SYMMETRIC MATRICES — Math 21b, O. Knill

SYMMETRIC MATRICES. A matrix A with real entries is **symmetric**, if $A^T = A$.

EXAMPLES. $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ is symmetric, $A = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix}$ is not symmetric.

EIGENVALUES OF SYMMETRIC MATRICES. Symmetric matrices A have real eigenvalues.

PROOF. The dot product is extend to complex vectors as $(v, w) = \sum_i \overline{v}_i w_i$. For real vectors it satisfies $(v, w) = v \cdot w$ and has the property $(Av, w) = (v, A^T w)$ for real matrices A and $(\lambda v, w) = \overline{\lambda}(v, w)$ as well as $(v, \lambda w) = \lambda(v, w)$. Now $\overline{\lambda}(v, v) = (\lambda v, v) = (Av, v) = (v, A^T v) = (v, Av) = (v, \lambda v) = \lambda(v, v)$ shows that $\overline{\lambda} = \lambda$ because $(v, v) \neq 0$ for $v \neq 0$.

EXAMPLE. $A = \begin{bmatrix} p & -q \\ q & p \end{bmatrix}$ has eigenvalues $p + iq$ which are real if and only if $q = 0$.

EIGENVECTORS OF SYMMETRIC MATRICES. Symmetric matrices have an orthonormal eigenbasis

PROOF. If $Av = \lambda v$ and $Aw = \mu w$. 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 $(v, w) = 0$ if $\lambda \neq \mu$.

WHY ARE SYMMETRIC MATRICES IMPORTANT? In applications, matrices are often symmetric. For example in **geometry** as **generalized dot products** $v \cdot Av$, or in **statistics** as **correlation matrices** $\text{Cov}[X_k, X_l]$ or in quantum mechanics as **observables** or in **neural networks** as **learning maps** $x \mapsto \text{sign}(Wx)$ or in graph theory as **adjacency matrices** etc. etc. Symmetric matrices play the same role as real numbers do among the complex numbers. Their eigenvalues often have physical or geometrical interpretations. One can also calculate with symmetric matrices like with numbers: for example, we can solve $B^2 = A$ for B if A is symmetric matrix and B is square root of A.) This is not possible in general: try to find a matrix B such that $B^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$...

RECALL. We have seen when an eigenbasis exists, a matrix A can be transformed to a diagonal matrix $B = S^{-1}AS$, where $S = [v_1, ..., v_n]$. The matrices A and B are **similar**. B is called the **diagonalization** of A. Similar matrices have the same characteristic polynomial $\det(B - \lambda) = \det(S^{-1}(A - \lambda)S) = \det(A - \lambda)$ and have therefore the same determinant, trace and eigenvalues. Physicists call the set of eigenvalues also **the spectrum**. They say that these matrices are isospectral. The spectrum is what you "see" (etymologically the name origins from the fact that in quantum mechanics the spectrum of radiation can be associated with eigenvalues of matrices.)

SPECTRAL THEOREM. Symmetric matrices A can be diagonalized $B = S^{-1}AS$ with an orthogonal S.

PROOF. If all eigenvalues are different, there is an eigenbasis and diagonalization is possible. The eigenvectors are all orthogonal and $B = S^{-1}AS$ is diagonal containing the eigenvalues. In general, we can change the matrix A to $A = A + (C - A)t$ where C is a matrix with pairwise different eigenvalues. Then the eigenvalues are different for all except finitely many t. The orthogonal matrices S_t converges for $t \to 0$ to an orthogonal matrix S and S diagonalizes A.

WAIT A SECOND ... Why could we not perturb a general matrix A_t to have disjoint eigenvalues and A_t could be diagonalized: $S_t^{-1} A_t S_t = B_t$? The problem is that S_t might become singular for $t \to 0$. See problem 5) first practice exam.

EXAMPLE 1. The matrix $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$ has the eigenvalues $a + b, a - b$ and the eigenvectors $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} / \sqrt{2}$ and $v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} / \sqrt{2}$. They are orthogonal. The orthogonal matrix $S = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ diagonalized A.

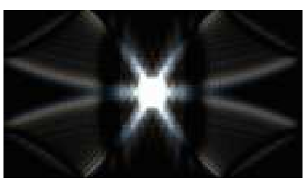
EXAMPLE 2. The 3×3 matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ has 2 eigenvalues 0 to the eigenvectors $\begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$ and one eigenvalue 3 to the eigenvector $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$. All these vectors can be made orthogonal and a diagonalization is possible even so the eigenvalues have multiplicities.

SQUARE ROOT OF A MATRIX. How do we find a square root of a given symmetric matrix? Because $S^{-1}AS = B$ is diagonal and we know how to take a square root of the diagonal matrix B, we can form $C = S\sqrt{B}S^{-1}$ which satisfies $C^2 = S\sqrt{B}S^{-1}S\sqrt{B}S^{-1} = SBS^{-1} = A$.

RAYLEIGH FORMULA. We write also $(\vec{v}, \vec{w}) = \vec{v} \cdot \vec{w}$. If $\vec{v}(t)$ is an eigenvector of length 1 to the eigenvalue $\lambda(t)$ of a symmetric matrix $A(t)$ which depends on t, differentiation of $(A(t) - \lambda(t))\vec{v}(t) = 0$ with respect to t gives $(A' - \lambda')v + (A - \lambda)v' = 0$. The symmetry of $A - \lambda$ implies $0 = (v, (A' - \lambda')v) + (v, (A - \lambda)v') = (v, (A' - \lambda')v)$. We see that the **Rayleigh quotient** $\boxed{\lambda' = (A'v, v)}$ is a polynomial in t if $A(t)$ only involves terms $t, t^2, \dots, t^m$. The formula shows how $\lambda(t)$ changes, when t varies. For example, $A(t) = \begin{bmatrix} 1 & t^2 \\ t^2 & 1 \end{bmatrix}$ has for $t = 2$ the eigenvector $\vec{v} = [1, 1]/\sqrt{2}$ to the eigenvalue $\lambda = 5$. The formula tells that $\lambda'(2) = (A'(2)\vec{v}, \vec{v}) = (\begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix} \vec{v}, \vec{v}) = 4$. Indeed, $\lambda(t) = 1 + t^2$ has at $t = 2$ the derivative $2t = 4$.

EXHIBITION. "Where do symmetric matrices occur?" Some informal motivation:

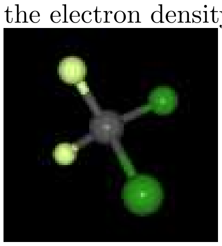
I) PHYSICS: In **quantum mechanics** a system is described with a vector $v(t)$ which depends on time t. The evolution is given by the **Schroedinger equation** $\dot{v} = i\hbar Lv$, where L is a symmetric matrix and $\hbar$ is a small number called the Planck constant. As for any linear differential equation, one has $v(t) = e^{i\hbar Lt}v(0)$. If $v(0)$ is an eigenvector to the eigenvalue λ, then $v(t) = e^{it\hbar\lambda}v(0)$. Physical observables are given by symmetric matrices too. The matrix L represents the energy. Given $v(t)$, the value of the observable $A(t)$ is $v(t) \cdot Av(t)$. For example, if v is an eigenvector to an eigenvalue λ of the energy matrix L, then the energy of $v(t)$ is λ.



This is called the **Heisenberg picture**. In order that $v \cdot A(t)v = v(t) \cdot Av(t) = S(t)v \cdot AS(t)v$ we have $A(t) = S(T)^*AS(t)$, where $S^* = \overline{S^T}$ is the correct generalization of the adjoint to complex matrices. $S(t)$ satisfies $S(t)^*S(t) = 1$ which is called **unitary** and the complex analogue of orthogonal. The matrix $A(t) = S(t)^*AS(t)$ has the same eigenvalues as A and is **similar** to A.

II) CHEMISTRY. The **adjacency matrix** A of a graph with n vertices determines the graph: one has $A_{ij} = 1$ if the two vertices i, j are connected and zero otherwise. The matrix A is symmetric. The eigenvalues λ_j are real and can be used to analyze the graph. One interesting question is to what extent the eigenvalues determine the graph.

In chemistry, one is interested in such problems because it allows to make rough computations of the electron density distribution of molecules. In this so called **Hückel theory**, the molecule is represented as a graph. The eigenvalues λ_j of that graph approximate the energies an electron on the molecule. The eigenvectors describe the electron density distribution.



The **Freon molecule** for example has 5 atoms. The adjacency matrix is $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$.

This matrix A has the eigenvalue 0 with multiplicity 3 ($\ker(A)$ is obtained immediately from the fact that 4 rows are the same) and the eigenvalues $2, -2$. The eigenvector to the eigenvalue ± 2 is $\begin{bmatrix} \pm 2 & 1 & 1 & 1 & 1 \end{bmatrix}^T$.

III) STATISTICS. If we have a random vector $X = [X_1, \cdots, X_n]$ and $E[X_k]$ denotes the expected value of X_k, then $[A]_{kl} = E[(X_k - E[X_k])(X_l - E[X_l])] = E[X_k X_l] - E[X_k]E[X_l]$ is called the **covariance matrix** of the random vector X. It is a symmetric $n \times n$ matrix. Diagonalizing this matrix $B = S^{-1}AS$ produces new random variables which are **uncorrelated**.

For example, if X is is the sum of two dice and Y is the value of the second dice then $E[X] = [(1+1) + (1+2) + ... + (6+6)]/36 = 7$, you throw in average a sum of 7 and $E[Y] = (1 + 2 + ... + 6)/6 = 7/2$. The matrix entry $A_{11} = E[X^2] - E[X]^2 = [(1+1) + (1+2) + ... + (6+6)]/36 - 7^2 = 35/6$ known as the **variance** of X, and $A_{22} = E[Y^2] - E[Y]^2 = (1^2 + 2^2 + ... + 6^2)/6 - (7/2)^2 = 35/12$ known as the **variance** of Y and $A_{12} = E[XY] - E[X]E[Y] = 35/12$. The covariance matrix is the symmetric matrix $A = \begin{bmatrix} 35/6 & 35/12 \\ 35/12 & 35/12 \end{bmatrix}$.