MATH 423

Linear Algebra II

Lecture 24:
Multiple eigenvalues.
Invariant subspaces.
Markov chains.

Example. $A = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$.

Characteristic polynomial:

$$\det(A - \lambda I) = (\lambda^2 + 1)^2 = (\lambda - i)^2(\lambda + i)^2.$$

The eigenvalues are i and $-i$. Both eigenspaces of A are one-dimensional. The eigenspace for i is spanned by $(0, 0, i, 1)$ and the eigenspace for $-i$ spanned by $(0, 0, -i, 1)$.

It follows that the matrix A is not diagonalizable in $\mathbb{C}^4$.

There is also an indirect way to show that A is not diagonalizable in $\mathbb{C}^4$. Assume the contrary. Then $A = UXU^{-1}$, where U is an invertible matrix with complex entries and

$$X = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}$$

(note that X should have the same characteristic polynomial as A). This would imply that $A^2 = UX^2U^{-1}$. But $X^2 = -I$ so that $A^2 = U(-I)U^{-1} = -I$.

Let us check if $A^2 = -I$.

$$A^2 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -2 & -1 & 0 \\ 2 & 0 & 0 & -1 \end{pmatrix}.$$

Since $A^2 \neq -I$, the matrix A is not diagonalizable in $\mathbb{C}^4$.

Remark. Note however that $(A^2 + I)^2 = O$ (this is an instance of the Cayley-Hamilton Theorem).

Multiple eigenvalues

Definition. Suppose λ is an eigenvalue of a matrix A. The **multiplicity** (or **algebraic multiplicity**) of this eigenvalue is its multiplicity as a root of the characteristic polynomial of A. The **geometric multiplicity** of λ is the dimension of the associated eigenspace.

Theorem 1 Geometric multiplicity of an eigenvalue cannot exceed its algebraic multiplicity.

Theorem 2 A square matrix is diagonalizable if and only if the following conditions are satisfied:

- the characteristic polynomial splits into factors of degree 1;
- the geometric multiplicity of each eigenvalue matches its algebraic multiplicity.

Invariant subspaces

Let $L : V \to V$ be a linear operator on a vector space V. Suppose W is a subspace of V. We say that the subspace W is **invariant** under the operator L (or that W is an **invariant subspace** of L) if $L(W) \subset W$.

If W is an invariant subspace of L, then the **restriction** of the operator L to W, denoted $L|_W$, can be regarded as an operator on W.

Example. Consider the vector space $\mathcal{P}$ of all polynomials, its subspace $\mathcal{P}_n$ (polynomials of degree at most n), and three operators L_1, L_2, L_3 on $\mathcal{P}$ given by

- $(L_1 p)(x) = p'(x)$,
- $(L_2 p)(x) = p(x+1)$,
- $(L_3 p)(x) = xp(x)$

for any polynomial $p \in \mathcal{P}$. Then the subspace $\mathcal{P}_n$ is invariant under operators L_1 and L_2, but not invariant under L_3.

Suppose $L : V \to V$ is a linear operator on an n-dimensional vector space V and W is an m-dimensional subspace of V that is invariant under L.

Let $\beta = [\mathbf{v}_1, \dots, \mathbf{v}_m]$ be a basis for W and $\mathbf{v}_{m+1}, \dots, \mathbf{v}_n$ be vectors that extend this basis to a basis for V (denoted α).

Theorem 1 The matrix A of the operator L relative to the basis α is a block matrix of the form

$$A = \begin{pmatrix} B & C \\ O & D \end{pmatrix},$$

where O is the $(n-m)\times m$ zero matrix and B is the matrix of the restriction $L|_W$ relative to the basis β.

Theorem 2 Using notation of the previous theorem, $\det(A) = \det(B)\,\det(D)$. Moreover, $\det(A - \lambda I_n) = \det(B - \lambda I_m)\,\det(D - \lambda I_{n-m})$ for any scalar λ.

Corollary The characteristic polynomial of the restriction $L|_W$ divides the characteristic polynomial of L.

Stochastic process

Stochastic (or **random**) **process** is a sequence of experiments for which the outcome at any stage depends on a chance.

Simple model:

- a finite number of possible outcomes (called **states**);
- discrete time

Let S denote the set of the states. Then the stochastic process is a sequence $s_0, s_1, s_2, \ldots,$ where all $s_n \in S$ depend on chance.

How do they depend on chance?

Bernoulli scheme

Bernoulli scheme is a sequence of independent random events.

That is, in the sequence $s_0, s_1, s_2, \ldots$ any outcome s_n is independent of the others.

For any integer $n \geq 0$ we have a probability distribution $p^{(n)}$ on S. This means that each state $s \in S$ is assigned a value $p_s^{(n)} \geq 0$ so that $\sum_{s \in S} p_s^{(n)} = 1$. Then the probability of the event $s_n = s$ is $p_s^{(n)}$.

The Bernoulli scheme is called **stationary** if the probability distributions $p^{(n)}$ do not depend on n.

Examples of Bernoulli schemes:

- Coin tossing

2 states: heads and tails. Equal probabilities: $1/2$.

- Die rolling

6 states. Uniform probability distribution: $1/6$ each.

- Lotto Texas

Any state is a 6-element subset of the set $\{1, 2, \ldots, 54\}$. The total number of states is $25,827,165$. Uniform probability distribution.

Markov chain

Markov chain is a stochastic process with discrete time such that the probability of the next outcome depends only on the previous outcome.

Let $S = \{1, 2, \dots, k\}$. The Markov chain is determined by **transition probabilities** $p_{ij}^{(t)}$, $1 \leq i, j \leq k$, $t \geq 0$, and by the **initial** probability distribution q_i, $1 \leq i \leq k$.

Here q_i is the probability of the event $s_0 = i$, and $p_{ij}^{(t)}$ is the conditional probability of the event $s_{t+1} = j$ provided that $s_t = i$. By construction, $p_{ij}^{(t)}, q_i \geq 0$, $\sum_i q_i = 1$, and $\sum_j p_{ij}^{(t)} = 1$.

We shall assume that the Markov chain is time-independent, i.e., transition probabilities do not depend on time: $p_{ij}^{(t)} = p_{ij}$.

Then a Markov chain on $S = \{1, 2, \dots, k\}$ is determined by a **probability vector** $\mathbf{x}_0 = (q_1, q_2, \dots, q_k) \in \mathbb{R}^k$ and a $k \times k$ **transition matrix** $P = (p_{ij})$. The entries in each row of P add up to 1.

Let $s_0, s_1, s_2, \dots$ be the Markov chain. Then the vector $\mathbf{x}_0$ determines the probability distribution of the initial state s_0.

Problem. Find the (unconditional) probability distribution for any s_n.