

# **MATH2999 Directed Studies in Mathematics**

## **Matrix theory and its applications**

## **Minimal Polynomials and its applications**

By

ZHANG SHIXIAO

2010800331

Supervisors: Prof. Chi-Kwong Li and Dr. J. T. Chan

# Contents

## Abstract

1. Introduction: background
2. Power of Matrices
  - 2.1 Characteristic polynomial and minimal polynomial of a matrix
  - 2.2 Efficient ways to compute minimal polynomials
  - 2.3 Three schemes to compute the power of matrices
  - 2.4 Application in controllability of systems

## Abstract

In this project, our focus will be the fundamental theory about minimal polynomials and power of matrices. It is important to state out both theoretical notations and applications in various fields. We will start with giving out mathematical characters description about general properties of minimal polynomials, and aim at figuring out some appropriate ways to compute the power of .

### 1. Introduction: Background

*Matrix Theory* (or *Linear Algebra*) sometimes is regarded as one of the two most important concepts in mathematics (the other is *Calculus* of course). Matrices are widely used in both science and engineering aspects. It has become a sufficient tool to represent different systems, simulate various processes and compute or make the final decision results. It is well-known that minimal polynomials are treated as one of the most pivotal concepts to reduce the power of matrices, which sometimes represent the long-term variation tendency. They have been probably full-understood and been transferred to not only mathematics applications, but also statistics, economics, dynamic systems, so on and so forth. Researchers have studied minimal polynomials for a long time, different measurements and methods have been developed to calculate the power of matrices, which is used to describe the flow direction of a particular system. Therefore, we are aiming at figuring out the underlying foundation of these methods and make a connection and comparison about different sections.

### 2. Power of matrices

#### 2.1 Characteristic polynomial and minimal polynomial of a matrix

2.2.1 First note that the analysis for functions could be easily transferred to matrices. Generally speaking, for analytic functions, if  $S$  is the set of all matrices  $T$  such that  $\sum_{n=0}^{\infty} \alpha_n T^n$  converges, then we define  $f$  to be  $f(T) = \sum_{n=0}^{\infty} \alpha_n T^n$  for all  $T$  belongs to  $S$  and we can see  $f$  is a matrix-valued function. As for a square matrix, there must exist some polynomials  $f(\lambda)$  such that when we plug in the matrix  $A$ ,  $f(A) = 0$ . One famous such polynomial is called *characteristic polynomial* of a matrix  $A$ , which is denoted as  $p_A(z) = \det(zI - A)$ . The result  $p_A(A) = 0$ , one should recall, is the famous *Cayley–*

*Hamilton Theorem*, which we will discuss in the latter section. On the other hand, among all such polynomials, there must one kind of polynomial with the lowest degree  $\mu$ , and when we divide such polynomials by the coefficient of  $\lambda^\mu$ , the result must be identical. Otherwise, if we obtain  $f_1(\lambda)$  and  $f_2(\lambda)$  which  $f_1(\lambda) \neq f_2(\lambda)$ , then  $f_1(\lambda) - f_2(\lambda) = \sum_{i=0}^{\mu-1} \alpha_i \lambda^i$ , substitute  $A$ , we will get  $f_1(A) - f_2(A) = \sum_{i=0}^{\mu-1} \alpha_i A^i = 0$ , which contradicts with the lowest degree assumption. Therefore, we denote such lowest degree polynomial to be the minimal polynomial of a matrix and write as  $m(A)$ .

### 2.1.2 Existence of minimal polynomial

To be a minimal polynomial of a matrix, there are three conditions must be satisfied, namely, (1)  $p(A)=0$  (2)  $p$  has the lowest degree which means if  $m'$  is another nonzero polynomial such that  $m'(A)=0$ ,  $\deg(m') \geq \deg(m)$ . (3)  $p$  is monic.

As a sequence of matrices, we denote  $m$  to be the smallest integer such that  $I, A, \dots, A^m$  are linearly dependent. By saying linearly dependent, we mean that there exist some coefficients  $\alpha_1, \alpha_2, \dots, \alpha_m$  which are not all equal to zero, and  $\sum_{i=1}^m \alpha_i A^i = 0$ .

Because of  $m$  is the minimal integer defined as above,  $\alpha_m \neq 0$ , then  $\alpha_m A^m + \sum_{i=1}^{m-1} \alpha_i A^i = 0$

could be rewrite as  $A^m + \sum_{i=1}^{m-1} \beta_i A^i = 0$  where  $\beta_i = \frac{\alpha_i}{\alpha_m}$  for  $i=1, 2, \dots, m-1$ . Let

$p_A(\lambda) = \lambda^m + \sum_{i=1}^{m-1} \beta_i \lambda^i$ , then observed that  $p_A(A) = 0$ , hence  $p$  is the minimal polynomial of  $A$ .

### 2.1.3 Uniqueness of minimal polynomial

If  $p$  and  $p'$  are both minimal polynomials of  $A$ , we divide  $p'$  by  $p$ , use the Euclidean Algorithm for Polynomials, we will get  $p'(\lambda) = p(\lambda)q(\lambda) + r(\lambda)$ , substitute  $A$ ,  $p'(A) = p(A)q(A) + r(A)$  which means  $r(A) = 0$ . On the other hand,  $\deg(r) < \deg(p)$ , hence  $r(A) = 0$  according to the lowest degree assumption. Therefore,  $p'(r) = p(r)q(r)$ . Moreover,  $p'$  and  $p$  must have the same degree, which means  $q(\lambda)$  is just a constant polynomial and equal to 1 according to the monic condition. That means  $p' = p$ , which is the uniqueness of minimal polynomial.

## 2.2 Efficient ways to compute minimal polynomials

The existence of minimal polynomial also highlights a method for us to find it. We solve the equation,  $A = \beta_0 I$ , if no solution, we move on to solve  $A^2 = \beta_0 I + \beta_1 A$ , do such an algorithm step by step until we get a solution such that  $A^m = \sum_{i=1}^{m-1} \beta_i A^i$ . Look at the property of Cayley–Hamilton theorem, such an algorithm must stop by some finite number of steps. On the other hand, for the matrix  $A = (a_{ij}) \in M_n$ , define

$$\text{vec}A = [a_{11} \ a_{12} \ \dots \ a_{1n} \ a_{21} \ a_{22} \ \dots \ a_{2n} \ \dots \ a_{n1} \ a_{n2} \ \dots \ a_{nn}]^T$$

$$A^0 = I \in M_n,$$

$$A^k = A^{k-1}A \quad (k = 1, 2, \dots),$$

$$A^k = [a_{ij}^{(k)}] \quad (k = 0, 1, 2, \dots),$$

$$a^{(k)} = \text{vec}A^k \quad (k = 0, 1, 2, \dots),$$

$$B_k = [a^{(0)} \ a^{(1)} \ \dots \ a^{(k)}] \quad (k = 0, 1, 2, \dots)$$

where  $a^{(k)}$  is  $k + 1$ -th column of the matrix  $B_k \in M_{n^2, k+1}$

then  $K = \{k \in N : \text{rank}(B_k) = \text{rank}(B_{k-1})\} \neq \emptyset$ . Moreover, if  $k_0 = \min K$  is the associated rank of matrix  $A$  and  $p$  is the minimal polynomial of  $A$ , then  $\text{rank}(B_k) = k + 1$  for  $k = 0, 1, \dots, k_0 - 1$ , and  $\text{rank}(B_k) = k_0$  for  $k \geq k_0$ ,  $\text{deg}(p) = k_0$  (S. BIAŁAS and M. BIAŁAS, 2008).

Now we consider  $B_n = (a^{(0)} \ a^{(1)} \ \dots \ a^{(n)}) \in M_{n^2, n+1}$ , while  $B_n = (b_{ij})$ ,  $b_{11} = 1, b_{12} = a_{11}^{(1)}, \dots, b_{1, n+1} = a_{11}^{(n)}, \dots, b_{n^2, n+1} = a_{nn}^{(n)}$ , then we calculate the rank of  $B_n$ , which is

$$\text{rank}(B_n) = \text{rank} \begin{pmatrix} 1 & b_{12} & \dots & b_{1, n+1} \\ 0 & b_{22}^{(1)} & \dots & b_{2, n+1}^{(1)} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & b_{n^2, 2}^{(1)} & \dots & b_{n^2, n+1}^{(1)} \end{pmatrix}, \text{ therefore, there exists } r \in N \text{ such that } r \leq N$$

and

$$\text{rank} B_n = \text{rank} \begin{bmatrix} 1 & b_{12} & b_{13} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & b_{1,n+1} \\ 0 & b_{12}^{(1)} & b_{23}^{(1)} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & b_{2,n+1}^{(1)} \\ 0 & 0 & b_{33}^{(2)} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & b_{3,n+1}^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & b_{rr}^{(r-1)} & b_{r,r+1}^{(r-1)} & \dots & \dots & \dots & b_{r,n+1}^{(r-1)} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & b_{r+1,r+2}^{(r-1)} & \dots & \dots & b_{r+1,n+1}^{(r-1)} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & b_{r+2,r+2}^{(r-1)} & \dots & \dots & b_{r+2,n+1}^{(r-1)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & b_{n^2,r+2}^{(r-1)} & \dots & \dots & b_{n^2,n+1}^{(r-1)} \end{bmatrix}$$

If we set  $\alpha_0 + \alpha_1 A + \dots + \alpha_{k_0-1} A^{k_0-1} + \alpha_{k_0} A^{k_0} = 0$ , where  $\alpha_i$ 's are the coefficients of the minimal polynomial of the matrix  $A$ , then the only solution is obtained by solving

$$\tilde{B}\alpha = \tilde{b} \text{ where}$$

$$\tilde{B} = \begin{bmatrix} 1 & b_{11} & \dots & \dots & b_{1r} \\ 0 & b_{22}^{(1)} & \dots & \dots & b_{2r}^{(1)} \\ 0 & 0 & b_{33}^{(2)} & \dots & b_{3r}^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & b_{rr}^{(r-1)} \end{bmatrix}, \tilde{b} = \begin{bmatrix} b_{1,r+1} \\ b_{2,r+1}^{(1)} \\ \vdots \\ b_{r,r+1}^{(r-1)} \end{bmatrix}$$

$$\alpha = [\alpha_0 \ \alpha_1 \ \dots \ \alpha_{k_0-1}]^T, \ r = k_0.$$

### 2.3 Three schemes to compute the power of matrices

#### 2.3.1 Diagonalizable matrices

In the simplest case, when the matrix is diagonalizable, we can always use the diagonality/similarity method to carry out the diagonal matrix. Recall that a  $n$  by  $n$  matrix is diagonalizable if it has  $n$  distinct eigenvalues. In general, we write

$$SPS^{-1} = \begin{pmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n \end{pmatrix}. \text{ Observed that if the first row of } S \text{ is normalized to } \pi, \text{ then the}$$

first column of  $S^{-1}$  must be normalized to  $\mathbf{1}$  since  $SS^{-1} = I$  and hence  $(SV)_{1l} = u_l v_l = \pi v_l = 1$  where  $u_i$  and  $v_i$ 's are the row vectors and column vector of  $S$  and  $S^{-1}$  respectively.

Denote  $\Lambda = \begin{pmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, P^t = S^{-1} \begin{pmatrix} 1 & & & \\ & \lambda_2^t & & \\ & & \ddots & \\ & & & \lambda_n^t \end{pmatrix} S$

$\xrightarrow{t \rightarrow \infty} \begin{pmatrix} v_{11}u_1 \\ v_{12}u_1 \\ \vdots \\ v_{1n}u_1 \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix}$ , where  $v_{ij}$ 's are the  $j$ th entry of the column vector  $v_i$  since  $P$  has

the unique largest eigenvalue  $\lambda_1 = 1$  by Perron-Frobenius Theorem. One may consider

$1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ , generally, given an initial distribution  $q_0, q^{(t)} = qP^t = \pi + \sum_{i=2}^n q_i' \lambda_i^t u_i$

, where  $q_i' = \frac{qu_i^T}{\|u_i\|^2}$ ,  $u_i$ 's are just the eigenvalue basis for  $P$ .

Numerical example:

Diagonalize  $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{4} & 0 \\ 0 & 1 \end{pmatrix}$

$\lim_{n \rightarrow \infty} P^n = \lim_{n \rightarrow \infty} \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} (-\frac{1}{4})^n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{2}{5} \end{pmatrix}$

$(I - zP)^{-1} = (1 + z + z^2 + z^3 + \dots) \begin{pmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{3}{5} & \frac{2}{5} \end{pmatrix} + [1 - \frac{z}{4} + (\frac{z}{4})^2 - (\frac{z}{4})^3 + \dots] \begin{pmatrix} \frac{2}{5} & -\frac{2}{5} \\ -\frac{3}{5} & \frac{3}{5} \end{pmatrix}$

2.3.2 Cayley-Hamilton Theorem: suppose  $A$  is a  $n$  by  $n$  matrix, then characteristic polynomial we define above is  $p_A(z) = \det(zI - A)$ , then  $p(A) = 0$ . Without loss of generality, let us write  $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_0$ , denote  $B(\lambda)$  to be the adjoint matrix of  $\lambda I - A$ , then

$B(\lambda)(\lambda I - A) = |\lambda I - A|I = p(\lambda)I = \lambda^n I + a_{n-1}\lambda^{n-1}I + \dots + a_0I$ , on the other hand, write  $B(\lambda)$  as  $B(\lambda) = \lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \dots + \lambda B_1 + B_0$ , where  $B_i$ 's are  $n$  by  $n$  matrices.

Then we get

$$\begin{aligned} B(\lambda)(\lambda I - A) &= (\lambda^{n-1}B_{n-1} + \lambda^{n-2}B_{n-2} + \dots + \lambda B_1 + B_0)(\lambda I - A) \\ &= \lambda^n B_{n-1} + \lambda^{n-1}(B_{n-2} - B_{n-1}A) + \lambda^{n-2}(B_{n-3} - B_{n-2}A) + \dots + \lambda(B_0 - B_1A) - B_0A \\ &= \lambda^n I + a_{n-1}\lambda^{n-1}I + \dots + a_0I \end{aligned}$$

Compare the coefficients appearing in the each side of the above equation, we conclude that

$$\begin{aligned} B_{n-1} &= I \\ B_{n-2} - B_{n-1}A &= a_{n-1}I \\ &\vdots \\ -B_0A &= a_0I \end{aligned}$$

Moreover, we multiply  $A^n, A^{n-1}, \dots, A, I$  to each of the above equation respectively, add them together, we can easily get  $p(A)=0$ . Cayley-Hamilton Theorem provides us with an efficient way to reduce or represent the high degree power of matrices with some lower degree power of matrices, which is similar to the previous minimal polynomial method. In particular,  $A^m = p(A)q(A) + r(A)$ , where  $p(A)$  is the characteristic polynomial of  $A$ . Then  $A^m = r(A) = r_{n-1}A^{n-1} + \dots + r_0I$ . As for the diagonalizable case, same as the previous notation, if we denote

$$\begin{aligned} S &= (u_1 \quad u_2 \quad \dots \quad u_n) \text{ and } S^{-1} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \\ \text{then } A^m &= r_{n-1}S \begin{pmatrix} \lambda_1^{n-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{n-1} \end{pmatrix} S^{-1} + r_{n-2}S \begin{pmatrix} \lambda_1^{n-2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{n-2} \end{pmatrix} S^{-1} + \dots + r_0I \\ &= S(r_{n-1} \begin{pmatrix} \lambda_1^{n-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{n-1} \end{pmatrix} + r_{n-2} \begin{pmatrix} \lambda_1^{n-2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{n-2} \end{pmatrix} + \dots + r_0I)S^{-1} \\ &= S \begin{pmatrix} \lambda_1^m & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^m \end{pmatrix} S^{-1} \end{aligned}$$

What we need to do is solve a batch of equations, namely

$$\begin{pmatrix} \lambda_1^m \\ \lambda_2^m \\ \vdots \\ \lambda_n^m \end{pmatrix} = \begin{pmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \lambda_n & \cdots & \lambda_n^{n-1} \end{pmatrix} \begin{pmatrix} r_0 \\ r_1 \\ \vdots \\ r_{n-1} \end{pmatrix}$$

### 2.3.3 Partial fraction decomposition

Write  $(I - zA)^{-1}$  as a power series expansion, and then compare the coefficients of  $A^m$  to get the representative.

$$(I - zA)^{-1} = \frac{1}{\det(I - zA)} \text{adj}(I - zA) = \left( \sum_{k=0}^{\infty} \gamma_k z^k \right) (z^{n-1} B_{n-1} + z^{n-2} B_{n-2} + \cdots + z B_1 + I)$$

This should involve the computation of polynomials, especially minimal polynomials of a univariate polynomial matrix. You may refer to (Nicolas, 2005).

## 2.4 Application in controllability of systems

2.4.1 In modern mathematical control theory, stability, controllability and observability are three essential characters describing a system. Here we mainly focus on controllability of systems, and try to highlight some applications of power of matrices and minimal polynomials involved. In general, the mathematical model for a linear time-invariant dynamical control system is represented by  $x'(t) = Ax(t) + Bu(t)$ , where  $x(t) \in R^n$  is a state vector,  $u(t) \in R^m$  is an input vector, A, B are real matrices of appropriate dimensions. By the Existence and Uniqueness theorem for differential equations, give an initial value  $x(0) \in R^n$  and control  $u(t) \in R^m$ , there must exist a unique solution for  $x(t; x(0), u) = e^{tA} x(0) + \int_0^t e^{(t-s)A} Bu(s) ds$ .

Definition: a system is said to be controllable (reachable) if for every initial condition  $x(0)$  and every vector  $x^I \in R^n$ , there exist a finite time  $t_I$  and control  $u(t) \in R^m$ ,  $t \in [0, t_I]$ , such that  $x(t_I; x(0), u) = x^I$ .

For a discrete time case, we write  $x_{t+1} = Ax_t + Bu_t$  to represent the system. The controllability concept could be transferred to the following equation:

$$x_t = \begin{pmatrix} B & AB & A^2B & \cdots & A^{t-2}B & A^{t-1}B \end{pmatrix} \begin{pmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{pmatrix}$$

where  $W_t = (B \ AB \ A^2B \ \dots \ A^{t-2}B \ A^{t-1}B)$  is said to be the controllability matrix. By the Cayley-Hamilton theorem, we can express each  $A^t$  for  $t \geq n$  as a linear combination of  $A^0, \dots, A^{n-1}$ , hence  $\text{rank}(W_t) = \text{rank}(W_n)$ . Thus, the system is controllable if and only if  $\text{rank}(W_t) = n$ .

For a continuous time case, the idea is similar while the proof is slightly different, you may refer to (Klamka 2008), a sketch proof is that the dynamical system is controllable if and only if for certain time  $t=t_1$  the range of integral operator

$\int_0^t e^{(t-s)A} Bu(s) ds = R^n$ . However, since  $\exp(t_1 A)$  is nonsingular for any  $t_1$  if and only if  $\int_0^{t_1} e^{-sA} BB^T e^{-sA^T} ds$  is nonsingular. Taking into account Taylor series expansion of  $\exp(-sA)$  and Cayley-Hamilton theorem we conclude that dynamical system is controllable if and only if  $\text{rank}(W_t) = n$ .

## References

1. Seneta, E. *Non-negative matrices and Markov chains*. New York: Springer, c2006. [2nd ed.], rev. print.
2. Watkins, David S. *Fundamentals of matrix computations*. New York: Wiley-Interscience, c2002. 2nd ed.
3. Pullman, Norman J. *Matrix theory and its applications: selected topics*. New York: M. Dekker, c1976.
4. Stewart, G. W. *Introduction to matrix*. New York: Academic Press, 1973.
5. NICHOLAS P. KARAMPETAKIS, PANAGIOTIS TZEKIS. *ON THE COMPUTATION OF THE MINIMAL POLYNOMIAL OF A POLYNOMIAL MATRIX*. Int. J. Appl. Math. Comput. Sci., 2005, Vol. 15, No. 3, 339–349.
6. S. BIAŁAS and M. BIAŁAS. *An algorithm for the calculation of the minimal polynomial*. Vol. 56, No. 4, 2008.
7. Stanford. *LQR, Controllability and Observability*. EE363 Review Session 1.
8. Klamka, J. *SYSTEM CHARACTERISTICS STABILITY, CONTROLLABILITY, OBSERVABILITY*. Institute of Automatic Control, Technical University, Gliwice, Poland. CONTROL SYSTEMS, ROBOTICS AND AUTOMATION - Vol. VII.
9. Klamka, J. *Controllability of dynamical systems*. MATEMATYKA STOSOWANA 9, 2008