MA 511, Session 28

Diagonalization

Let A be a $n \times n$ matrix. Suppose that A has n linearly independent eigenvectors, $v_1, \ldots, v_n$, corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively.

Let us define the eigenvalue matrix Λ and an eigenvector matrix S as follows:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ & \cdot & & \\ & \cdot & & \\ & \cdot & & \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}, \quad S = \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix}.$$

The eigenvalues may be repeated, and may be real or complex. Also, since the columns of S are assumed independent, S is invertible and we have

Theorem: $A = S\Lambda S^{-1}$, that is, if A has n linearly independent eigenvectors, then A is diagonalizable.

Proof: Since $Av_j = \lambda_j v_j$, $1 \leq j \leq n$, it follows that $AS = S\Lambda$.

We say that A has been <u>diagonalized</u> by S. Not all matrices can be diagonalized. The matrix A from last session is (trivially) diagonalizable since it is already diagonal, but B and C are not diagonalizable.

To better understand this concept, let us introduce the concepts of <u>algebraic multiplicity</u> and <u>geometric multiplicity</u> of an eigenvalue λ of A. The former is the multiplicity of λ as root of the characteristic polynomial $p(\lambda)$; the latter is the dimension of the eigenspace S_λ, which is always less than or equal to the former.

Remark: By definition, when λ is an eigenvalue of A, a nonzero solution of $Ax = \lambda x$ exists and, therefore, $\dim S_\lambda$ is at least 1. As a consequence, it follows immediately that, when all n eigenvalues of A are distinct, each of the corresponding eigenspaces have dimension exactly equal to 1.

Example: Consider the following three matrices:

$$\tilde{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \tilde{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

$$\tilde{C} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

All three share the same characteristic polynomial $p(\lambda) = (1-\lambda)^3(2-\lambda)$, and the same eigenvalues, $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\lambda_4 = 2$. This means that the algebraic multiplicity of $\lambda = 1$ is 3 in all cases, and that of $\lambda = 2$ is 1 in all cases. Concerning the eigenspaces, S_2 is the line spanned by $(0,0,0,1)^T$ in all cases. However, S_1 is quite different in the three cases: it is a 3-dimensional subspace of $\mathbb{R}^4$ for $\tilde{A}$, spanned by $(1,0,0,0)^T$, $(0,1,0,0)^T$ and $(0,0,1,0)^T$, it is a 2-dimensional subspace of $\mathbb{R}^4$ (a plane) for $\tilde{B}$, spanned by $(1,0,0,0)^T$ and $(0,0,1,0)^T$, and it is a 1-dimensional subspace of $\mathbb{R}^4$ (a line) for $\tilde{C}$, spanned by $(1,0,0,0)^T$.

Theorem: Let $v_1, \ldots, v_k$ be eigenvectors of A corresponding to different eigenvalues $\lambda_1, \ldots, \lambda_k$, respectively. Then, $v_1, \ldots, v_k$ are linearly independent.

<u>Proof</u>: By contradiction, assume there is a nontrivial linear combination

$$c_1 v_1 + \cdots + c_k v_k = 0. \tag{*}$$

Since at least one of the coefficients must be different from zero, we may assume $c_1 \neq 0$ (otherwise, change the order of the vectors and renumber them). Now

$$\begin{aligned} 0 = A0 &= A(c_1 v_1 + \cdots + c_k v_k) \\ &= c_1 A(v_1) + \cdots + c_k A(v_k) \\ &= \lambda_1 c_1 v_1 + \cdots + \lambda_k c_k v_k. \end{aligned} \tag{**}$$

Multiplying (*) by λ_k we obtain the relation

$$\lambda_k c_1 v_1 + \cdots + \lambda_k c_k v_k = 0,$$

which, subtracted from (**) leads to

$$(\lambda_1 - \lambda_k) c_1 v_1 + \cdots + (\lambda_{k-1} - \lambda_k) c_{k-1} v_{k-1} = 0.$$

Repeating this process $k-1$ times we arrive at the relation

$$(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1}) \ldots (\lambda_1 - \lambda_2) c_1 v_1 = 0,$$

which implies $c_1 = 0$, since all the other factors are different from zero by assumption. This is a contradiction and thus, there cannot exist a nontrivial linear combination of $v_1, \ldots, v_k$ that produces the zero vector, i.e. they are linearly independent.

The process of diagonalization is very useful in making changes of variables. When we do this later, the following observation will be very important.

Remark: Let x be an eigenvector of A corresponding to an eigenvalue λ. Then,

$$A^k x = A^{k-1} A x = \lambda A^{k-1} x = \lambda^2 A^{k-2} x = \cdots = \lambda^k x,$$

that is, x is also an eigenvector of A^k corresponding to the eigenvalue λ^k. Moreover, if S diagonalizes A (i.e. $S^{-1} A S = \Lambda$), then

$$\Lambda^k = (S^{-1} A S)(S^{-1} A S) \ldots (S^{-1} A S) = S^{-1} A^k S,$$

that is, S diagonalizes A^k too.