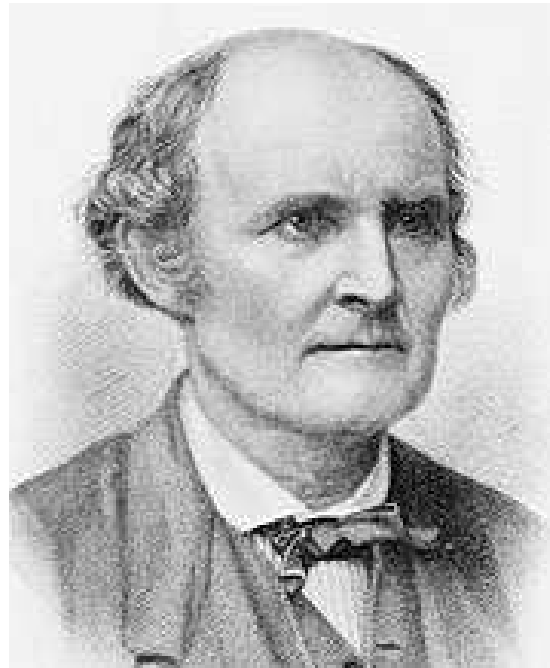


Lecture 6: The Group Inverse



The matrix index

Let $A \in \mathbb{C}^{n \times n}$, k positive integer. Then $R(A^{k+1}) \subset R(A^k)$. The **index** of A , denoted $\text{Ind } A$, is the smallest integer k such that

$$R(A^k) = R(A^{k+1}), \text{ or equivalently, } \text{rank } A^k = \text{rank } A^{k+1} \quad (1)$$

holds. Then

$$R(A^j) = R(A^{\text{Ind } A}), \quad \forall j > \text{Ind } A.$$

A matrix A is **range Hermitian** (or **EP matrix**) if

$$R(A) = R(A^*)$$

Special cases: (a) Nonsingular matrices, (b) normal matrices, in particular (c) $A = O$.

If A is range Hermitian, then $\text{Ind } A = 1$.

The converse is not true, e.g., $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

Index and complementary subspaces

Theorem. Let $A \in \mathbb{C}^{n \times n}$ have **index** k , and let ℓ be a positive integer. Then $R(A^\ell)$ and $N(A^\ell)$ are **complementary subspaces** if and only if $\ell \geq k$.

Proof. The theorem is obvious for nonsingular matrices.

Let $A \in \mathbb{C}^{n \times n}$. Then, for any positive integer ℓ ,

$$\dim R(A^\ell) + \dim N(A^\ell) = \text{rank } A^\ell + \text{nullity } A^\ell = n .$$

$$\therefore \mathbb{C}^n = R(A^\ell) \oplus N(A^\ell) \iff R(A^\ell) \cap N(A^\ell) = \{\mathbf{0}\} .$$

From

$$R(A^{\ell+1}) \subset R(A^\ell) \text{ and } N(A^\ell) \subset N(A^{\ell+1}) ,$$

it follows that

$$\mathbb{C}^n = R(A^\ell) \oplus N(A^\ell) \iff R(A^\ell) = \dim R(A^{\ell+1}) . \square$$

Nilpotent matrices

A matrix N is **nilpotent** if $N^k = O$ for some integer $k \geq 0$. The smallest such k is called the **index of nilpotency** of N .

Let $J_k(\lambda)$ be a $k \times k$ Jordan block,

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & \cdots & \cdots & 0 \\ \vdots & \lambda & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \lambda & 1 \\ 0 & \cdots & \cdots & \cdots & \lambda \end{bmatrix}$$

Then $J_k(\lambda)$ is nilpotent if $\lambda = 0$,

$$\text{ind } J_k(\lambda) = \begin{cases} 1, & \text{if } \lambda \neq 0; \\ k, & \text{if } \lambda = 0. \end{cases}$$

Jordan blocks

Let $A \in \mathbb{C}^{n \times n}$, $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix} \in \mathbb{C}_k^{n \times k}$ and $\lambda \in \mathbb{C}$ satisfy

$$AX = XJ_k(\lambda), \quad \text{where } J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{k \times k},$$

$$\therefore A\mathbf{x}_1 = \lambda \mathbf{x}_1,$$

$$A\mathbf{x}_j = \lambda \mathbf{x}_j + \mathbf{x}_{j-1}, \quad j = 2, \dots, k.$$

$$\therefore \forall j \in \overline{1, k}: (A - \lambda I)^j \mathbf{x}_j = \mathbf{0}, \quad (A - \lambda I)^{j-1} \mathbf{x}_j = \mathbf{x}_1 \neq \mathbf{0},$$

\mathbf{x}_1 is an **eigenvector** of A corresponding to λ

\mathbf{x}_j is a λ -**vector** (**principal vector**, **generalized eigenvector**)
of A of **grade** j

Powers of the matrix $(J_k(\lambda) - \lambda I) \in \mathbb{C}^{k \times k}$

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{bmatrix}, \quad J_k(\lambda) - \lambda I = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}$$

$$(J_k(\lambda) - \lambda I)^2 = \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ \vdots & & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}, \quad (J_k(\lambda) - \lambda I)^k = O.$$

The Jordan normal form

Theorem. Let $A \in \mathbb{C}^{n \times n}$. Then A is similar to a block diagonal matrix J with Jordan blocks on its diagonal, i.e. \exists nonsingular $X \ni$

$$X^{-1}AX = J = \begin{bmatrix} J_{k_1}(\lambda_1) & O & \cdots & O \\ O & J_{k_2}(\lambda_2) & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & J_{k_p}(\lambda_p) \end{bmatrix}. \quad (1)$$

The matrix J , called the **Jordan normal form** of A , is unique up to a rearrangement of its blocks. \square

The scalars $\{\lambda_1, \dots, \lambda_p\}$ in (1) are the eigenvalues of A . The set of eigenvalues, or **spectrum** of A , is denoted $\Lambda(A)$.

For A as above,

$$(A - \lambda_1 I)^{k_1} (A - \lambda_2 I)^{k_2} \cdots (A - \lambda_p I)^{k_p} = O. \quad (2)$$

Vanishing polynomials

The polynomial

$$c(z) = (z - \lambda_1)^{k_1} (z - \lambda_2)^{k_2} \cdots (z - \lambda_p)^{k_p}$$

is the **characteristic polynomial** of A . An eigenvalue λ may be repeated, say

$$c(z) = p(z)(z - \lambda)^{k_1} (z - \lambda)^{k_2} \cdots (z - \lambda)^{k_m}, \quad p(\lambda) \neq 0$$

The **algebraic multiplicity** of the eigenvalue λ is then

$$k_1 + k_2 + \cdots + k_m$$

and $\max \{k_i : i \in \overline{1, m}\}$ is the **geometric multiplicity**, or **index** of the eigenvalue λ , denoted $\nu(\lambda)$. The polynomial

$$m(z) = \prod_{\lambda \in \Lambda(A)} (z - \lambda)^{\nu(\lambda)}$$

is the **minimal polynomial** of A .

Spectral inverses

Let $A \in \mathbb{C}^{n \times n}$. Then X is an S -inverse of A if they share the property that, for every $\lambda \in \mathbb{C}$ and every vector \mathbf{x} ,

$$\left\{ \begin{array}{l} \mathbf{x} \text{ is a } \lambda\text{-vector of } A \\ \text{of grade } p \end{array} \right\} \iff \left\{ \begin{array}{l} \mathbf{x} \text{ is a } \lambda^\dagger\text{-vector of } X \\ \text{of grade } p \end{array} \right\}$$

where $\lambda^\dagger = 0$ if $\lambda = 0$, and otherwise $1/\lambda$.

Ex. If $A \in \mathbb{C}^{n \times n}$ is **diagonable**,

$$A = PJP^{-1}, \quad J = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

then

$$X = PJ^\dagger P^{-1}, \quad J^\dagger = \text{diag}(\lambda_1^\dagger, \lambda_2^\dagger, \dots, \lambda_n^\dagger)$$

is an S -inverse of A . It is a $\{1, 2\}$ -inverse, and **commutes** with A ,

$$AX = XA.$$

The group inverse

Theorem. Let $A \in \mathbb{C}^{n \times n}$. Then the system

$$AXA = A, \quad (1)$$

$$XAX = X, \quad (2)$$

$$AX = XA, \quad (5)$$

has a solution X iff

$$\text{Ind } A = 1$$

and the solution is unique. It is called the **group inverse**, or **{1, 2, 5}–inverse**, of A , and is denoted $A^\#$.

Proof.

$$X \in A\{1, 2\} \iff AX = P_{R(A), N(X)}, \quad XA = P_{R(X), N(A)}$$

$$\therefore X \in A\{1, 2, 5\} \iff AX = P_{R(A), N(A)}$$

$$\mathbb{C}^n = R(A) \oplus N(A) \iff \text{Ind } A = 1 \quad \square$$

The group inverse (cont'd)

Theorem. A is range-Hermitian if, and only if, $A^\# = A^\dagger$.

Proof. $A^\dagger = A_{R(A), N(A)}^{(1,2)}$, $A^\# = A_{R(A^*), N(A^*)}^{(1,2)}$. □

Theorem (Erdélyi). Let A have index 1 and Jordan form

$$A = PJP^{-1},$$

Then

$$A^\# = PJ^\dagger P^{-1}.$$

Proof. The relations (1) $AXA = A$, (2) $XAX = X$, and (5) $AX = XA$, are similarity invariants. Therefore

$$J^\# = P^{-1}A^\#P$$

and since $\text{Ind } J = 1$,

$$J^\# = J^\dagger. \quad \square$$

The group inverse (cont'd)

Theorem (Cline). Let a square matrix A have the FRF

$$A = CR .$$

Then A has group inverse if and only if RC is nonsingular, and

$$A^\# = C(RC)^{-2}R .$$

Ex. Let $A \in \mathbb{C}^{n \times n}$. Then A has **index 1** if and only if the **limit**

$$\lim_{\lambda \rightarrow 0} (\lambda I_n + A)^{-1} A$$

exists, in which case

$$\lim_{\lambda \rightarrow 0} (\lambda I_n + A)^{-1} A = AA^\# .$$

Proof. A full rank factorization $A = CR$ gives

$$(\lambda I_n + A)^{-1} A = C(\lambda I_r + RC)^{-1} R , \text{ etc.}$$

Properties of the group inverse

- (a) If A is nonsingular, $A^\# = A^{-1}$.
- (b) $A^{\#\#} = A$.
- (c) $A^{*\#} = A^{\#*}$.
- (d) $A^{T\#} = A^{\#T}$.
- (e) $(A^\ell)^\# = (A^\#)^\ell$ for every positive integer ℓ .
- (f) Let A have index 1 and denote

$$A^{-j} := (A^\#)^j, \quad \forall j = 1, 2, \dots,$$
$$A^0 := AA^\#.$$

Then,

$$A^\ell A^m = A^{\ell+m}, \quad \forall \ell, m.$$

The “powers” of A , positive, negative and zero, constitute an **Abelian group** under matrix multiplication (\therefore “group inverse”)

Lemma on λ -vectors

Lemma 1. Let \mathbf{x} be a λ -vector of A with $\lambda \neq 0$. Then $\mathbf{x} \in R(A^\ell)$ where ℓ is an arbitrary positive integer.

Proof. Let $(A - \lambda I)^p \mathbf{x} = \mathbf{0}$ for some positive integer p , or, using the binomial expansion,

$$\mathbf{x} = c_1 A \mathbf{x} + c_2 A^2 \mathbf{x} + \cdots + c_p A^p \mathbf{x}, \quad c_i = (-1)^{i-1} \lambda^{-i} \binom{p}{i}. \quad (1)$$

$$\begin{aligned} \therefore A \mathbf{x} &= c_1 A^2 \mathbf{x} + c_2 A^3 \mathbf{x} + \cdots + c_p A^{p+1} \mathbf{x}, \\ A^2 \mathbf{x} &= c_1 A^3 \mathbf{x} + c_2 A^4 \mathbf{x} + \cdots + c_p A^{p+2} \mathbf{x}, \\ \dots &= \dots \end{aligned} \quad (2)$$

$$A^{\ell-1} \mathbf{x} = c_1 A^\ell \mathbf{x} + c_2 A^{\ell+1} \mathbf{x} + \cdots + c_p A^{p+\ell-1} \mathbf{x},$$

Successive substitutions of (2) in (1) give

$$\mathbf{x} = A^\ell q(A) \mathbf{x}, \quad \text{where } q \text{ is some polynomial.}$$

Lemma on spectral inverses

Lemma 2. Let A be a square matrix and let

$$XA^{\ell+1} = A^{\ell} \quad (1)$$

for some positive integer ℓ . Then every λ -vector of A of grade p for $\lambda \neq 0$ is a λ^{-1} -vector of X of grade p .

Proof by induction on the grade p .

$p = 1$. Let $A\mathbf{x} = \lambda\mathbf{x}$, $\lambda \neq 0$. Then

$$A^{\ell+1}\mathbf{x} = \lambda^{\ell+1}\mathbf{x}$$

$$\therefore \mathbf{x} = \lambda^{-\ell-1}A^{\ell+1}\mathbf{x}$$

$$\therefore X\mathbf{x} = \lambda^{-\ell-1}XA^{\ell+1}\mathbf{x} = \lambda^{-1}\mathbf{x}.$$

Assume true for $p \in \overline{1, r}$. Let \mathbf{x} be a λ -vector of A of grade $r + 1$.

Lemma 2 (cont'd)

Then, by Lemma 1, $\mathbf{x} = A^\ell \mathbf{y}$, for some \mathbf{y} .

$$\begin{aligned} \therefore (X - \lambda^{-1}I) \mathbf{x} &= (X - \lambda^{-1}I) A^\ell \mathbf{y} = X(A^\ell - \lambda^{-1}A^{\ell+1}) \mathbf{y} \\ &= X(I - \lambda^{-1}A) A^\ell \mathbf{y} = -\lambda^{-1}X(A - \lambda I) \mathbf{x} . \end{aligned}$$

By the induction hypothesis, $(A - \lambda I) \mathbf{x}$ is a λ^{-1} -vector of X of grade r .

$$\begin{aligned} \therefore (X - \lambda^{-1}I)^r (A - \lambda I) \mathbf{x} &= \mathbf{0} , \\ \mathbf{z} &= (X - \lambda^{-1}I)^{r-1} (A - \lambda I) \mathbf{x} \neq \mathbf{0} \\ X\mathbf{z} &= \lambda^{-1}\mathbf{z} . \end{aligned}$$

$$\begin{aligned} \therefore (X - \lambda^{-1}I)^{r+1} \mathbf{x} &= -\lambda^{-1}X(X - \lambda^{-1}I)^r (A - \lambda I) \mathbf{x} = \mathbf{0} , \\ (X - \lambda^{-1}I)^r \mathbf{x} &= -\lambda^{-1}X\mathbf{z} = -\lambda^{-2}\mathbf{z} \neq \mathbf{0} . \end{aligned}$$

$\therefore \mathbf{x}$ is a λ^{-1} -vector of X of grade $r + 1$. \square

The group inverse is spectral

Recall: Let $A \in \mathbb{C}^{n \times n}$. Then X is an S -inverse of A if they share the property that, for every $\lambda \in \mathbb{C}$ and every vector \mathbf{x} ,

$$\left\{ \begin{array}{l} \mathbf{x} \text{ is a } \lambda\text{-vector of } A \\ \text{of grade } p \end{array} \right\} \iff \left\{ \begin{array}{l} \mathbf{x} \text{ is a } \lambda^\dagger\text{-vector of } X \\ \text{of grade } p \end{array} \right\}$$

where $\lambda^\dagger = 0$ if $\lambda = 0$, and otherwise $1/\lambda$.

Theorem. Let $A \in \mathbb{C}^{n \times n}$ have index 1. Then $A^\#$ is the unique S -inverse of A in $A\{1\} \cup A\{2\}$. If A is diagonalizable, $A^\#$ is the only S -inverse of A .

Proof. $A^\#$ is an S -inverse of A . Since $X = A^\#$ satisfies $XA^{\ell+1} = A^\ell$ with $\ell = 1$, it follows from Lemma 2 that $A^\#$ satisfies the \implies part of the definition of S -inverse for $\lambda \neq 0$. Replacing A by $A^\#$ establishes the \impliedby part for $\lambda \neq 0$, since $A^{\#\#} = A$.

The rest of the proof omitted.

Application to finite Markov chains

A **system** is observed at times $t = 0, 1, 2, \dots$. The system has N **states**, denoted $\{1, 2, \dots, N\}$, the **state at time** t is denoted \mathbf{X}_t

The system is a **(finite) Markov chain** (or **chain**) if \exists a matrix $P = (p_{ij}) \in \mathbb{R}^{N \times N}$ such that

$$\text{Prob} \{ \mathbf{X}_{t+1} = j \mid \mathbf{X}_t = i \} = p_{ij} , \quad \forall i, j \in \overline{1, N} , \quad \forall t = 0, 1, \dots .$$

The numbers p_{ij} are called the **transition probabilities**, and the matrix P is called the **transition matrix**.

$$\therefore \sum_{j=1}^N p_{ij} = 1 , \quad i \in \overline{1, N} , \quad (\text{a})$$

$$p_{ij} \geq 0 , \quad i, j \in \overline{1, N} . \quad (\text{b})$$

A square matrix $P = [p_{ij}]$ satisfying (a),(b) is called **stochastic**.

Condition (a) is $P\mathbf{e} = \mathbf{e}$, i.e., $\mathbf{1} \in \Lambda(P)$.

Markov chains (cont'd)

Let $p_{ij}^{(n)}$ denote the n -step transition probability

$$p_{ij}^{(n)} = \text{Prob} \{ \mathbf{X}_{t+n} = j \mid \mathbf{X}_t = i \} , \quad p_{ij}^{(0)} := \delta_{ij} .$$

The **Chapman–Kolmogorov equations**

$$p_{ij}^{(m+n)} = \sum_{k=1}^N p_{ik}^{(m)} p_{kj}^{(n)} , \quad \forall m, n \in \mathbb{Z}_+ . \quad (1)$$

$\therefore p_{ij}^{(n)}$ is the (i, j) -th element of P^n .

More terminology:

(a) A state i **leads** to state j , denoted by $i \rightsquigarrow j$, if $p_{ij}^{(n)} > 0$ for some $n \geq 1$.

(b) Two states i, j **communicate**, denoted by $i \longleftrightarrow j$ if each state leads to the other.

(c) A set S of states is **closed** if $i, j \in S \implies i \longleftrightarrow j$

Terminology (cont'd)

(d) A chain is **irreducible** if its only closed set is the set of all states $\overline{1, N}$, and is **reducible** otherwise.

(e) A single state forming a closed set is an **absorbing state**. A reducible chain is **absorbing** if each of its closed sets consists of a single (necessarily absorbing) state.

(f) A chain is **regular** if for some k , P^k is a positive matrix.

(g) A state i has **period** τ if $p_{ii}^{(n)} = 0$ except when $n = \tau, 2\tau, 3\tau, \dots$. The period of i is denoted $\tau(i)$. If $\tau(i) = 1$, i is **aperiodic**.

(h) Let $f_{ij}^{(n)}$ = **probability that starting from i , state j is reached for 1st time at step n** , and let

$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ **probability that j eventually reached from i ,**

$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$ **the mean 1st passage time from i to j .**

Terminology (cont'd)

- (h) $\mu_{kk} := \sum_{n=1}^{\infty} n f_{kk}^{(n)}$ is the **mean return time** of state k .
- (i) A state i is **recurrent** if $f_{ii} = 1$, and is **transient** otherwise. A chain is **recurrent** if all its states are recurrent.
- (j) If i is recurrent and $\mu_{ii} = \infty$, i is called a **null state**. All states in a closed set are of the same type.
- (k) A state is **ergodic** if **recurrent** and **aperiodic**, but not a **null state**. If a chain is irreducible, all its states have the same period. An irreducible aperiodic chain is called **ergodic**.

Theorem (Chung). For any states i, j of an irreducible recurrent chain

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n p_{ij}^{(k)} = \frac{1}{\mu_{jj}}. \quad (1)$$

Proof. Both sides of (1) give the expected number of returns to state j in unit time.

Stationary probabilities

The probabilities $\{\pi_1, \pi_2, \dots, \pi_N\}$ are **stationary** if

$$\pi_k = \sum_{i=1}^N p_{ik} \pi_i, \quad k \in \overline{1, N},$$

or, in terms of vector of stationary probabilities,

$$P^T \boldsymbol{\pi} = \boldsymbol{\pi}.$$

$\boldsymbol{\pi}$ is called a **stationary distribution** or **steady state**. In an irreducible recurrent chain, the stationary distribution is simply,

$$\pi_k = \frac{1}{\mu_{kk}}, \quad k \in \overline{1, N}.$$

If a chain is **ergodic**, the system converges to its stationary distribution from **any** initial distribution.

In terms of P , every row of P^n converges to $\boldsymbol{\pi}^T$.

Summary for ergodic chains

Theorem (Feller). Let $P \in \mathbb{R}^{N \times N}$ be the transition matrix of an ergodic chain. Then:

(a) $\forall j, k \in \overline{1, N}$ the following limit exists, independent of j ,

$$\lim_{n \rightarrow \infty} p_{jk}^{(n)} = \pi_k > 0 \quad (1)$$

(b) $\pi_k = \frac{1}{\mu_{kk}}$

(c) The numbers $\{\pi_k : k \in \overline{1, N}\}$ are probabilities,

$$\pi_k \geq 0, \quad \sum_{k=1}^N \pi_k = 1, \quad (2)$$

and are a stationary distribution of the chain:

$$\pi_k = \sum_{i=1}^N p_{ik} \pi_i, \quad k \in \overline{1, N}. \quad (3)$$

The distribution $\{\pi_k\}$ is uniquely determined by (2) and (3). \square

Solution using the group inverse

Theorem (Meyer). Let P be the transition matrix of a **finite Markov chain**, and let $Q = I - P$.

Then Q has a **group inverse**, and:

$$I - QQ^\# = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n P^k, \quad (1)$$

$$= \lim_{n \rightarrow \infty} (\alpha I + (1 - \alpha)P)^n, \text{ for any } 0 < \alpha < 1, \quad (2)$$

and if the chain is **regular**, or **absorbing**,

$$I - QQ^\# = \lim_{n \rightarrow \infty} P^n. \quad (3)$$

For proofs and other details see original work.