Note Open Access
Special Issue Dedicated to Charles R. Johnson

Pietro Paparella*

Frobenius normal forms of doubly stochastic matrices

https://doi.org/10.1515/spma-2019-0015
Received August 15, 2019; accepted November 22, 2019

Abstract: An elementary proof of a fundamental result on doubly stochastic matrices in Frobenius normal form is given. This result is used to establish several well-known results concerning permutations, including a theorem due to Ruffini.

Keywords: Frobenius normal form, doubly stochastic matrix, permutation, permutation matrix

MSC: 15A21; 05A05

1 Introduction

In 1965, Perfect and Mirsky [5, Lemma 3] stated, without proof, that a doubly stochastic matrix is permutationally similar to a direct sum of irreducible, doubly stochastic matrices. Furthermore, they state that the result "is almost certainly well-known" and omit a proof because it "follows very easily from the definitions" [5, p. 38]. This result is fundamental in the Perfect–Mirsky conjecture, but, to the best of our knowledge, it appears sparingly in the literature: Liu and Lai [4, Theorem 2.7.4] prove the weaker result that a doubly stochastic matrix is either irreducible or permutationally similar to a direct sum of two doubly stochastic matrices; and Hartfiel and Spellman [2, Lemma 1(b)] give a proof via strong induction.

In this work, an elementary proof of this fundamental result is provided that relies on weak induction and the *Frobenius normal form* of a matrix. To demonstrate its import and utility, we apply it to permutation matrices and and characterize the Frobenius normal forms of a permutation matrix. This, in turn, is used to derive two well-known results concerning permutations, including the *disjoint cyclic form* and the result due to Ruffini that the order of a permutation in disjoint cyclic form is the least common multiple of the lengths of its disjoint cycles.

2 Background

For $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$, we let $\mathsf{M}_n(\mathbb{F})$ denote the set of n-by-n matrices with entries over $\mathbb{F}$, and $\mathbb{F}^n$ denote the collection of all column vectors of length n over $\mathbb{F}$. We let I_n denote the n-by-n identity matrix and e denotes the all-ones vector (the size of which is determined by the context in which it appears). Finally, for $n \in \mathbb{N}$, we let $\langle n \rangle := \{1, \dots, n\}$.

A *directed graph* (or simply *digraph*) $\Gamma = (V, E)$ consists of a finite, nonempty set V of *vertices*, together with a set of *arcs* $E \subseteq V \times V$. A digraph Γ is *strongly connected* if, for any two vertices u and v of V, there is a *(directed) walk* in Γ from u to v Every vertex of V is considered strongly connected to itself so that strong

***Corresponding Author: Pietro Paparella:** University of Washington Bothell, E-mail: pietrop@uw.edu

connectivity defines an equivalence relation, and hence a partition, of the vertices into *strongly connected components*. Notice that Γ is strongly connected if and only if Γ possesses one strongly connected component.

If $A \in \mathsf{M}_n(\mathbb{C})$, then the *digraph of* A, denoted by $\Gamma(A)$, has vertices $V = \langle n \rangle$ and arcs $E = \{(i,j) \in V \times V \mid a_{ij} \neq 0\}$. For $n \geq 2$, a matrix $A \in \mathsf{M}_n(\mathbb{C})$, is *reducible* if there is a permutation matrix P such that

$$P^\top AP = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

in which A_{11} and A_{22} are nonempty square matrices and 0 is a zero block. If A is not reducible, then A is called *irreducible*. It is well-known that A is irreducible if and only if $\Gamma(A)$ is strongly connected (see, e.g., Brualdi and Ryser [1, Theorem 3.2.1]).

If $A \in \mathsf{M}_n(\mathbb{C})$, then there is a permutation matrix P such that

$$P^\top AP = \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ & \ddots & \vdots \\ & & A_{kk} \end{bmatrix},$$

in which the matrices $A_{11}, \ldots, A_{kk}$ are irreducible [1, Theorem 3.2.4]. The matrix $P^\top AP$ is called a *Frobenius normal form* (FNF) of A and is not unique. The irreducible matrices $A_{11}, \ldots, A_{kk}$, called the *irreducible components of* A, are unique up to permutation similarity.

3 Main Result

Recall that a nonnegative matrix $A \in \mathsf{M}_n(\mathbb{R})$ is called *doubly stochastic* if $Ae = e = A^\top e$. With this definition, we are now ready to present the main result.

Theorem 3.1. Let $A \in \mathsf{M}_n(\mathbb{R})$ and suppose that

$$P^\top AP = \begin{bmatrix} A_{11} & \cdots & A_{1k} \\ & \ddots & \vdots \\ & & A_{kk} \end{bmatrix}$$

is a Frobenius normal form of A. If A is doubly stochastic, then $A_{ij} = 0$, $i < j$, i.e.,

$$P^\top AP = \bigoplus_{i=1}^{k} A_{ii} = \begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{kk} \end{bmatrix}$$

and the irreducible components $A_{11} \in \mathsf{M}_{n_1}(\mathbb{R}), \ldots, A_{kk} \in \mathsf{M}_{n_k}(\mathbb{R})$ are doubly stochastic.

Proof. Proceed by induction on k, the number of irreducible blocks in any FNF of A. When $k = 1$, then A is irreducible and the result is clear.

For the induction-step, assume that $k \geq 2$ and that the result holds for any doubly stochastic matrix having $k - 1$ irreducible blocks in any FNF.

Since $Pe = P^\top e = e$, it follows that $P^\top AP$ is doubly stochastic. Consequently, $A_{11}^\top e = e$ and $A_{11}e \leq e$ (inequality here is considered entrywise). We claim that $A_{11}e = e$. Otherwise,

$$n_1 = e^\top e = \left(e^\top A_{11}\right) e = e^\top (A_{11}e) < e^\top e = n_1,$$

a contradiction. Thus, A_{11} is doubly stochastic and

$$P^\top A P = \begin{bmatrix} A_{11} & & & \\ & A_{22} & \cdots & A_{2k} \\ & & \ddots & \vdots \\ & & & A_{kk} \end{bmatrix}.$$

The submatrix

$$\begin{bmatrix} A_{22} & \cdots & A_{2k} \\ & \ddots & \vdots \\ & & A_{kk} \end{bmatrix}$$

of $P^\top AP$ is doubly stochastic and has $k-1$ irreducible blocks; as such, the induction-hypothesis applies and the result is established. □

Remark 3.2. For $\alpha > 0$ let $\mathcal{CS}^n_\alpha := \{A \in \mathsf{M}_n(\mathbb{R}) \mid A \geq 0,\ Ae = \alpha e = A^\top e\}$. Notice that $A \in \mathcal{CS}^n_\alpha$ if and only if $r_i = c_j = \alpha,\ \forall i, j \in \langle n \rangle$. Furthermore, $A \in \mathcal{CS}^n_\alpha$ if and only if $B := A/\alpha \in \mathcal{CS}^n_1$, therefore Theorem 3.1 applies to matrices in $\mathcal{CS}^n_\alpha$, $\alpha > 0$.

4 Permutations and Permutation Matrices.

Denote by $\mathcal{S}_n$ the *symmetric group* of $\langle n \rangle$. For $\sigma \in \mathcal{S}_n$, the *permutation matrix corresponding to* σ, denoted by P_σ, is the the n-by-n matrix such that (i, j)-entry is $\delta_{i,\sigma^{-1}(i)}$, where δ_{ij} denotes the Kronecker delta. When the context is clear, P_σ is abbreviated to P. As is well-known, $P_\sigma P_\gamma = P_{\sigma\gamma}$ and so the set of all n-by-n permutation matrices, denoted by $\mathcal{P}_n$, forms a group under matrix multiplication and the map $\Phi : \mathcal{S}_n \longrightarrow \mathcal{P}_n$, defined by $\Phi(\sigma) = P_\sigma$ is an isomorphism.

Observation 4.1. If $P \in \mathsf{M}_n(\mathbb{R})$ is a permutation matrix, then P is irreducible if and only if $\Gamma(P)$ is an n-cycle.

Proof. Let $v \in \langle n \rangle$. Since $\Gamma(P)$ is strongly connected, $\Gamma(P)$ has a k-cycle of distinct vertices $v := v_1 \longrightarrow \cdots \longrightarrow v_k \longrightarrow v_1$ with $1 \leq k \leq n$.

We claim that $k = n$; otherwise, if $k < n$, then $\Gamma(P)$ clearly contains more than one connected component, which is a contradiction. Thus, $k = n$, i.e., $\Gamma(P)$ is an n-cycle.

Conversely, the matrix P is clearly irreducible when $\Gamma(P)$ is an n-cycle. □

Observation 4.2. For a positive integer $n \geq 2$, the matrix $C_n := \begin{bmatrix} 0 & 1 \\ I_{n-1} & 0 \end{bmatrix}$ is called a *basic circulant*. Notice that C_n is the permutation matrix corresponding to the *cyclic permutation* $\begin{pmatrix} 1 & \cdots & n \end{pmatrix}$. Thus, if P is an irreducible permutation matrix, then there is a permutation matrix Q such that $Q^\top PQ = C_n$.

Corollary 4.3 (FNF of a permuation matrix)**.** If P is a permutation matrix, then there is a permutation matrix Q such that $Q^\top PQ = \bigoplus_{i=1}^k C_{n_i},\ 1 \leq k \leq n$.

Lemma 4.4. If $\alpha = \begin{pmatrix} \alpha_1 & \cdots & \alpha_k \end{pmatrix} \in \mathcal{S}_n$ and $\beta = \begin{pmatrix} \beta_1 & \cdots & \beta_\ell \end{pmatrix} \in \mathcal{S}_n$ are disjoint, cyclic permutations (i.e., $\alpha_i \neq \beta_j$) and $\gamma \in \mathcal{S}_n$, then $\gamma\alpha\gamma^{-1}$ and $\gamma\beta\gamma^{-1}$ are disjoint, cyclic permutations.

Proof. If $y_i := \gamma(\alpha_i)$, $1 \leq i \leq k$, then $\gamma(\alpha(\gamma^{-1}(y_i))) = \gamma(\alpha(\alpha_i)) = \gamma(\alpha_{i+1}) = y_{i+1}$, where, for convenience, $\alpha_{k+1} := \alpha_1$ and $y_{k+1} := y_1$. If $y \in \langle n \rangle \backslash \{y_1, \ldots, y_k\}$ and $x := \gamma^{-1}(y)$, then $(\gamma\alpha\gamma^{-1})(y) = \gamma(\alpha(\gamma^{-1}(y))) = \gamma(\alpha(x)) = \gamma(x) = y$. Thus, $\gamma\alpha\gamma^{-1} = \begin{pmatrix} y_1 & \cdots & y_k \end{pmatrix}$. A similar argument demonstrates that $\gamma\beta\gamma^{-1} = \begin{pmatrix} z_1 & \cdots & z_\ell \end{pmatrix}$ with $z_i := \gamma(\beta_i)$, $1 \leq i \leq \ell$.

For contradiction, if $\gamma\alpha\gamma^{-1}$ and $\gamma\beta\gamma^{-1}$ are not disjoint, then there are positive integers i and j such that $\gamma(\alpha_i) = y_i = z_j = \gamma(\beta_j)$, i.e., $\alpha_i = \beta_j$, a contradiction. □

Theorem 4.5 (disjoint cyclic form). If $\sigma \in \mathcal{S}_n$, then there are disjoint, cyclic permutations $\sigma_1, \ldots, \sigma_k \in \mathcal{S}_n$, $1 \le k \le n$, such that $\sigma = \prod_{i=1}^k \sigma_i$.

Proof. If P is the permutation matrix corresponding to σ, then, following Corollary 4.3, there is a permutation matrix Q corresponding to $\gamma \in \mathcal{S}_n$ such that

$$P = Q\left(\bigoplus_{i=1}^k C_{n_i}\right)Q^\top = \bigoplus_{i=1}^k QC_{n_i}Q^\top,\ 1 \le k \le n.$$

If $\hat{\sigma}_i$ is the cyclic permutation corresponding to C_{n_i} and σ_i is the permutation corresponding to $QC_{n_i}Q^\top$, i.e., $\sigma_i = \gamma\hat{\sigma}_i\gamma^{-1}$, then $\sigma_1, \ldots, \sigma_k$ are pairwise disjoint (Lemma 4.4) and $\sigma = \prod_{i=1}^k \sigma_i$, as desired. □

Remark 4.6. Johnson [3] arrives at the FNF of a permutation matrix, however the disjoint cyclic form of a permutation is assumed.

5 Ruffini's theorem.

Before we derive a matricial proof of Ruffini's theorem, we will require the following fundamental fact.

Lemma 5.1. If C_n is a basic circulant, then $|C_n| = n$.

Proof. The matrix C_n is the companion matrix of the polynomial $p(t) = t^n - 1$. As such, p is the characteristic and minimal polynomial of C_n. By the Cayley-Hamilton theorem, $p(C_n) = 0$, i.e., $(C_n)^n = I_n$ and $|C_n| \le n$. For contradiction, if $|C_n| = m < n$, then the polynomial $q(t) = t^m - 1$ annihilates C_n, contradicting the minimality of p. Thus, $|C_n| = n$. □

The following well-known result is due to Ruffini.

Theorem 5.2. Let $\sigma \in \mathcal{S}_n$. If there are disjoint cyclic permutations $\sigma_1, \ldots, \sigma_k \in \mathcal{S}_n$, $1 \le k \le n$, such that $\sigma = \prod_{i=1}^k \sigma_i$, then $|\sigma| = \operatorname{lcm}(|\sigma_1|, \ldots, |\sigma_k|)$.

Proof. Let P be the permutation matrix corresponding to σ and let Q be a permutation matrix such that $Q^\top PQ = \bigoplus_{i=1}^k C_{n_i}$ is a Frobenius normal form of P. Since the irreducible components of a matrix are unique (up to permutation) and correspond to disjoint cycles of σ, without loss of generality, it may be assumed that σ_i is the permutation corresponding to $QC_{n_i}Q^\top$.

Since $|\sigma| = |P|$ and since $P \longmapsto Q^\top PQ$ is an inner automorphism, it follows that $|\sigma| = |Q^\top PQ| = |\bigoplus_{i=1}^k C_{n_i}|$.

By the mechanics of block matrix multiplication,

$$\left(\bigoplus_{i=1}^k C_{n_i}\right)^m = \bigoplus_{i=1}^k (C_{n_i})^m,$$

whence

$$\left|\bigoplus_{i=1}^k C_{n_i}\right| = \operatorname{lcm}(|C_{n_1}|, \ldots, |C_{n_k}|) = \operatorname{lcm}(|\sigma_1|, \ldots, |\sigma_k|),$$

i.e., $|\sigma| = \operatorname{lcm}(|\sigma_1|, \ldots, |\sigma_k|)$. □

Acknowledgments: The author thanks the University of Washington Royalty Research Fund for financial support, Geoffrey R. Robinson for the idea of the proof of Theorem 3.1, and the anonymous referee who alerted the author to reference [2].

Supported by University of Washington Royalty Research Fund Award # A139510.

References

[1] R. A. Brualdi and H. J. Ryser. *Combinatorial matrix theory*, volume 39 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1991.
[2] D. J. Hartfiel and J. W. Spellmann. A role for doubly stochastic matrices in graph theory. *Proc. Amer. Math. Soc.*, 36:389–394, 1972.
[3] C. R. Johnson. An inclusion region for the field of values of a doubly stochastic matrix based on its graph. *Aequationes Math.*, 17(2-3):305–310, 1978.
[4] B. Liu and H.-J. Lai. *Matrices in combinatorics and graph theory*, volume 3 of *Network Theory and Applications*. Kluwer Academic Publishers, Dordrecht, 2000. With a foreword by Richard A. Brualdi.
[5] H. Perfect and L. Mirsky. Spectral properties of doubly-stochastic matrices. *Monatsh. Math.*, 69:35–57, 1965.