The Spectral Theorem

17.1. A real or complex matrix A is called **symmetric** or **self-adjoint** if $A^* = A$, where $A^* = \overline{A}^T$. For a real matrix A, this is equivalent to $A^T = A$. A real or complex matrix is called **normal** if $A^*A = AA^*$. Examples of normal matrices are symmetric or anti-symmetric matrices. Normal matrices appear often in applications. Correlation matrices in statistics or operators belonging to observables in quantum mechanics, adjacency matrices of networks are all self-adjoint. Orthogonal and unitary matrices are all normal.

17.2.

> **Theorem:** Symmetric matrices have only real eigenvalues.

Proof. We extend the dot product to complex vectors as $(v, w) = v \cdot w = \sum_i \overline{v}_i w_i$ which extends the usual dot product $(v, w) = \overline{v} \cdot w$ for real vectors. This dot product has the property $(A^*v, w) = (v, Aw)$ and $(\lambda v, w) = \overline{\lambda}(v, w)$ as well as $(v, \lambda w) = \lambda(v, w)$. Now $\overline{\lambda}(v, v) = (\lambda v, v) = (Av, v) = (A^*v, v) = (v, Av) = (v, \lambda v) = \lambda(v, v)$ shows that $\overline{\lambda} = \lambda$ because $(v, v) = \overline{v} \cdot v = |v_1|^2 + \cdots + |v_n|^2$ is non-zero for non-zero vectors v. □

17.3.

> **Theorem:** If A is symmetric, then eigenvectors to different eigenvalues are perpendicular.

Proof. Assume $Av = \lambda v$ and $Aw = \mu w$. If $\lambda \neq \mu$, then the relation $\lambda(v, w) = (\lambda v, w) = (Av, w) = (v, A^T w) = (v, Aw) = (v, \mu w) = \mu(v, w)$ is only possible if $(v, w) = 0$. □

17.4. If A is a $n \times n$ matrix for which all eigenvalues are different, we say such a matrix has **simple spectrum**. The “wiggle-theorem” tells that we can approximate a given matrix with matrices having simple spectrum:

> **Theorem:** A symmetric matrix can be approximated by symmetric matrices with simple spectrum.

Proof. We show that there exists a curve $A(t) = A(t)^T$ of symmetric matrices with $A(0) = A$ such that $A(t)$ has simple for small positive t.
Use induction with respect to n. For $n = 1$, this is clear. Assume it is true for n, let A be a $(n+1) \times (n+1)$ matrix. It has an eigenvalue λ_1 with eigenvector v_1 which we assume to have length 1. The still symmetric matrix $A + tv_1 \cdot v_1^T$ has the same eigenvector v_1 with eigenvalue $\lambda_1 + t$. Let $v_2, \dots, v_n$ be an orthonormal basis of $V^\perp$ the space perpendicular to $V = \text{span}(v_1)$. Then $A(t)v = Av$ for any v in $V^\perp$. In that basis, the matrix $A(t)$ becomes $B(t) = \begin{bmatrix} \lambda_1 + t & C \\ 0 & D \end{bmatrix}$. Let S be the orthogonal matrix which contains the orthonormal basis $\{v_1, v_2, \dots, v_n\}$ of $\mathbb{R}^n$. Because $B(t) = S^{-1}A(t)S$ with orthogonal S, also $B(t)$ is symmetric implying that $C = 0$. So, $B(t)$ preserves D and $B(t)$ restricted to D does not depend on t. In particular, all the eigenvalues are different from $\lambda_1 + t$. By induction we find a curve $D(t)$ with $D(0) = D$ such that all the eigenvalues of $D(t)$ are different and also different from $\lambda_1 + t$. ☐

17.5. This immediately implies the **spectral theorem**

Theorem: Every symmetric matrix A has an orthonormal eigenbasis.

Proof. Wiggle A so that all eigenvalues of $A(t)$ are different. There is now an orthonormal basis $\mathcal{B}(t)$ for $A(t)$ leading to an orthogonal matrix $S(t)$ such that $S(t)^{-1}A(t)S(t) = B(t)$ is diagonal for every small positive t. Now, the limit $S(t) = \lim_{t \to 0} S(t)$ and also the limit $S^{-1}(t) = S^T(t)$ exists and is orthogonal. This gives a diagonalization $S^{-1}AS = B$. The ability to diagonalize is equivalent to finding an eigenbasis. As S is orthogonal, the eigenbasis is orthonormal. ☐

17.6. What goes wrong if A is not symmetric? Why can we not wiggle then? The proof applied to the magic matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ gives $A(t) = A + te_1 \cdot e_1^T = \begin{bmatrix} t & 1 \\ 0 & 0 \end{bmatrix}$ which has the eigenvalues $0, t$. For every $t > 0$, there is an eigenbasis with eigenvectors $[1, 0]^T, [1, -t]$. We see that for $t \to 0$, these two vectors collapse. This can not happen in the symmetric case because eigenvectors to different eigenvalues are orthogonal there. We see also that the matrix $S(t)$ converges to a singular matrix in the limit $t \to 0$.

17.7. First note that if A is normal, then A has the same eigenspaces as the symmetric matrix $A^*A = AA^*$: if $A^*Av = \lambda v$, then $(A^*A)Av = AA^*Av = A\lambda v = \lambda Av$, so that also Av is an eigenvector of A^*A. This implies that if A^*A has simple spectrum, (leading to an orthonormal eigenbasis as it is symmetric), than A also has an orthonormal eigenbasis, namely the same one. The following result follows from a Wiggling theorem for normal matrices:

17.8.

Theorem: Any normal matrix can be diagonalized using a unitary S.

Examples

17.9. A matrix A is called doubly stochastic if the sum of each row is 1 and the sum of each column is 1. Doubly stochastic matrices in general are not normal, but they are in the case $n = 2$. Find its eigenvalues and eigenvectors. The matrix must have the form

$$A = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

It is symmetric and therefore normal. Since the rows sum up to 1, the eigenvalue 1 appears to the eigenvector $[1,1]^T$. The trace is $2a$ so that the second eigenvalue is $2a - 1$. Since the matrix is symmetric and for $a \neq 0$ the two eigenvalues are distinct, by the theorem, the two eigenvectors are perpendicular. The second eigenvector is therefore $[-1,1]^T$.

17.10. We have seen the quaternion matrix belonging to $z = p + iq + jr + ks$: $\begin{bmatrix} p & -q & -r & -s \\ q & p & s & -r \\ r & -s & p & q \\ s & r & -q & p \end{bmatrix}$. As an orthogonal matrix, it is normal. Let $v = [q,r,s]$ be the space vector defined by the quatenion. Then the eigenvalues of A are $p \pm i|v|$, both with algebraic multiplicity 2. The characteristic polynomial is $p_A(\lambda) = (\lambda^2 - 2p\lambda + |z|^2)^2$.

17.11. Every normal 2×2 matrix is either symmetric or a rotation-dilation matrix. Proof: just write down $AA^T = A^T A$. This gives a system of quadratic equations for four variables a, b, c, d. This gives $c = b$ or $c = -b, d = a$.