AN ANALOGUE OF BIRKHOFF'S PROBLEM 111 FOR INFINITE MARKOV MATRICES[1]

Choo-Whan Kim

(received November 20, 1968)

1. Introduction. A celebrated theorem of Birkhoff ([1], [6]) states that the set of $n \times n$ doubly stochastic matrices is identical with the convex hull of the set of $n \times n$ permutation matrices. Birkhoff [2, p. 266] proposed the problem of extending his theorem to the set of infinite doubly stochastic matrices. This problem, which is often known as Birkhoff's Problem 111, was solved by Isbell ([3], [4]), Rattray and Peck [7], Kendall [5] and Révész [8]. From the viewpoint of probability theory, it is interesting to know analogues of the Birkhoff's theorem and Birkhoff's Problem 111 for Markov transition matrices. An analogue of the Birkhoff theorem for the set of $n \times n$ Markov (transition) matrices is known [6, p. 133]: the set of $n \times n$ Markov matrices is identical with the convex hull of the set of $n \times n$ Markov matrices with exactly one entry 1 in each row. The purpose of this paper is to give a solution (Theorems 1, 2 and 3 below) to a version of Birkhoff's Problem 111 for infinite Markov matrices.

An infinite matrix A with non-negative entries a_{ij} is called sub-Markov (Markov) if $\Sigma_j a_{ij} \leq 1$ $(= 1)$ for each i. By a matrix we

[1] This work was supported by the National Science Foundation under Grant GP-5270, Syracuse University.

shall always mean an infinite matrix unless the contrary is noted. We denote the set of sub-Markov matrices by m_σ and the set of Markov matrices by m. Clearly m_σ and m are convex sets and $m \subset m_\sigma$. A Markov matrix $A = (a_{ij})$ is called doubly stochastic if $\Sigma_j a_{ij} = \Sigma_i a_{ij} = 1$ for each i and for each j. Let p be the set of those Markov matrices with exactly one entry 1 in each row. In order to treat our problem we shall introduce a topological vector space. Let V be the vector space of real infinite matrices $A = (a_{ij})$ such that $\Sigma_j |a_{ij}| < \infty$ for each i. We define a neighborhood base at the zero matrix 0 by sets of the form $\{(a_{ij}) \in V : \Sigma_j |a_{ij}| < \varepsilon,\ i \leq N\}$, where $\varepsilon > 0$ and N is a positive integer. Topologize V with translates of these neighborhoods. Then V is a topological vector space with contains m_σ. For each subset S of V, S^- denotes the closure of the set S in the topology defined above. We note the following facts.

LEMMA 1. m_σ <u>is a closed convex subset of</u> V.

<u>Proof</u>. It is enough to prove $V - m_\sigma \subset V - m_\sigma^-$. We note that $V - m_\sigma$ is a disjoint union of the two sets:

$$\{(a_{ij}) \in V : a_{ij} < 0 \text{ for some } i \text{ and some } j\}$$

and

$$\{(a_{ij}) \in V : a_{ij} \geq 0,\ \Sigma_j a_{ij} > 1 \text{ for some } i\}.$$

Suppose that $A = (a_{ij})$ is in the first set with $a_{st} < 0$. Choose ε such that $-a_{st} > \varepsilon > 0$. Then for each $B = (b_{ij})$ in m_σ we have

$$\Sigma_j |a_{sj} - b_{sj}| \geq |a_{st} - b_{st}| = -a_{st} + b_{st} > \varepsilon ,$$

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press

so A is in $\mathcal{V} - \mathcal{m}_\sigma^-$. If $A = (a_{ij})$ is in the second set we may assume $\mu_1 = \Sigma_j a_{1j} > 1$. Since for each $B = (b_{ij}) \in \mathcal{m}_\sigma$

$$\Sigma_j |a_{1j} - b_{1j}| \geq |\Sigma_j a_{1j} - \Sigma_j b_{1j}| \geq |\mu_1 - 1| > 0 ,$$

A is also in $\mathcal{V} - \mathcal{m}_\sigma^-$.

LEMMA 2. $\mathcal{m}$ <u>is a closed convex subset of</u> $\mathcal{V}$.

<u>Proof</u>. We shall show $\mathcal{m}_\sigma - \mathcal{m} \subset \mathcal{m}_\sigma - \mathcal{m}^-$. Let $A = (a_{ij}) \in \mathcal{m}_\sigma - \mathcal{m}$ and $0 \leq \Sigma_j a_{ij} = \lambda_i < 1$ for some i. We have for each $B = (b_{ij}) \in \mathcal{m}$,

$$\Sigma_j |a_{ij} - b_{ij}| \geq |\Sigma_j a_{ij} - \Sigma_j b_{ij}| = |\lambda_i - 1| > 0 ,$$

and thus $A \in \mathcal{m}_\sigma - \mathcal{m}^-$. The assertion follows immediately.

2. <u>Extreme Points and Approximation</u>. We begin by proving

THEOREM 1. <u>A Markov matrix</u> P <u>is an extreme point of</u> $\mathcal{m}$ <u>iff</u> P <u>is in</u> $\mathcal{p}$.

<u>Proof</u>. ($\Rightarrow$): Let $A = (a_{ij})$ be a Markov matrix which is not in $\mathcal{p}$. We may assume $0 < a_{11} < 1$. We shall show that there are Markov matrices B and C that are distinct from A with $A = (B + C)/2$. Since $\Sigma_j a_{1j} = 1$ and $0 < a_{11} < 1$, there is a_{1t} such that $0 < a_{1t} < 1$. Choose ε such that $0 < \varepsilon < \min(a_{11}, a_{1t})$. Define the matrix $E = (e_{ij})$ by $e_{11} = \varepsilon$, $e_{1t} = -\varepsilon$ and $e_{ij} = 0$ otherwise. Then $B = A + E$ and $C = A - E$ are Markov matrices with $A = (B + C)/2$.

($\Leftarrow$): Let $P = (p_{ij}) \in \mathcal{p}$. For each i, there is j_i such that $p_{ij_i} = 1$ and $p_{ij} = 0$ for $j \neq j_i$. Suppose that $P = (A + B)/2$ where

$A = (a_{ij})$ and $B = (b_{ij})$ are Markov matrices. It follows from $p_{ij} = (a_{ij} + b_{ij})/2$ that, for each i, $p_{ij_i} = a_{ij_i} = b_{ij_i} = 1$ and $p_{ij} = a_{ij} = b_{ij} = 0$ for $j \neq j_i$. Thus $P = A = B$ and P is an extreme point of $\mathcal{M}$.

For extreme points of $\mathcal{M}_\sigma$ we have

PROPOSITION. A sub-Markov matrix P is an extreme point of $\mathcal{M}_\sigma$ if and only if $P = 0$ or $P \in \mathcal{P}$.

Proof. ($\Rightarrow$): Suppose that $A = (a_{ij})$ is a non-zero sub-Markov matrix which is not in $\mathcal{P}$. We may assume $0 < a_{11} < 1$. If $\lambda_1 = \Sigma_j a_{1j} > a_{11}$, then there is a_{1t} with $0 < a_{1t} \leq \lambda_1 - a_{11}$. Choose ε such that $0 < \varepsilon < \min(a_{11}, a_{1t})$. Define the matrix $E = (e_{ij})$ by $e_{11} = \varepsilon$, $e_{1t} = -\varepsilon$ and $e_{ij} = 0$ otherwise. Then $A + E$ and $A - E$ are sub-Markov matrices with $A = ((A + E) + (A - E))/2$. Hence A is not an extreme point of $\mathcal{M}_\sigma$. However, if $\lambda_1 = a_{11}$ i.e., $a_{1j} = 0$ for $j \geq 2$, then we choose ε with $0 < \varepsilon < a_{11}$. Define the matrix $E = (e_{ij})$ by $e_{11} = \varepsilon$ and $e_{ij} = 0$ otherwise. Then $A + E$ and $A - E$ are sub-Markov matrices with $A = ((A + E) + (A - E))/2$. Hence A is not an extreme point of $\mathcal{M}_\sigma$.

($\Leftarrow$): Plainly the zero matrix 0 is an extreme point of $\mathcal{M}_\sigma$. We may readily show that each $P \in \mathcal{P}$ is an extreme point of $\mathcal{M}_\sigma$ by a similar argument given in the proof of Theorem 1.

Let us denote by $co(\mathcal{P})$ the convex hull of the set $\mathcal{P}$. A simple example can be furnished to show $co(\mathcal{P}) \subsetneq \mathcal{M}$. For example, let $A = (a_{ij})$ be such that for each $i = 1,2,\ldots$, $a_{ij} = 1/i$ for $j = 1,\ldots,i$, and

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press

$a_{ij} = 0$ for $j > i$. What are those Markov matrices in co(p)? We have the following answer.

THEOREM 2. *Let* $A = (a_{ij})$ *be a Markov matrix*. $A \in$ co(p) *if and only if* a_{ij} *takes only finitely many distinct values*.

The following two lemmas will be useful in the proof of Theorem 2.

LEMMA 3. *If for each positive integer* m, *an infinite matrix* $A = (a_{ij})$ *with non-negative integers* a_{ij} *satisfies the conditions*: $0 \leq a_{ij} \leq m$ *and* $\Sigma_j a_{ij} = m$ *for each* i, *then the matrix* A/m *is in* co(p).

Proof. Let $\lambda_i = \min_j\{a_{ij}: a_{ij} > 0\}$, $i = 1,2,\ldots$, and $\lambda = \min_i \lambda_i$. Then $1 \leq \lambda \leq \lambda_i \leq m$, $i = 1,2,\ldots$. If $\lambda = m$ then $m = \lambda_i$ for each i. It follows that for each i, there exists j_i such that $a_{ij_i} = m$ and $a_{ij} = 0$ for $j \neq j_i$. Hence $A = mP$ for some $P \in p$. Now suppose that $1 \leq \lambda < m$. Let t be the smallest positive integer with $\lambda = \lambda_t$. For each i let j_i be the smallest positive integer with $a_{ij_i} = \lambda_i$. Define the matrix $P = (p_{ij})$ by $p_{ij_i} = 1$ and $p_{ij} = 0$ for $j \neq j_i$. Then the matrix $C = A - \lambda P = (c_{ij})$ satisfies the conditions: $c_{ij} \geq 0$ and $0 < \Sigma_j c_{ij} = m - \lambda < m$ for each i. Clearly the assertion holds for $m = 1$. If the assertion is true for each $k < m$ we have $C/(m - \lambda) \in$ co(p), so $A/m \in$ co(p). Hence the lemma follows by the induction.

COROLLARY. *If* $A = (a_{ij})$ *is a Markov matrix such that* a_{ij} *takes only finitely many distinct rational numbers, then* $A \in$ co(p).

We state a lemma of Isbell whose proof will be outlined in the proof of Theorem 2 for ease of our argument.

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press

LEMMA 4. (Isbell [4, p. 3]). <u>For any finite set of positive real numbers</u> $\lambda_1,\ldots,\lambda_n$, <u>there exists a Hamel (vector) basis</u> $\{b_\alpha\}$ <u>for the reals over the rationals such that each</u> λ_i is $\Sigma r_{ij} b_{\alpha_j}$ <u>with non-negative rational coefficients</u> r_{ij}.

<u>Proof of Theorem 2</u>. Since $(\Rightarrow)$ is obvious, it remains to prove $(\Leftarrow)$. Suppose that the entries a_{ij} of a Markov matrix A take $n + 1$ distinct values: $\lambda_0,\lambda_1,\ldots,\lambda_n$, where $\lambda_0 = 0$, $0 < \lambda_i \leq 1$, $i = 1,\ldots,n$. Let C be the convex cone (over the rationals) which consists of those $\Sigma_{i=1}^n r_i\lambda_i$ with non-negative rationals r_i such that $\Sigma_{i=1}^n r_i > 0$. Then C is a cone with vertex at 0 and $0 \notin C$. Since the cone C has an interior point, the set $C - C$ is the subspace of the reals over the rationals which is spanned by $\lambda_1,\ldots\lambda_n$. We pick a basis $\{b_1,\ldots,b_t\}$ for the subspace $C - C$ from $\{\lambda_1,\ldots,\lambda_n\}$ and extend it to a Hamel basis (see [4, p. 3]). In particular we have, for each i, $\lambda_i = \Sigma_{j=1}^t r_{ij}b_j$ with non-negative rationals r_{ij} and $t \leq n$. Since each a_{ij} can be identified with some λ_k, and $\Sigma_j a_{ij} = 1$ for each i, it follows readily that 1 is in the cone C and $1 = \Sigma_{j=1}^t s_j b_j$ with positive rationals s_j. It is also evident that $A = \Sigma_{j=1}^t b_j B_j$ where the matrix B_j has entries r_{ij}, $i = 1,\ldots,n$, and each B_j/s_j is a Markov matrix. It is easily seen from the corollary to Lemma 3 that each B_j/s_j is in $co(p)$ and so $A \in co(p)$. This completes the proof.

We establish the following approximation theorem by using an argument of Rattray and Peck [7, p. 56].

THEOREM 3. $m = co(p)^-$.

<u>Proof</u>. Since we have $co(p)^- \subset m$ from Lemma 2 and Theorem 2,

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press

it remains to prove $m \subset \mathrm{co}(p)^-$. Let $A = (a_{ij})$ be a Markov matrix and N a positive integer. Given $\varepsilon > 0$, there is a positive integer n such that $\Sigma_{j>n} a_{ij} < \varepsilon/4$ for each $i \leq N$. Then $\Sigma_{j\leq n} a_{ij} > 1 - \varepsilon/4$ for each $i \leq N$. Choose a positive integer m such that $n/m < \varepsilon/4$. Let p_{ij} be a non-negative integer such that $p_{ij}/m < a_{ij} \leq (p_{ij} + 1)/m$ when $a_{ij} > 0$ and $p_{ij} = 0$ when $a_{ij} = 0$. Define the sub-Markov matrix $B = (b_{ij})$ by $b_{ij} = p_{ij}/m$. If we set $r_i = \Sigma_j p_{ij}$ for each i, then $0 < 1 - \Sigma_j p_{ij}/m = 1 - r_i/m < 1$ and $1 \leq s_i = m - r_i < m$. For each i, we increase some of b_{ij} to c_{ij} by addition of $1/m$ in s_i places. Clearly $c_{ij} \geq b_{ij}$. The matrix $C = (c_{ij})$ is Markov such that c_{ij} takes only finitely many distinct rationals. Hence, by the corollary to Lemma 3, we can identify the matrix C with a matrix in $\mathrm{co}(p)$. We have, for each $i \leq N$,

$$\begin{aligned}\Sigma_j |a_{ij} - b_{ij}| &= \Sigma_{j\leq n}(a_{ij} - b_{ij}) + \Sigma_{j>n}(a_{ij} - b_{ij}) \\ &< \Sigma_{j\leq n} 1/m + \Sigma_{j>n} a_{ij} \\ &\leq n/m + \varepsilon/4 < \varepsilon/2.\end{aligned}$$

On the other hand it follows that, for each $i \leq N$,

$$\begin{aligned}\Sigma_j b_{ij} > \Sigma_{j<n} b_{ij} &\geq \Sigma_{j\leq n}(a_{ij} - 1/m) \\ &> 1 - \varepsilon/4 - n/m \\ &> 1 - \varepsilon/2,\end{aligned}$$

and thus

$$\begin{aligned}\Sigma_j |b_{ij} - c_{ij}| &= \Sigma_j c_{ij} - \Sigma_j b_{ij} \\ &< 1 - (1 - \varepsilon/2) = \varepsilon/2.\end{aligned}$$

The assertion follows from

$$\Sigma_j |a_{ij} - c_{ij}| \leq \Sigma_j |a_{ij} - b_{ij}| + \Sigma_j |b_{ij} - c_{ij}| < \varepsilon, \quad i \leq N.$$

Remark. If we define an infinite column sub-Markov (Markov) matrix $A = (a_{ij})$ by the condition: $a_{ij} \geq 0$ and $\Sigma_i a_{ij} \leq 1$ $(= 1)$ for each j, and the set $\mathcal{P}$ by those matrices with exactly one entry 1 in each column, the above results remain to be true with the obvious modification of the topological vector space $\mathcal{V}$. The results are found to be useful in approximation of Markov operators in $L_1(-\infty,\infty)$ which will be discussed elsewhere.

REFERENCES

1. G. Birkhoff, Tres observaciones sobre el algebra lineal. Rev. Univ. nac. Tucuman A. 5 (1946) 147-151.

2. G. Birkhoff, Lattice theory (revised edition). (New York, 1948).

3. J.R. Isbell, Birkhoff's problem 111. Proc. Amer. Math. Soc. 6 (1955) 217-218.

4. J.R. Isbell, Infinitely doubly stochastic matrices. Canad. Math. Bull. 5 (1962) 1-4.

5. D.G. Kendall, On infinite doubly stochastic matrices and Birkhoff's Problem 111. J. Lon. Math. Soc. 35 (1960) 81-84.

6. M. Marcus and H. Minc, A survey of matrix theory and matrix inequalities. (Allyn and Bacon, Boston, 1964).

7. B.A. Rattray and J.E.L. Peck, Infinite doubly stochastic matrices. Trans. Roy. Soc. Canada III (3) 49 (1955) 55-57.

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press

8. P. Revész, A probabilistic solution of Problem 111 of G. Birkhoff. Acta. Math. Acad. Sci. Hungar. 13 (1962) 187-198.

Simon Fraser University

https://doi.org/10.4153/CMB-1969-080-7 Published online by Cambridge University Press