough for a matrix to be symmetric? Amazingly, symmetry is all you need! This is the content of the following

Theorem 2.2. *(Spectral Theorem) A matrix is orthogonally diagonalizable if and only if it is symmetric.*

Proving this result requires quite a bit of work. We'll proceed in several steps.

- For distinct eigenvalues λ_1, λ_2 of a symmetric matrix A, E_{λ_1} is orthogonal to E_{λ_2}.
- A symmetric matrix has n real eigenvalues (counted with multiplicity).
- The geometric multiplicity of each eigenvalue is equal to the algebraic multiplicity.

Theorem 2.3. *If λ_1, λ_2 are distinct eigenvalues for a symmetric matrix A, then E_{λ_1} is orthogonal to E_{λ_2}.*

Proof. Take $\overrightarrow{v_1}, \overrightarrow{v_2}$ in $E_{\lambda_1}, E_{\lambda_2}$, respectively. Then we'll compute $\overrightarrow{v_1} \cdot A\overrightarrow{v_2}$ in two different ways.
$$\overrightarrow{v_1} \cdot A\overrightarrow{v_2} = \lambda_2(\overrightarrow{v_1} \cdot \overrightarrow{v_2})$$
$$\overrightarrow{v_1} \cdot A\overrightarrow{v_2} = \overrightarrow{v_1}^T A\overrightarrow{v_2} = (A^T\overrightarrow{v_1})^T\overrightarrow{v_2} = (A\overrightarrow{v_1})^T\overrightarrow{v_2} = \lambda_1(\overrightarrow{v_1}^T\overrightarrow{v_2}) = \lambda_1(\overrightarrow{v_1} \cdot \overrightarrow{v_2})$$
But since $\lambda_1 \neq \lambda_2$ this means $\overrightarrow{v_1} \cdot \overrightarrow{v_2} = 0$. □

Theorem 2.4. *All eigenvalues of a symmetric matrix A are real.*

Proof. If we were allowed to use complex numbers then A would have n eigenvalues since char(A) would split into linear factors. So we just need to show that complex eigenvalues aren't too complex at all: they are actually real. So suppose that $p \pm iq$ are complex eigenvalues of A with corresponding eigenvectors $\overrightarrow{v} \pm i\overrightarrow{w}$. We'll compute a dot product in two ways (just like before).
$$\begin{aligned}(\overrightarrow{v} + i\overrightarrow{w})^T A(\overrightarrow{v} - i\overrightarrow{w}) &= (\overrightarrow{v} + i\overrightarrow{w})(p - iq)(\overrightarrow{v} - i\overrightarrow{w}) \\ &= (p - iq)(\overrightarrow{v} + i\overrightarrow{w})(\overrightarrow{v} - i\overrightarrow{w}) \\ &= (p - iq)(\|\overrightarrow{v}\|^2 + \|\overrightarrow{w}\|^2)\end{aligned}$$
and
$$\begin{aligned}(\overrightarrow{v} + i\overrightarrow{w})^T A(\overrightarrow{v} - i\overrightarrow{w}) &= (A^T(\overrightarrow{v} + i\overrightarrow{w}))^T(\overrightarrow{v} - i\overrightarrow{w}) \\ &= (A(\overrightarrow{v} + i\overrightarrow{w}))^T(\overrightarrow{v} - i\overrightarrow{w}) \\ &= (p + iq)(\overrightarrow{v} + i\overrightarrow{w})^T(\overrightarrow{v} + i\overrightarrow{w}) \\ &= (p + iq)(\|\overrightarrow{v}\|^2 + \|\overrightarrow{w}\|^2)\end{aligned}$$
But this means that $q = 0$, and hence our eigenvalues are actually real. □

Proof of Spectal Theorem. We'll prove the result by induction on the size of the matrix A. (So in the language from the preface, $S(n)$ is the statement 'The spectral theorem holds for $n \times n$ matrices.')

For the base case we need to show that a 1×1 matrix which is orthogonally diagonalizable is also symmetric. But all 1×1 matrices are symmetric, so we win.

Now for the induction step, assume that all $(k-1)\times(k-1)$ which are symmetric are orthogonally diagonalizable, and we'll show that all $k\times k$ matrices which are symmetric are orthogonally diagonalizable. So let A be a $k\times k$ matrix which is symmetric. Start by selecting an eigenvalue λ of A and a corresponding eigenvector $\vec{v_1}$ of length 1. Now complete $\vec{v_1}$ to an orthonormal basis of $\mathbb{R}^k$: $\vec{v_1},\cdots,\vec{v_k}$, and let P be the (orthogonal) matrix which has columns $\vec{v_1},\cdots,\vec{v_n}$.

Now notice two things about the matrix $P^{-1}AP = P^TAP$:

- The first column is $\lambda\vec{e_1}$
- It is symmetric, since
$$(P^TAP)^T = P^TA^T(P^T)^T = P^TAP.$$

This means that
$$P^TAP = \begin{pmatrix} \lambda & \\ & B \end{pmatrix},$$
where B is some symmetric $(k-1)\times(k-1)$ matrix. By induction, there exists Q which is $(k-1)\times(k-1)$ and orthogonal so that $B = QDQ^T$. Then notice that
$$\begin{pmatrix} 1 & \\ & Q \end{pmatrix}^{-1} P^{-1}AP \begin{pmatrix} 1 & \\ & Q \end{pmatrix} = \begin{pmatrix} 1 & \\ & Q^{-1} \end{pmatrix}\begin{pmatrix} \lambda & \\ & B \end{pmatrix}\begin{pmatrix} 1 & \\ & Q \end{pmatrix} = \begin{pmatrix} \lambda & \\ & D \end{pmatrix}.$$
□

Theorem 2.5. *You can orthogonally diagonalize a symmetric matrix A by*

- *finding eigenvalues and a basis for the corresponding eigenspaces;*
- *using Gram-Schmidt to compute an orthonormal basis for each eigenspace;*
- *concatenating these orthonormal bases.*

3. Singular Value Decomposition

The theory which described is important for a lot of reasons, but we'll focus on one particular application:

Theorem 3.1 (The Singular Value Decomposition)**.** *Every $n\times m$ matrix A has a decomposition*
$$A = U\Sigma V^T$$
where

- *V is an orthogonal $m\times m$ matrix;*
- *Σ is a diagonal $n\times m$ matrix; and*
- *U is an orthogonal $n\times n$ matrix.*

Definition 3.1. For an $n\times m$ matrix A, the singular values of A are the square roots of the eigenvalues of the symmetric matrix A^TA. The singular values are usually denoted $\sigma_1,\cdots,\sigma_m$ and listed in decreasing order.

Theorem 3.2. *Let A be an $n\times m$ matrix. If $\vec{v_1},\cdots,\vec{v_m}$ is an orthonormal basis which diagonalizes A^TA (and $\vec{v_i}$ corresponds to eigenvalues $\lambda_i = \sigma_i^2$), then*

- *the vectors $A\vec{v_1},\cdots,A\vec{v_m}$ are orthogonal and*
- *$\|A\vec{v_i}\| = \sigma_i\vec{v_i}$.*

Proof. For the first,

$$A\overrightarrow{v_i} \cdot A\overrightarrow{v_j} = (A\overrightarrow{v_i})^T A\overrightarrow{v_j} = \overrightarrow{v_i}^T A^T A\overrightarrow{v_j} = \lambda_j \overrightarrow{v_i}^T \overrightarrow{v_j} = \lambda_j(\overrightarrow{v_i} \cdot \overrightarrow{v_j}).$$

For the second,

$$\|A\overrightarrow{v_i}\|^2 = (A\overrightarrow{v_i} \cdot A\overrightarrow{v_i}) = (A\overrightarrow{v_i})^T A\overrightarrow{v_i} = \overrightarrow{v_i}^T A^T A\overrightarrow{v_i} = \lambda_i \overrightarrow{v_i}^T \overrightarrow{v_i} = \lambda_i = \sigma_i^2.$$

□

Theorem 3.3. *If $r = rank(A)$, then $\sigma_{r+1}, \cdots, \sigma_m$ are zero.*

Proof. Let s be the largest number so that $\sigma_s \neq 0$. Since $\overrightarrow{v_1}, \cdots, \overrightarrow{v_m}$ are orthonormal, they form a basis for $\mathbb{R}^m$ and hence the image of A. But notice that for an arbitrary $\overrightarrow{x} \in \mathbb{R}^m$ we have

$$A\overrightarrow{x} = A(c_1\overrightarrow{v_1} + \cdots + c_s\overrightarrow{v_s} + \cdots + c_m\overrightarrow{v_m}) = c_1 A\overrightarrow{v_1} + \cdots + c_s A\overrightarrow{v_s}$$

since $\|A\overrightarrow{v_i}\| = \sigma_i$ (note that therefore if $i > s$ then $\|A\overrightarrow{v_i}\| = 0$ so that $A\overrightarrow{v_i} = \overrightarrow{0}$. But since $A\overrightarrow{v_1}, \cdots, A\overrightarrow{v_s}$ are nonzero and mutually orthogonal they are linearly independent and hence form a basis for the image of A. But im(A) has dimension rank(A), and so $s = \text{rank}(A)$. □

Proof of the Singular Value Decomposition. For $i \leq \text{rank}(A)$, let $\overrightarrow{u_i} = \frac{1}{\sigma_i} A\overrightarrow{v_i}$ for $i = 1, \cdots, r = \text{rank}(A)$, and let $\overrightarrow{u_{r+1}}, \cdots, \overrightarrow{u_n}$ be a collection of vectors that completes $\overrightarrow{u_1}, \cdots, \overrightarrow{u_r}$ as an orthonormal basis of $\mathbb{R}^n$. Then we have

$$A \underbrace{\left(\overrightarrow{v_1} \middle| \cdots \middle| \overrightarrow{v_r} \middle| \overrightarrow{v_{r+1}} \middle| \cdots \middle| \overrightarrow{v_m} \right)}_{V} = \left(\sigma_1\overrightarrow{u_1} \middle| \cdots \middle| \sigma_r\overrightarrow{u_r} \middle| \overrightarrow{0} \middle| \cdots \middle| \overrightarrow{0} \right)$$

$$= \left(\overrightarrow{u_1} \middle| \cdots \middle| \overrightarrow{u_r} \middle| \overrightarrow{0} \middle| \cdots \middle| \overrightarrow{0} \right) \begin{pmatrix} \sigma_1 & & \\ & \ddots & 0 \\ & & \sigma_r \\ 0 & & 0 \end{pmatrix}$$

$$= \underbrace{\left(\overrightarrow{u_1} \middle| \cdots \middle| \overrightarrow{u_r} \middle| \overrightarrow{u_{r+1}} \middle| \cdots \middle| \overrightarrow{u_n} \right)}_{U} \underbrace{\begin{pmatrix} \sigma_1 & & \\ & \ddots & 0 \\ & & \sigma_r \\ 0 & & 0 \end{pmatrix}}_{\Sigma}$$

Theorem 3.4. *To find an singular value decomposition of a matrix A,*

- *Find eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m$ of the symmetric $m \times m$ matrix $A^T A$ and a corresponding orthonormal colletion of eigenvalues $\overrightarrow{v_1}, \cdots, \overrightarrow{v_m}$;*
- *The matrix V is the matrix whose columns are $\overrightarrow{v_1}, \cdots, \overrightarrow{v_m}$;*
- *The matrix Σ is diagonal with entries the singular values $\sigma_i = \sqrt{\lambda_i}$*
- *The matrix U is the matrix whose first r columns are $\frac{1}{\sigma_i} A\overrightarrow{v_i}$ and whose final $n - r$ columns complete an orthonormal basis for $\mathbb{R}^n$.*

□