proof of Birkhoff-von Neumann theorem

Canonical name	ProofOfBirkhoffvonNeumannTheorem
Date of creation	2013-03-22 13:10:03
Last modified on	2013-03-22 13:10:03
Owner	Andrea Ambrosio (7332)
Last modified by	Andrea Ambrosio (7332)
Numerical id	14
Author	Andrea Ambrosio (7332)
Entry type	Proof
Classification	msc 15A51

First, we prove the following lemma:

Lemma:

A convex combination of doubly stochastic matrices is doubly stochastic.

Proof:

Let $\{A_i\}_{i=1}^m$ be a collection of $n \times n$ doubly-stochastic matrices, and suppose $\{\lambda_i\}_{i=1}^m$ is a collection of scalars satisfying $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for each $i = 1, \ldots, m$. We claim that $A = \sum_{i=1}^m \lambda_i A_i$ is doubly stochastic.

Take any $i \in \{1, \ldots, m\}$. Since A_i is doubly stochastic, each of its rows and columns sum to 1. Thus each of the rows and columns of $\lambda_i A_i$ sum to λ_i.

By the definition of elementwise summation, given matrices $N = M_1 + M_2$, the sum of the entries in the ith column of N is clearly the sum of the sums of entries of the ith columns of M_1 and M_2 respectively. A similar result holds for the jth row.

Hence the sum of the entries in the ith column of A is the sum of the sums of entries of the ith columns of $\lambda_k A_k$ for each i, that is, $\sum_{k=1}^m \lambda_k = 1$. The sum of entries of the jth row of A is the same. Hence A is doubly stochastic.

Observe that since a permutation matrix has a single nonzero entry, equal to 1, in each row and column, so the sum of entries in any row or column must be 1. So a permutation matrix is doubly stochastic, and on applying the lemma, we see that a convex combination of permutation matrices is doubly stochastic. This provides one direction of our proof; we now prove the more difficult direction: suppose B is doubly-stochastic. Define a weighted graph $G = (V, E)$ with vertex set $V = \{r_1, \ldots, r_n, c_1, \ldots, c_n\}$, edge set E, where $e_{ij} = (r_i, c_i) \in E$ if $B_{ij} \neq 0$, and edge weight ω, where $\omega(e_{ij}) = B_{ij}$. Clearly G is a bipartite graph, with partitions $R = \{r_1, \ldots, r_n\}$ and $C = \{c_1, \ldots, c_n\}$, since the only edges in E are between r_i and c_j for some $i, j \in \{1, \ldots, n\}$. Furthermore, since $B_{ij} \geq 0$, then $\omega(e) > 0$ for every $e \in E$. For any $A \subset V$ define $N(A)$, the neighbourhood of A, to be the set of vertices $u \in V$ such that there is some $v \in A$ such that $(u, v) \in E$. We claim that, for any $v \in V$, $\sum_{u \in N(\{v\})} \omega(u, v) = 1$. Take any $v \in V$; either $v \in R$ or $v \in C$. Since G is bipartite, $v \in R$ implies $N(\{v\}) \subset C$, and $v \in C$ implies $N(\{v\}) \subset R$. Now,

$$v = r_i \implies \sum_{u \in N(r_i)} \omega(r_i, u) = \sum_{\substack{j=1, \\ e_{ij} \in E}}^{n} \omega(r_i, c_j) = \sum_{\substack{j=1, \\ B_{ij} \neq 0}}^{n} B_{ij} = \sum_{j=1}^{n} B_{ij} = 1$$

$$v = c_j \implies \sum_{u \in N(c_j)} \omega(u, c_j) = \sum_{\substack{i=1, \\ e_{ij} \in E}}^{n} \omega(r_i, c_j) = \sum_{\substack{i=1, \\ B_{ij} \neq 0}}^{n} B_{ij} = \sum_{i=1}^{n} B_{ij} = 1$$

since B is doubly stochastic. Now, take any $A \subset R$. We have

$$\sum_{\substack{v \in A \\ w \in N(A)}} \omega(v,w) = \sum_{v \in A} \sum_{w \in N(\{v\})} \omega(v,w) = \sum_{v \in A} 1 = |A|$$

Let $B = N(A)$. But then clearly $A \subset N(B)$, by definition of neighbourhood. So

$$|N(A)| = |B| = \sum_{\substack{v \in B \\ w \in N(B)}} \omega(v,w) \geq \sum_{\substack{v \in B \\ w \in A}} \omega(v,w) = \sum_{\substack{w \in A \\ v \in N(A)}} \omega(v,w) = |A|$$

So $|N(A)| \geq |A|$. We may therefore apply the graph-theoretic version of Hall's marriage theorem to G to conclude that G has a perfect matching. So let $M \subset E$ be a perfect matching for G. Define an $n \times n$ matrix P by

$$P_{ij} = \begin{cases} 1 & \text{if } e_{ij} \in M \\ 0 & \text{otherwise} \end{cases}$$

Note that $B_{ij} = 0$ implies $P_{ij} = 0$: if $B_{ij} = 0$, then $(r_i, c_j) \notin E$, so $(r_i, c_j) \notin M$, which implies $P_{ij} = 0$. Further, we claim that P is a permutation matrix:

Let i be any row of P. Since M is a perfect matching of G, there exists $e_0 \in M$ such that r_i is an end of e_0. Let the other end be c_j for some i; then $P_{ij} = 1$.

Suppose $i_i, i_2 \in \{1, \ldots, n\}$ with $i_1 \neq i_2$ and $P_{i_1,j} = P_{i_2,j} = 1$ for some j. This implies $(r_{i_1}, c_j), (r_{i_2}, c_j) \in M$, but this implies the vertex i is the end of two distinct edges, which contradicts the fact that M is a matching.

Hence, for each row and column of P, there is exactly one nonzero entry, whose value is 1. So P is a permutation matrix. Define $\lambda = \min_{i,j \in \{1,\ldots,n\}} \{B_{ij} \mid P_{ij} \neq 0\}$. We see that $\lambda > 0$ since $B_{ij} \geq 0$, and $P_{ij} \neq 0 \implies B_{ij} \neq 0$. Further, $\lambda = B_{pq}$ for some p, q. Let $D = B - \lambda P$. If $D = 0$, then $\lambda = 1$ and B is a permutation matrix, so we are done. Otherwise, note that D is non-negative; this is clear since $\lambda P_{ij} \leq \lambda \leq B_{ij}$ for any $B_{ij} \neq 0$. Notice that $D_{pq} = B_{pq} - \lambda P_{pq} = \lambda - \lambda \cdot 1 = 0$. Note that since every row and column of B and P sums to 1, that every row and column of $D = B - \lambda P$ sums to $1 - \lambda$. Define $B' = \frac{1}{1-\lambda} D$. Then every row and column of B' sums to 1, so B' is doubly stochastic. Rearranging, we have $B = \lambda P + (1 - \lambda) B'$. Clearly $B_{ij} = 0$ implies that $P_{ij} = 0$ which implies that $B'_{ij} = 0$, so the zero entries of B' are a superset of those of B'. But notice that $B'_{pq} = \frac{1}{1-\lambda} D_{pq} = 0$, so the zero entries of B' are a strict superset of those of B. We have decomposed B into a convex combination of a permutation matrix and another doubly stochastic matrix with strictly more zero entries than B. Thus we may apply this procedure repeatedly on the doubly stochastic matrix obtained from the previous step, and the number of zero entries will increase with each step. Since B has at most n^2 nonzero entries, we will obtain a convex combination of permutation matrices in at most n^2 steps. Thus B is indeed expressible as a convex combination of permutation matrices.

References

G. Birkhoff, "Tres observaciones sobre el algebra lineal," Univ. Nac. Tucumán Rev, Ser. A, no. 5, pp147–151. (1946)