특성다항식과 Jordan Normal Form

崔, E-mail Address

June 2, 2023

Abstract

In this paper, we will explore some properties of the characteristic polynomial of an matrix in $M_n(\mathbb{C})$. Moreover, we will introduce the Jordan Decomposition Theorem, and discuss some of its most important consequences.

1 Matrix Polynomials

행렬 다항식(Matrix Polynomial)은 정사각행렬을 변수로 가지는 다항식이다. 예를 들어, 다음과 같은 다항식 $\mathbf{P}$가 주어져 있다고 해보자.

$$P(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

여기서 모든 a_i $(0 \le i \le n)$는 복소수이고 x 역시 복소수이다. 즉, $\mathbf{P}$는 $\mathbb{C}$를 $\mathbb{C}$로 매핑하는 함수이다. 이때 임의의 정사각행렬 $A \in M_n(\mathbb{C})$에 대하여 x 자리에 A를 대입하고, 상수항에 I_n을 곱하면 $P(A)$를 다음과 같이 정의할 수 있다. (단, $A^0 = I_n$으로 정의하였다.)

$$P(A) = \sum_{i=0}^{n} a_i A^i = a_0 I_n + a_1 A + a_2 A^2 + \cdots + a_n A^n$$

$p(A)$의 이러한 정의는 $\mathbb{C} \to \mathbb{C}$의 매핑으로서 정의되던 다항식 $\mathbf{P}$의 여러 성질과 계산 법칙을 보존하는, 아주 자연스러운 정의이다. 또한, 다항식 $\mathbf{P}$가 일차 이하가 아닌 이상 $P(A)$에서 A^2 항이 등장하므로, 행렬 A를 정사각행렬로 한정하는 것 역시 타당한 조건이다. (어떤 행렬의 거듭제곱이 정의되기 위해서 그 행렬은 반드시 정사각행렬이어야 한다.)

2 The Characteristic Polynomial

$n \times n$ 행렬 A의 특성다항식은 $p_A(\lambda) = \det(\lambda I_n - A)$로 정의되는 scalar-valued 다항식이다. (정의 방식에 따라 $\det(A - \lambda I_n)$을 특성다항식으로 정의하는 경우도 있다.) 특성다항식의 n개의 복소 영점은 행렬 A의 고윳값이며, 특성다항식을 행렬 다항식으로 확장하게 되면 다음이 성립한다.

$$p_A(A) = \mathbf{0}$$

여기서 $\mathbf{0}$은 $n \times n$ 영행렬이다. 즉, 행렬 다항식으로 확장된 특성다항식에 A 자신을 대입하면 그 결과가 항상 영행렬이라는 것이다. 이를 Cayley-Hamilton Theorem이라 한다. 또한 A를 대입하였을 때 영행렬이 나오도록 하는 최소 차수의 monic polynomial(최고차항의 계수가 1인 다항식)을 행렬 A의 minimal polynomial이라고 하며, $P(A) = \mathbf{0}$을 만족시키는 모든 행렬 다항식 $\mathbf{P}$는 minimal polynomial μ_A의 배수이므로 행렬 A의 특성다항식 p_A 역시 μ_A의 배수가 된다. Cayley-Hamilton Theorem의 증명은 여기서는 자세하게 다루지 않겠다.

놀랍게도 행렬 A의 특성다항식 $p_A(\lambda)$의 모든 계수는 A의 원소들에 관한 정보로 표현할 수 있으며, 더 나아가 오직 $\det(A)$와 $\mathrm{tr}(A)$에 관한 식으로 나타낼 수 있다.

예를 들어, 다음과 같이 주어진 2×2 행렬 A를 고려하자.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

여기서 a, b, c, d는 모두 복소수이다. 이때 A의 특성다항식은

$$\begin{aligned} p_A(\lambda) = \det(\lambda I_2 - A) &= \lambda^2 - (a+d)\lambda + ad - bc \\ &= \lambda^2 - \operatorname{tr}(A)\lambda + \det(A) \end{aligned}$$

이므로 2×2 행렬의 특성다항식의 모든 계수를 상수 또는 $\operatorname{tr}(A)$와 $\det(A)$로 표현할 수 있음을 알 수 있다. 이 다항식을 행렬 다항식으로 확장하여 A를 대입하면

$$p_A(A) = A^2 - \operatorname{tr}(A)A + \det(A) = \mathbf{0}_{2\times 2}$$

이며, A를 직접 대입하여 계산하여도 동일한 결과를 얻음을 쉽게 확인할 수 있다.

한편 $n \times n$ 행렬 A에 대하여 특성다항식의 계수를 구하는 것은 더 깊은 수준의 이론을 요구한다. 추상대수학과 미분기하학에서 다뤄지는 외대수(Exterior Algebra) 또는 그라스만 대수(Grassmann Exterior)의 결과를 이용하면 A의 kth exterior power $\Lambda^k A$에 대하여 $p_A(\lambda)$를 다음과 같이 쓸 수 있다.

$$p_A(\lambda) = \sum_{k=0}^{n} \lambda^{n-k}(-1)^k \operatorname{tr}(\Lambda^k A)$$

또한 모든 순서체(Ordered Field)와 복소수체 $\mathbb{C}$는 0의 characteristic을 가지므로 $\operatorname{tr}(\Lambda^k A)$는 다음의 단일 행렬식으로 계산할 수 있다.

$$\operatorname{tr}(\Lambda^k A) = \frac{1}{k!} \begin{vmatrix} \operatorname{tr}(A) & k-1 & 0 & \cdots & 0 \\ \operatorname{tr}(A^2) & \operatorname{tr}(A) & k-2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \operatorname{tr}(A^{k-1}) & \operatorname{tr}(A^{k-2}) & & \cdots & 1 \\ \operatorname{tr}(A^k) & \operatorname{tr}(A^{k-1}) & & \cdots & \operatorname{tr}(A) \end{vmatrix}$$

즉, 행렬 A의 특성다항식은 다음과 같다.

$$p_A(\lambda) = \sum_{k=0}^{n} \lambda^{n-k} \frac{(-1)^k}{k!} \begin{vmatrix} \operatorname{tr}(A) & k-1 & 0 & \cdots & 0 \\ \operatorname{tr}(A^2) & \operatorname{tr}(A) & k-2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \operatorname{tr}(A^{k-1}) & \operatorname{tr}(A^{k-2}) & & \cdots & 1 \\ \operatorname{tr}(A^k) & \operatorname{tr}(A^{k-1}) & & \cdots & \operatorname{tr}(A) \end{vmatrix}$$

마지막으로 특성다항식에 관한 유용한 정리 하나를 소개하겠다.

Thm. Let A be a square $n \times n$ matrix and let $f(\lambda)$ be a polynomial. If the characteristic polynomial of A has a factorization

$$p_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

then the characteristic polynomial of the matrix $f(A)$ is given by

$$p_{f(A)}(\lambda) = (\lambda - f(\lambda_1))(\lambda - f(\lambda_2))\cdots(\lambda - f(\lambda_n))$$

증명에 앞서 Jordan Decomposition Theorem을 소개하겠다.

Jordan Decomposition Theorem

Every square complex matrix A is similar to a block diagonal matrix

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix}$$

where each block J_i is a square matrix of the form

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

and each λ_i is an eigenvalue of A. Hence, there exists an invertible matrix P such that $P^{-1}AP = J$ is such that the only non-zero entries of J are on the diagonal and the super-diagonal. J is called the **Jordan normal form** of A, and each J_i is called a **Jordan block** of A.

이제 앞서 제시한 정리를 증명해보자. $A \in M_n(\mathbb{C})$이고, 대수학의 기본 정리에 의해 임의의 n차 복소 계수 다항식은 복소 범위에서 중복을 고려하여 n개의 영점을 가지므로 A의 특성다항식은 복소 범위에서 다음과 같이 인수분해된다.

$$p_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

여기서 $\lambda_1, \lambda_2, \ldots, \lambda_n$은 A의 고윳값이다. 또한 Jordan Decomposition Theorem에 의해 가역행렬 S와 상삼각행렬 U가 존재하여 $A = S^{-1}US$이다. (여기서 U의 주대각선에는 A의 고윳값 $\lambda_1, \lambda_2, \ldots, \lambda_n$이 있다.)

$f(\lambda) = \sum_i \alpha_i \lambda^i$라 하자. 이제 f를 행렬 다항식으로 확장하고 A를 대입하면

$$\begin{aligned} f(A) = \sum \alpha_i A^i &= \sum \alpha_i (S^{-1}US)^i \\ &= \sum \alpha_i (S^{-1}USS^{-1}US \cdots S^{-1}US) \\ &= \sum \alpha_i S^{-1}U^i S \\ &= S^{-1}(\sum \alpha_i U^i)S \\ &= S^{-1}f(U)S \end{aligned}$$

이때 U는 주대각선에 $\lambda_1, \lambda_2, \ldots, \lambda_n$이 놓인 상삼각행렬이므로 U^i는 주대각선에 $\lambda_1^i, \lambda_2^i, \ldots, \lambda_n^i$이 놓인 상삼각행렬이 된다.

따라서

$$f(U) = \sum_i \alpha_i U^i = \sum_i \alpha_i \begin{bmatrix} \lambda_1^i & & \times \\ & \ddots & \\ 0 & & \lambda_n^i \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i \alpha_i \lambda_1^i & & \times \\ & \ddots & \\ 0 & & \sum_i \alpha_i \lambda_n^i \end{bmatrix}$$

$$= \begin{bmatrix} f(\lambda_1) & & \times \\ & \ddots & \\ 0 & & f(\lambda_n) \end{bmatrix}$$

이고 $f(U)$는 주대각선에 $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$이 놓인 상삼각행렬이다. 즉, $f(U)$의 고윳값은 $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$이다. 한편 $f(A) = S^{-1} f(U) S$이고 S는 자명하게 가역이므로 $f(A)$는 $f(U)$와 닮아 있고, 같은 대수적 중복도의 동일한 고윳값을 가지게 된다. 따라서 고윳값의 정의에 의해 $f(A)$의 특성다항식은

$$p_{f(A)}(\lambda) = (\lambda - f(\lambda_1))(\lambda - f(\lambda_2)) \cdots (\lambda - f(\lambda_n))$$

으로 구해진다. ■

3 Consequences of the Jordan Normal Form

3.1 Characteristic Polynomial

행렬 $A \in M_n(\mathbb{C})$의 특성다항식은 $p_A(\lambda) = \det(\lambda I_n - A)$로 정의되는 scalar-valued polynomial이다. 한편 서로 닮음인 두 행렬은 동일한 특성다항식을 가지므로 A의 Jordan normal form J에 대하여

$$p_A(\lambda) = p_J(\lambda) = \prod_i (\lambda - \lambda_i)^{m_i}$$

이다. 여기서 λ_i는 p_J의 i번째 영점이며 m_i는 그 대수적 중복도이다.

3.2 Cayley-Hamilton Theorem

Cayley-Hamilton Theorem은 임의의 행렬 $A \in M_n(\mathbb{C})$와 그 특성다항식 p_A에 대하여 $p_A(A) = \mathbf{0}_{n \times n}$이 성립한다는 내용의 정리이며, Jordan normal form에서 직접적인 계산을 통해 어렵지 않게 보일 수 있다. A의 한 고윳값 λ_i의 대수적 중복도를 m_i라 하면, A의 Jordan block J_i는 주대각선에 λ_i가 놓인 $m_i \times m_i$ 상삼각행렬이므로 $(J_i - \lambda_i I_{m_i})^{m_i} = \mathbf{0}_{m_i \times m_i}$이다. 또한 A 각각의 diagonal block(Jordan block)은 서로 영향을 주지 않으므로 $(A - \lambda_i I_n)^{m_i}$의 i번째 diagonal block은 $(J_i - \lambda_i I_{m_i})^{m_i} = \mathbf{0}_{m_i \times m_i}$이다. 즉,

$$p_A(A) = \prod_i (A - \lambda_i I_n)^{m_i} = \mathbf{0}_{n \times n}$$

이 되어 증명이 완료되었다. ■

3.3 Matrix Functions

$\mathbb{C}$를 정의역으로 가지는 해석함수 $f(z)$를 고려하고, f의 멱급수 표현을 $f(z) = \sum_{n=0}^{\infty} c_n z^n$이라 하자. f를 $M_n(\mathbb{C})$를 $M_n(\mathbb{C})$로 매핑하는 matrix function으로 확장하여 λ를 고윳값으로 가지는 $n \times n$ Jordan block J을 대입하면 다음과 같다.

$$f(J) = \sum_{n=0}^{\infty} c_n J^n = \sum_{n=0}^{\infty} c_n \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}^n$$

한편 직접 계산을 통해 J^n의 일반적인 공식이 다음과 같음을 어렵지 않게 유추할 수 있고, n에 대한 수학적 귀납법을 이용하여 증명할 수 있다. (J를 $k \times k$ 행렬이라 하자.)

$$J^n = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}^n = \begin{bmatrix} \lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \cdots & \binom{n}{k-1}\lambda^{n-k+1} \\ 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \cdots & \binom{n}{k-2}\lambda^{n-k+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{n}{1}\lambda^{n-1} \\ 0 & 0 & 0 & \cdots & \lambda^n \end{bmatrix} \tag{1}$$

순수한 수식적 유도를 위해 superdiagonal(주대각선 바로 위의 대각선)의 원소가 모두 1이고 나머지는 모두 0으로 채워진 다음과 같은 행렬 $J - \lambda I_k = N$을 정의하자.

$$N = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

이러한 형태의 행렬 N은 nilpotent임이 잘 알려져 있고, 더 나아가 $N^k = \mathbf{0}_{k\times k}$이다. 따라서 이항정리에 의해 다음이 성립한다.

$$J^n = (\lambda I_k + N)^n = \sum_{r=0}^{n} \binom{n}{r} \lambda^{n-r} N^r = \sum_{r=0}^{\min(n,k-1)} \binom{n}{r} \lambda^{n-r} N^r \tag{2}$$

또한 N^r은 직접 계산을 통해 다음과 같이 구해짐을 유추할 수 있다. (단, $0 \le r \le k$)

$$[N^r]_{ij} = \begin{cases} 1 & (j - i = r) \\ 0 & (\text{otherwise}) \end{cases} \tag{3}$$

$r = 0$일 때 $N^0 = I_k$가 됨은 자명하므로 $r = m$일 때 성립한다고 가정한 후 $r = m+1$일 때를 살펴보자. 행렬 곱셈의 정의에 의해

$$[N^{m+1}]_{ij} = [N \times N^m]_{ij} = \sum_{p=1}^{k} [N]_{ip}[N^m]_{pj}$$

이다. $p - i = 1$인 경우 $[N]_{ip} = 1$이고, $j - p = m$인 경우 $[N^m]_{pj} = 1$이므로 $[N]_{ip}[N^m]_{pj} = 1$이 되기 위해서는 이 두 조건이 모두 만족되어야 한다. 그렇지 않을 경우 두 항 중 적어도 하나가 0이 되어 $[N]_{ip}[N^m]_{pj} = 0$이 된다. 즉, $p = i + 1 = j - m$이고 $j - i = m + 1$인 경우만 $[N]_{ip}[N^m]_{pj} = 1$이 되고, 이때의 p는 $i + 1$로 유리하므로 $\sum_{p=1}^{k}[N]_{ip}[N^m]_{pj}$의 값 역시 1이 된다. (그 외의 경우는 모두 0이다.) 따라서

$$[N^{m+1}]_{ij} = \begin{cases} 1 & (j - i = m + 1) \\ 0 & (\text{otherwise}) \end{cases}$$

이므로 수학적 귀납법에 의해 (3)의 추측이 증명되었고, 이를 식 (2)에 대입하여 계산하면 식 (1)이 성립한다는 것 역시 쉽게 보여진다. ■

따라서 $f(J)$는 다음과 같이 계산할 수 있다.

$$f(J)=\sum_{n=0}^{\infty}c_nJ^n=\sum_{n=0}^{\infty}c_n\begin{bmatrix}\lambda^n & \binom{n}{1}\lambda^{n-1} & \binom{n}{2}\lambda^{n-2} & \cdots & \binom{n}{k-1}\lambda^{n-k+1}\\ 0 & \lambda^n & \binom{n}{1}\lambda^{n-1} & \cdots & \binom{n}{k-2}\lambda^{n-k+2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \binom{n}{1}\lambda^{n-1}\\ 0 & 0 & 0 & \cdots & \lambda^n\end{bmatrix}$$

$$=\begin{bmatrix}\sum_n c_n\lambda^n & \sum_n c_n\binom{n}{1}\lambda^{n-1} & \sum_n c_n\binom{n}{2}\lambda^{n-2} & \cdots & \sum_n c_n\binom{n}{k-1}\lambda^{n-k+1}\\ 0 & \sum_n c_n\lambda^n & \sum_n c_n\binom{n}{1}\lambda^{n-1} & \cdots & \sum_n c_n\binom{n}{k-2}\lambda^{n-k+2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \sum_n c_n\binom{n}{1}\lambda^{n-1}\\ 0 & 0 & 0 & \cdots & \sum_n c_n\lambda^n\end{bmatrix}$$

$$=\boxed{\begin{bmatrix}\frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(k-1)}(\lambda)}{(k-1)!}\\ 0 & \frac{f(\lambda)}{0!} & \frac{f'(\lambda)}{1!} & \cdots & \frac{f^{(k-2)}(\lambda)}{(k-2)!}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \frac{f'(\lambda)}{1!}\\ 0 & 0 & 0 & \cdots & \frac{f(\lambda)}{0!}\end{bmatrix}}$$

이러한 정의를 이용하여 matrix function의 정의역을 spectral radius가 멱급수의 수렴 반경보다 작은 행렬들의 집합으로 확장할 수 있다.

References

[1] Steven Roman (1992). Advanced linear algebra (2 ed.). Springer. p. 137. ISBN 3540978372.

[2] Horn, Roger A.; Johnson, Charles R. (2013). Matrix Analysis (2nd ed.). Cambridge University Press. pp. 108–109, Section 2.4.2. ISBN 978-0-521-54823-6.

[3] Beauregard, Raymond A.; Fraleigh, John B. (1973), A First Course In Linear Algebra: with Optional Introduction to Groups, Rings, and Fields, Boston: Houghton Mifflin Co., ISBN 0-395-14017-X, pp. 310-316.

[4] Cullen, Charles G. (1966), Matrices and Linear Transformations, Reading: Addison-Wesley, LCCN 66021267, p. 114.