

## A GEOMETRIC PROOF OF MARKOV ERGODIC THEOREM

R. Z. YEH

ABSTRACT. A geometric approach combined with topological results leads to a criterion for ergodic stability of Markov transformations. The matrix representation of this criterion provides an alternative proof for the well-known theorem of Markov in probability.

Given a finite Markov chain with an  $n$ -by- $n$  transition matrix  $P$ , if for some positive integer  $k$   $P^k$  is positive in the sense that all entries in  $P^k$  are positive, then regardless of how the initial probability distribution  $p_0$  among the  $n$  states is, the subsequent probability distributions  $p_i$  tend to a fixed probability distribution  $q$  as  $i$  tends to infinity; furthermore  $q$  is positive in the sense that all  $n$  components of  $q$  are positive. This well-known result is originally due to Markov [1]. The purpose of this paper is to present a geometric proof, in which we deal with the totality of probability distributions rather than with individual states. Our starting point is to interpret  $P$  as an affine transformation of a certain *stochastic simplex*; we then proceed geometrically until at the very end when we reinterpret our main result (Theorem 1) in terms of  $P$  to obtain Markov's theorem (Corollary 1). The geometric approach also provides a simple classification of finite Markov chains (Definition 3).

DEFINITION 1. By a stochastic  $n$ -simplex  $\Delta^n$  for  $n = 1, 2, 3, \dots$  we mean the subset of an  $n$ -dimensional Euclidean space  $R^n$  given by

$$\Delta^n = \left\{ \sum_{i=1}^n x_i \mathbf{u}_i; 0 \leq x_i, \sum_i x_i = 1 \right\}$$

where  $\mathbf{u}_i$  are the orthonormal basis vectors of  $R^n$ .

Clearly each element of  $\Delta^n$  represents a possible initial probability distribution of a Markov chain with  $n$  states.

Geometrically  $\Delta^n$  has dimension  $n - 1$ , being a subset of the  $(n - 1)$ -dimensional hyperplane

$$H^{n-1} = \left\{ \sum_{i=1}^n x_i \mathbf{u}_i; \sum_i x_i = 1 \right\}$$

---

Received by the editors October 27, 1969.

AMS 1969 subject classifications. Primary 6065, 1565; Secondary 5025, 5210, 5485.

Key words and phrases. Stochastic simplex, Markov transformation, sequence of convex sets, permutation of vertices, boundary, interior, fixed points, cyclic points, acyclic transformation, irreducible transformation, ergodic stability, transition matrix, hyperplane.

contained in  $R^n$ . The boundary  $\partial\Delta^n$  of  $\Delta^n$  as a subset of  $H^{n-1}$  is the union of  $n$  stochastic  $(n-1)$ -simplices:

$$\Delta_r^{n-1} = \left\{ \sum_{i=1}^n x_i \mathbf{u}_i : 0 \leq x_i, \sum_i x_i = 1, x_r = 0 \right\}$$

for  $r=1, 2, \dots, n$ . Each of these stochastic  $(n-1)$ -simplices as a subset of an  $(n-2)$ -dimensional hyperplane has in turn its boundary consisting of  $n-1$  stochastic  $(n-2)$ -simplices, and so on. All these simplices, from  $\Delta^n$  itself on down to the  $n$  stochastic 1-simplices, shall be referred to as stochastic subsimplices of  $\Delta^n$ . The interior of each stochastic subsimplex of  $\Delta^n$  is what remains after its boundary is removed from it. In particular, the interior of  $\Delta^n$  is given by

$$\Delta^n - \partial\Delta^n = \left\{ \sum_{i=1}^n x_i \mathbf{u}_i : 0 < x_i, \sum_i x_i = 1 \right\}.$$

It is important to note that each point of  $\Delta^n$  is an interior point of exactly one stochastic subsimplex of  $\Delta^n$ . We hasten to add that a stochastic 1-simplex shall be regarded as having empty boundary so that its interior is itself.

**DEFINITION 2.** By a Markov transformation  $f$  we mean a linear transformation of  $R^n$  into  $R^n$  which maps  $\Delta^n$  into  $\Delta^n$ .

Given an  $n$ -by- $n$  transition matrix  $P$ , the linear transformation  $f$  of  $R^n$  into  $R^n$  determined by  $f(\mathbf{u}_i) = \sum_j p_{ij} \mathbf{u}_j$  where  $p_{ij}$  is the entry in the  $i$ th row and  $j$ th column of  $P$  is clearly a Markov transformation. Conversely, given a Markov transformation  $f$  its matrix representation  $P$  is a transition matrix.

Since  $\Delta^n$  (henceforth denoted  $\Delta$  whenever convenient) represents all possible initial probability distributions, the image sets  $f(\Delta)$ ,  $f^2(\Delta) = f[f(\Delta)]$ ,  $f^3(\Delta)$ ,  $\dots$  respectively represent the totality of probability distributions at the transition time  $i=1, 2, 3, \dots$ . Since clearly  $f(\Delta)$ ,  $f^2(\Delta)$ ,  $f^3(\Delta)$ ,  $\dots$  form a monotone decreasing sequence of closed bounded sets, their intersection  $\sigma = \bigcap_i f^i(\Delta)$  is nonempty. It is natural that we should investigate the set  $\sigma$  since  $\sigma$  represents the totality of possible "ultimate" probability distributions. Since  $\Delta$ ,  $f(\Delta)$ ,  $f^2(\Delta)$ ,  $\dots$  is a decreasing sequence of  $r$ -convex sets (convex polytopes having  $r \leq n$  vertices) with  $r$  also possibly decreasing ( $f$  being not necessarily one-to-one), it is known [2] that  $\sigma$  is actually an  $r$ -convex set with some  $r \leq n$ .

Beside  $\sigma$  we need to look at another subset  $\tau$  of  $\Delta$  consisting of all fixed points of  $f$ . According to the well-known Brouwer's fixed point

theorem, since  $f$  is a continuous mapping of  $\Delta$  into  $\Delta$ ,  $f$  must have at least one fixed point in  $\Delta$ , hence  $\tau$  is nonempty. Actually  $\tau$  is a subset of  $\sigma$  since if  $\mathbf{p} \in \tau$  then  $f(\mathbf{p}) = \mathbf{p}$  whence  $f^i(\mathbf{p}) = \mathbf{p}$  so that  $\mathbf{p} \in f^i(\Delta)$  for all  $i$  and  $\mathbf{p} \in \sigma$ . Now it can be shown (see [4]) that  $f(\sigma) = \sigma$  so that the effect of  $f$  on  $\sigma$  can be described by a permutation of vertices of  $\sigma$  and hence, for some integer  $j$ ,  $f^j$  will send each vertex back to itself and thus act as an identity mapping on  $\sigma$ .

Each point  $\mathbf{p}$  in  $\Delta$  may now be classified according to its relation to  $\tau$  and  $\sigma$ . For example, we may call  $\mathbf{p}$  *stationary* if  $\mathbf{p}$  is in  $\tau$ , *cyclic* if  $\mathbf{p}$  is in  $\sigma$  but not in  $\tau$ , *ultimately stationary* if  $\mathbf{p}$  is not in  $\sigma$  but  $f^i(\mathbf{p})$  tends to some  $\mathbf{q}$  in  $\tau$  as  $i$  tends to infinity, and *ultimately cyclic* if  $\mathbf{p}$  is none of those considered above. In particular,  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  may likewise be classified as stationary, cyclic, ultimately stationary, and ultimately cyclic.

DEFINITION 3. A Markov transformation  $f$  is said to be *cyclic* if  $f$  has a cyclic point in  $\Delta$ , i.e.,  $\sigma$  contains  $\tau$  properly. Otherwise  $f$  is said to be *acyclic*; in particular,  $f$  is said to be *stable* if  $\sigma = \tau$  consists of a single point; if furthermore this point lies in the interior of  $\Delta$ , then  $f$  is said to be *ergodically stable*.

Gantmacher [3] has an algebraic proof of Markov's theorem based on matrix theory. From the matrix point of view he defines (p. 88) *cyclic* and *acyclic* Markov chains. These however are not quite equivalent to cyclic and acyclic Markov transformations defined above. Gantmacher's definitions presuppose irreducibility of matrices (p. 80) whereas irreducibility consideration comes later in this paper.

We now make two useful observations regarding  $\sigma$  and  $\tau$ .

LEMMA 1. A Markov transformation  $f$  on  $\Delta$  cannot have a cyclic point in  $\Delta$  without having one on the boundary of  $\Delta$ , i.e.,  $\sigma - \tau \neq \emptyset$  implies  $(\sigma - \tau) \cap \partial\Delta \neq \emptyset$ .

PROOF. Let  $\mathbf{p} \in \sigma - \tau$  and  $\mathbf{q} \in \tau$ , then as we pointed out earlier, for some  $j$  (and this  $j$  must be greater than 1 since  $\mathbf{p} \notin \tau$ )  $f^j(\mathbf{p}) = \mathbf{p}$ , and of course  $f^j(\mathbf{q}) = \mathbf{q}$  for any  $j$ , but since  $f$  is a linear transformation, so is  $f^j$ , and  $f^j$  will leave the line connecting  $\mathbf{p}$  and  $\mathbf{q}$  pointwise fixed. This line must pierce (intersect) the boundary  $\partial\Delta$ , say at  $\mathbf{p}'$ . Now  $f^j(\mathbf{p}') = \mathbf{p}'$  so that  $f^{mj}(\mathbf{p}') = \mathbf{p}'$  for  $m = 1, 2, 3, \dots$ . This insures that  $\mathbf{p}' \in \sigma$ . On the other hand  $f(\mathbf{p}') \neq \mathbf{p}'$ , for otherwise  $f$  would leave  $\mathbf{p}'$  and  $\mathbf{q}$  fixed and consequently also  $\mathbf{p}$ , contradicting  $\mathbf{p} \notin \tau$ . Thus  $\mathbf{p}' \in (\sigma - \tau) \cap \partial\Delta$  and  $(\sigma - \tau) \cap \partial\Delta \neq \emptyset$ .

LEMMA 2. A Markov transformation  $f$  on  $\Delta$  cannot have more than

one stationary point without having one on the boundary of  $\Delta$ , i.e., unless  $\tau$  is a singleton set,  $\tau \cap \partial\Delta \neq \emptyset$ .

PROOF. If  $\mathbf{p} \neq \mathbf{q}$  and  $f(\mathbf{p}) = \mathbf{p}$  and  $f(\mathbf{q}) = \mathbf{q}$ , then since  $f$  is a linear transformation  $f$  will leave the line connecting  $\mathbf{p}$  and  $\mathbf{q}$  pointwise fixed. This line must intersect  $\partial\Delta$  and hence there will be a fixed point on  $\partial\Delta$ .

In order to arrive at a criterion for ergodic stability of Markov transformation the concept of *irreducibility* must be considered.

DEFINITION 4. A Markov transformation  $f$  on the stochastic  $n$ -simplex  $\Delta$  is said to be *invariant* over a stochastic subsimplex  $\Delta'$  of  $\Delta$  if  $f$  maps  $\Delta'$  into  $\Delta'$ .

DEFINITION 5. A Markov transformation  $f$  on  $\Delta$  is said to be *reducible* over  $\Delta$  to  $\Delta'$  if  $f$  is invariant over a proper subsimplex  $\Delta'$  of  $\Delta$ .  $f$  is said to be *irreducible* over  $\Delta$  if it is not reducible over  $\Delta$ .

The concept of irreducibility of  $f$  is an effective one since it is discernible from the transition matrix  $P$ . Noting that  $f$  is reducible over  $\Delta$  if and only if  $P$  contains a proper subtransition matrix  $P'$  we see that if  $P$  does not contain such a submatrix then  $f$  is irreducible. We now state in connection with the irreducibility of  $f$  two useful lemmas regarding  $\tau$ .

LEMMA 3. A Markov transformation  $f$  has a stationary point on the boundary  $\partial\Delta$  if and only if  $f$  is reducible over  $\Delta$ .

PROOF. Suppose  $\mathbf{p}$  is a stationary point on the boundary  $\partial\Delta$ . Without loss of generality we may suppose  $\mathbf{p}$  is a linear combination of  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$  with  $r < n$

$$\mathbf{p} = \sum_{i=1}^r x_i \mathbf{u}_i \quad \text{where } 0 < x_i \quad \text{and} \quad \sum_{i=1}^r x_i = 1$$

i.e.,  $\mathbf{p}$  is an interior point of the subsimplex  $\Delta'$  determined by  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ . We shall show  $f(\Delta') \subset \Delta'$  and hence  $f$  is reducible over  $\Delta$  to  $\Delta'$ . So suppose on the contrary that  $\Delta'$  were not invariant under  $f$ . Then at least one of  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ , say  $\mathbf{u}_1$  is transformed out of  $\Delta'$ , i.e.,  $f(\mathbf{u}_1) \cdot \mathbf{u}_{r+1} > 0$  say. From this it follows that

$$f(\mathbf{p}) \cdot \mathbf{u}_{r+1} = \left[ \sum_{i=1}^r x_i f(\mathbf{u}_i) \right] \cdot \mathbf{u}_{r+1} \geq x_1 f(\mathbf{u}_1) \cdot \mathbf{u}_{r+1} > 0$$

i.e.,  $\mathbf{p}$  is transformed out of  $\Delta'$  and hence cannot be a stationary point.

Conversely, if  $f$  is reducible over  $\Delta$ , say  $f$  is invariant over  $\Delta'$ ,

$f(\Delta') \subset \Delta'$ , then by Brouwer's fixed point theorem  $f$  has a fixed point in  $\Delta'$ , but  $\Delta' \subset \partial\Delta$ . This completes the proof.

LEMMA 4. *A Markov transformation  $f$  is irreducible over  $\Delta$  if and only if  $f$  has a unique stationary point in the interior of  $\Delta$ .*

PROOF. If  $f$  has a stationary point on the boundary  $\partial\Delta$ , then  $f$  is reducible by Lemma 3. If  $f$  has more than one stationary point, it will have a stationary point on the boundary  $\partial\Delta$  by Lemma 2, hence  $f$  will be reducible by Lemma 3 again. Finally, if  $f$  has a unique stationary point in the interior of  $\Delta$ ,  $f$  must be irreducible, for otherwise  $f$  would be invariant over some subsimplex  $\Delta'$ ,  $f(\Delta') \subset \Delta'$ , and by Brouwer's fixed point theorem  $f$  would have a fixed point in  $\Delta'$ , but  $\Delta' \subset \partial\Delta$ , and this contradicts the fact that  $f$  has a stationary point only in the interior.

Now that we know if  $f$  is irreducible over  $\Delta$  then  $f$  has a unique stationary point in the interior of  $\Delta$ , if we know further that  $f$  is acyclic ( $\sigma = \tau$ ) then  $f$  is ergodically stable. Conversely, if  $f$  is ergodically stable ( $\sigma = \tau = \text{an interior point of } \Delta$ ), then  $f$  is acyclic and by Lemma 4  $f$  is irreducible. Consequently we have actually proved the following

PROPOSITION 1. *A Markov transformation  $f$  is ergodically stable if and only if  $f$  is irreducible and acyclic.*

The above criterion for ergodic stability of  $f$  is not very effective in as much as acyclicity of  $f$  is not easily discernible. So we show next that irreducibility and acyclicity together is equivalent to *interiority* of  $f$ .

DEFINITION 6. A Markov transformation  $f$  is said to have the *interiority* property over the stochastic simplex  $\Delta$  if for some positive integer  $k$   $f^k(\Delta)$  lies in the interior of  $\Delta$ ,  $f^k(\Delta) \subset \subset \Delta$ .

Interiority of  $f$  is discernible from the transition matrix  $P$  since  $f$  has the interiority property if and only if  $P^k$  is positive for some  $k$ .

PROPOSITION 2. *A Markov transformation  $f$  is irreducible and acyclic if and only if  $f$  has the interiority property.*

PROOF. Assume that  $f^k(\Delta)$  lies in the interior of  $\Delta$ . Then from  $\tau \subset \sigma \subset f^k(\Delta) \subset \subset \Delta$  follows  $\sigma \cap \partial\Delta = \emptyset$ , and by Lemma 1  $\sigma = \tau$ , i.e.,  $f$  is acyclic. Also,  $\tau \cap \partial\Delta = \emptyset$  and by Lemma 2  $\tau$  consists of a single point. Since this unique stationary point is in the interior of  $\Delta$ , by Lemma 4  $f$  is irreducible over  $\Delta$ .

Conversely assume that  $f$  is irreducible and acyclic over  $\Delta$ . Acyclic-

ity means  $\sigma = \tau$ , and irreducibility implies  $\tau \cap \partial\Delta = \emptyset$  by Lemma 3. Hence  $\sigma \cap \partial\Delta = \emptyset$ . Now since  $\sigma = \bigcap_i f^i(\Delta)$ , for some integer  $k$   $f^k(\Delta) \subset \subset \Delta$ , for otherwise  $f^i(\Delta) \cap \partial\Delta \neq \emptyset$  for every  $i$  would imply  $\bigcap_i [f^i(\Delta) \cap \partial\Delta] \neq \emptyset$  with  $f^i(\Delta) \cap \partial\Delta$  forming a decreasing sequence of closed bounded (compact) sets, but then

$$\emptyset \neq \bigcap_i [f^i(\Delta) \cap \partial\Delta] = \bigcap_i f^i(\Delta) \cap \partial\Delta = \sigma \cap \partial\Delta$$

contradicting  $\sigma \cap \partial\Delta = \emptyset$ .

Combining Propositions 1 and 2 we obtain

**THEOREM 1.** *A Markov transformation  $f$  is ergodically stable if and only if  $f^k(\Delta)$  lies in the interior of  $\Delta$  for some positive integer  $k$ .*

In terms of transition matrix this "interiority criterion" of  $f$  may be stated as follows:

**COROLLARY 1.** *A Markov transformation  $f$  (or a finite Markov chain) is ergodically stable if and only if its (transition) matrix  $P$  is such that for some positive integer  $k$   $P^k$  has no zero entries.*

#### REFERENCES

1. A. A. Markov, *Investigation of a noteworthy case of dependent trials*, Izv. Ros. Akad. Nauk 1 (1907) (Russian) or B. V. Gnedenko, *Course in the theory of probability*, Fizmatgiz, Moscow, 1961; English transl., Chelsea, New York, 1962, pp. 142-145. MR 25 #2622.
2. V. Borovikov, *On the intersection of a sequence of simplexes*, Uspehi Mat. Nauk 7 (1952), no. 6 (52), 179-180. (Russian) MR 14, 784.
3. F. R. Gantmacher, *The theory of matrices*, GITTL, Moscow, 1953; English transl., Vol. 2, Chelsea, New York, 1959, pp. 50-93. MR 16, 438; MR 21 #6372c.
4. R. Z. Yeh, *On the effect of an affine transformation on a certain  $k$ -convex set*, (to appear).

UNIVERSITY OF HAWAII, HONOLULU, HAWAII 96822