

# Lecture XIII

## Nonnegative Matrices

### Key Points:

- Markov matrices
- Stochastic and doubly stochastic matrices
- Theorem of Perron on convergence
- Perron-Frobenius theory
- Examples and applications

# Nonnegative and Positive Matrices

A matrix  $M = (m_{ij})$  is said to be **nonnegative**, if  $m_{ij} \geq 0$  for all  $i, j$ .

It is said to be **positive**, if  $m_{ij} > 0$  for all  $i, j$ .

# Comments

Positive matrix  $\neq$  Positive-definite matrix

Nonnegative matrix  $\neq$  Nonnegative definite matrix (also known as positive semi-definite matrix)

# A Motivating Example

Consider a particle taking values from the set  $\{1, 2, \dots, N\}$  and moving at discrete points in time  $n = 0, 1, 2, \dots$ .

Let  $M = (m_{ij})$  be the (presumably time-invariant) **transition matrix**,

with  $m_{ij}$  the probability the particle jumping from state  $j$  at time  $n$  to state  $i$  at time  $n + 1$ .

# A Motivating Example (cont'd)

Such a stochastic process is usually called a **discrete Markov process**.

$M = (m_{ij})$  is a (nonnegative) **Markov matrix** satisfying the following conditions:

i)  $m_{ij} \geq 0$ ;

ii)  $\sum_{i=1}^N m_{ij} = 1, \forall j = 1, 2, \dots, N.$



## A Motivating Example (cont'd)

Let  $x_i(n)$  be the probability the particle is in state  $i$  at time  $n$ . Then, the following relations hold:

$$x_i(n+1) = \sum_{j=1}^N m_{ij} x_j(n), \quad 1 \leq i \leq N$$

or, in compact matrix notation,

$$x(n+1) = Mx(n), \quad n = 0, 1, 2, \dots \quad x \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

# *Fundamental Question:*

What is the limiting behavior of  $x(n)$ ?

*Remark:* For any  $n$ ,  $x(n)$  is a **probability vector**, i.e., all components  $x_i(n)$  are nonnegative,

and satisfy  $\sum_{i=1}^N x_i(n) = 1$ .

# Question 1: Markov matrices

What is the range of parameters  $\lambda$  so that a linear combination  $\lambda P + (1 - \lambda)Q$  of two Markov matrices  $P, Q$  remains to be Markov?



## Question 2: Probability Vector

Assume  $M$  is a Markov matrix and  $x$  is a probability vector. Is  $Mx$  a probability vector? Why?

# Remarkable Result

For any *positive* Markov matrix  $M$  and any probability vector  $x(0)$ , the solution  $x(n)$

to  $x(n+1) = Mx(n)$

settles at a fixed probability vector  $y$ , that is,

$$\lim_{n \rightarrow \infty} x(n) = y$$

In addition,  $y$  is independent of  $x(0)$ , and is an eigenvector of  $M$  associated with eigenvalue 1.

## Comment 1

It should be noted that this result does not hold, if  $M$  is only a nonnegative matrix, but not positive. A counter-example is

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Indeed, in this case,  $M^n x = x$  is dependent on initial condition  $x$ !

## Comment 2

**This type of matrix-theoretic results has been applied recently to address engineering and bio-problems, such as**

- Coordination and control of groups of robots
- Consensus in biological multi-agents: bird flocking, fish schooling ...

**See: “Jadbabaie, Lin and Morse, IEEE Transactions on Automat. Control, 2003” and many references on distributed control and computation.**

# Sketch of Proof

Noting that

$$\langle x(n), b \rangle = \langle M^n x(0), b \rangle = \langle x(0), (M^T)^n b \rangle,$$

it suffices to show that, for any  $b$ ,

$$(M^T)^n b \rightarrow b_l, \text{ as } n \rightarrow \infty.$$

With this in mind, consider

$$z(n) = (M^T)^n b$$

obeying the difference equation

$$z(n+1) = M^T z(n), \text{ with } z(0) = b.$$



## Sketch of Proof (cont'd)

Let  $u(n)$  be the largest component of  $z(n)$ , and  $v(n)$  the smallest component of  $z(n)$ .

Now, we only need to prove that

$$u(n) - v(n) \rightarrow 0.$$

Since

$$z_i(n+1) = \sum_{j=1}^N m_{ji} z_j(n),$$

with  $\sum_{j=1}^N m_{ji} = 1, m_{ji} > 0$ , we have

## Sketch of Proof (cont'd)

we have

$$u(n+1) \leq u(n) \text{ and } v(n+1) \geq v(n).$$

So, the decreasing sequence  $u(n)$  and the increasing sequence  $v(n)$  *both* converge:

$$u(n) \rightarrow u, \quad v(n) \rightarrow v.$$

Next, we need to show that  $u = v$ .

## Sketch of Proof (cont'd)

Let  $d$  be the positive lower bound for  $m_{ij}$ .

Using the component form of  $z(n+1) = M^T z(n)$ , it follows that

$$(*) \begin{cases} u(n+1) \leq (1-d)u(n) + dv(n), \\ v(n+1) \geq (1-d)v(n) + du(n). \end{cases}$$

$\Rightarrow$

$$u(n+1) - v(n+1) \leq (1-2d)(u(n) - v(n))$$

## **Remark:** Detailed derivations of (\*)

To illustrate the idea, let's first look at the Case of  $N = 2$ . It suffices to prove the following

$$c = \alpha a + \beta b \leq (1-d)a + db$$

for any  $a \geq b \geq 0$  and  $\alpha + \beta = 1$ .

Of course, the above is equivalent to

$$0 \leq (1-\alpha-d)a - (\beta-d)b \Leftrightarrow (\beta-d)(a-b) \geq 0.$$

The latter clearly holds, because  $d \leq \beta$ .

To prove the general case  $N \geq 2$ , note that we can assume that  $v = u_N$ . Also note that

$$\sum_{j=1}^N m_{ji} u_j \leq (1-d)u + dv$$

holds if

$$0 \leq \left( 1 - \sum_{j=1}^{N-1} m_{ji} - d \right) u - (m_{Ni} - d) v$$

$$\Leftrightarrow (m_{Ni} - d)(u - v) \geq 0.$$

The latter clearly holds, because  $d \leq m_{Ni}$ .



## Sketch of Proof (cont'd)

Since  $d \leq 0.5$  when  $N \geq 2$ ,

$$u(n+1) - v(n+1) \leq (1 - 2d)(u(n) - v(n))$$

implies that  $u(n) - v(n) \rightarrow 0$ , as  $n \rightarrow \infty$ .

As a result,  $z(n)$  converges to a vector  $z$  with the components being **equal**. That is,

$$z = (a_1, \dots, a_1)^T. \text{ So, } x(n) \rightarrow y.$$

Using  $\langle x(n), b \rangle = \langle x(0), (M^T)^n b \rangle$ , and letting  $c = x(0)$ ,  
 $\langle y, b \rangle = \langle c, z \rangle = a_1 (c_1 + \dots + c_N) = a_1.$

Since  $a_1$  depends only on  $b$ ,  $y$  is independent of  $c$ !

# Sketch of Proof (end)

Finally, we need to show that  $y$  is an eigenvector of  $M$  associated with eigenvalue 1. Indeed,

$$y = \lim_{n \rightarrow \infty} M^{n+1}c = M \lim_{n \rightarrow \infty} M^n c = My.$$

*Note that  $y > 0$ .*

# Property 1 of Markov Matrices

For any eigenvalue  $\lambda$  of a Markov matrix  $M$ ,  
 $|\lambda| \leq 1$ .

# Proof

Since  $M^T$  shares the same eigenvalues with  $M$ , let us take an eigenvector  $x$  of  $M^T$  associated with the eigenvalue  $\lambda$ . That is,

$$(*) \quad \lambda x = M^T x.$$

Let  $m$  be the absolute value of a component of  $x$  of greatest magnitude, then using  $l_\infty$ -norm,

(\*) implies:

$$|\lambda| m \leq m \sum_{j=1}^N m_{ji} = m \Rightarrow |\lambda| \leq 1.$$

## ***Remark (Jie Du, 2013.12.10)***

- We can prove it using the Gersgorin disk Theorem.



# Property 2 of Markov Matrices

If  $M$  is a *positive* Markov matrix, then  $\lambda = 1$  is the only eigenvalue of absolute value one.

# Proof

Assume that  $\mu$  is another eigenvalue with  $|\mu| = 1$ , and  $\omega + jz$  an associated eigenvector.

Choose a big enough  $t_1$  to make  $\omega + t_1 y$  and  $z + t_1 y$  both positive vectors, with  $y$  the limit of  $M^n x(0)$ .

It can be directly checked that

$$M(\omega + jz + t_1(1+j)y) = \mu(\omega + jz) + t_1(1+j)y.$$

So,

$$M^n(\omega + jz + t_1(1+j)y) = \mu^n(\omega + jz) + t_1(1+j)y.$$

## Proof (Cont'd)

On the other hand, as  $n \rightarrow \infty$ ,

$$M^n(\omega + jz + t_1(1+j)y) = M^n((\omega + t_1y) + j(z + t_1y)).$$

As shown previously,  $M^n(\omega + t_1y)$  and  $M^n(z + t_1y)$  converge to a scalar multiple of ones, resp.

However,  $\mu^n(\omega + jz)$  converges only when  $\mu = 1$ ,  
*under the constraint*  $|\mu| = 1$ .

# Exercise

For any positive Markov matrix, if  $\lambda$  is an eigenvalue with an associated *positive* eigenvector, then  $\lambda = 1$ .



# Detailed Solution

By hypothesis,  $Mx = \lambda x$ , with  $M$  a positive Markov matrix and  $x$  a positive eigenvector.

Clearly,  $\lambda$  can only be a positive real eigenvalue.

Without loss of generality, we can assume that  $x$  is a positive probability vector.

By the remarkable result,  $M^n x = \lambda^n x$  must converge to a probability vector. If  $\lambda < 1$ , then a contradiction occurs.



## Another Proof (by Matt)

Let  $x = (x_1, x_2, \dots, x_n)^T$  be the positive eigenvector associated with  $\lambda$ . Then,  $Mx = \lambda x$  implies:

$$m_{i1}x_1 + \dots + m_{in}x_n = \lambda x_i, \quad \forall i = 1, 2, \dots, n.$$

Summing up these equations, and using the fact that  $M$  is a Markov matrix, it holds:

$$x_1 + \dots + x_n = \lambda (x_1 + \dots + x_n)$$

$$\Rightarrow \lambda = 1.$$

**Note:** indeed, we only need to assume

$$x_1 + \dots + x_n \neq 0.$$

# Positive Matrices

The following linear equations often occur as a simple model for the growth of a set of biological objects:

$$x_i(n+1) = \sum_{j=1}^N a_{ij} x_j(n), \quad i = 1, 2, \dots, N.$$

Or, in compact matrix notation,

$$x(n+1) = Ax(n),$$

where  $A = (a_{ij})_{N \times N}$  is a **positive** matrix, i.e.  $a_{ij} > 0$ .

# ***Problem***

Determine the behavior of  $x_i[n]$  as  $n \rightarrow \infty$ ?

# Theorem of Perron (1907)

- (1) If  $A$  is a positive matrix, there is a unique eigenvalue of  $A$ , denoted as  $\lambda_{\max}(A)$ , which has the greatest absolute value.
- (2) This eigenvalue  $\lambda_{\max}(A)$  is positive and simple, and its associated eigenvector may be taken positive.

***Proof. See the textbook.***

# Application: A Limit Theorem

Let  $c \neq 0$  be any nonnegative vector. Then,

$$v = \lim_{n \rightarrow \infty} A^n c / \lambda_{\max}(A)^n$$

exists and is an eigenvector of  $A$  associated with  $\lambda_{\max}(A)$ , unique up to a scalar multiple determined by the choice of  $c$ , but otherwise independent of the initial state  $c$ .

***See the textbook.***



As a result,

the solution of  $x(n+1) = Ax(n)$ , with  $A = (a_{ij})$   
a positive matrix, asymptotically looks like

$$x(n) \sim \lambda^n \gamma, \quad \lambda = \lambda_{\max}(A): \text{Perron root}$$

where  $\gamma$  is a (positive) eigenvector associated  
with  $\lambda$ , or a positive multiple of a special  
normalized eigenvector.

For the population example described by

$$x(n+1) = Ax(n),$$

the above result implies that:

regardless of the initial population, we will approach a steady-state situation where the total population grows exponentially, but the proportions of the total various species remain constant.

# Continuous Growth Processes

Starting with a discrete-time process with a **small** time interval  $\Delta$ , then

$$x_i(t + \Delta) = (1 + a_{ii}\Delta) x_i(t) + \Delta \sum_{j \neq i} a_{ij} x_j(t)$$

$i = 1, 2, \dots, N$ ;  $a_{ij}$  = rates of production.

Letting  $\Delta \rightarrow 0$ , we have

$$\dot{x}_i = a_{ii}x_i + \sum_{j \neq i} a_{ij}x_j, \quad 1 \leq i \leq N.$$

# Continuous Version of Perron's Theorem

If  $a_{ij} > 0$ ,  $i \neq j$ , then the eigenvalue of  $A$  with largest real part, denoted as  $\rho(A)$ , is real and simple. There is an associated positive eigenvector which is unique up to a multiplicative constant.

*Note:* The asymptotic behavior of  $x_i(t)$ , as  $t \rightarrow \infty$ , is determined by  $\rho(A)$ .



# Exercise from Mathematical Economics

(K.D. Arrow & A.C. Enthoven, 1956)

If  $A$  has all negative diagonal elements, and no negative off-diagonal elements, if  $D$  is a diagonal matrix, and if the real parts of the eigenvalues of both  $A$  and  $DA$  are negative, then the diagonal elements of  $D$  are positive.



# Other Notions and Extensions

- **Irreducible matrix**
- **Perron-Frobenius theory**
- **Stochastic and doubly stochastic matrices**

# Irreducible Matrix

A nonnegative matrix  $A \in \mathbb{R}_+^{n \times n}$  is said to be **irreducible**, if for every pair  $(i, j)$ ,  $\exists k \geq 1$  such that the  $(i, j)$  entry of  $A^k$  is positive.

## Example 1: Positive (Markov) matrices

**Example 2:** Any matrix  $A \in \mathbb{R}_+^{n \times n}$  with the property that  $A^k > 0$  for some  $k \geq 1$ . Such a matrix is known as "primitive matrix".

**Is the converse true?**

**Remark:** Irreducible Matrices may not be primitive

## A Counter-Example:

The following matrix

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

is a (nonnegative) irreducible matrix, but

$A^k$  is **not** a positive matrix for *any*  $k$ .

(Do you know why?)

# Comment (*Primitive matrices*)

**Definition:** A matrix  $A \in \mathbb{R}_+^{n \times n}$  is **primitive**, if it is irreducible and has only one nonzero eigenvalue of maximum modulus.

## Theorem

For any matrix  $A \in \mathbb{R}_+^{n \times n}$ ,  $A$  is **primitive**, if and only if  $A^m$  is a positive matrix for some  $m \geq 1$ .

(See the text of Horn-Johnson, 2013, page 540, for a proof.)

# Examples

- Show that the matrix

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

is irreducible.

- How about an upper-triangular matrix

like  $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ ?



# Interesting Result

Let  $A \in \mathbb{R}^{n \times n}$  be a nonnegative irreducible matrix.

Then,  $(I_n + A)^{n-1}$  is a positive matrix.

**Proof:** follows from the following identity

$$(I_n + A)^{n-1} = \sum_{k=1}^{n-1} \binom{n-1}{k} A^{n-1-k}$$

and Caley-Hamilton Theorem.

# Exercise

For any nonnegative matrix  $A \in \mathbb{R}^{n \times n}$  and a diagonal matrix  $D \in \mathbb{R}^{n \times n}$  with positive entries in the diagonal.

The following statements are equivalent:

- 1)  $A$  is irreducible;
- 2)  $AD$  is irreducible;
- 3)  $DA$  is irreducible.

# Perron-Frobenius Theory

For any nonnegative irreducible matrix  $A \in \mathbb{R}^{n \times n}$ , it holds:

- 1) the spectral radius  $\rho(A) \triangleq \max \{|\lambda| : \lambda \in \sigma(A)\}$  is an eigenvalue of  $A$ .
- 2)  $\exists u \in \mathbb{R}_{>0}^n$ , such that  $Au = \rho(A)u$ .
- 3) The algebraic multiplicity of  $\rho(A)$  is one.

# Stochastic and Doubly Stochastic Matrices

Consider a nonnegative matrix  $P = (p_{ij}) \in \mathbb{R}_+^{n \times n}$ .

1) It is **stochastic**, if  $\sum_{j=1}^n p_{ij} = 1, \forall i$ .

2) It is **doubly stochastic**, if both  $P$  and  $P^T$  are stochastic.



# Theorem of Birkhoff-von Neumann

Any doubly stochastic matrix  $P \in \mathbb{R}^{n \times n}$  is a convex combination of finitely many permutation matrices. That is,

$$P = \sum_{i=1}^m \lambda_i P_i, \quad \text{with } \lambda_i \geq 0, \sum_i \lambda_i = 1,$$

where  $P_i$  is a permutation matrix derived from  $I_n$  after interchanging some of the rows.