

Convex Set of Doubly Substochastic Matrices

Lei Deng

College of Electronic and Information Engineering
Shenzhen University
Shenzhen, China
ldeng@szu.edu.cn

Abstract—Denote \mathcal{A} as the set of all doubly substochastic $m \times n$ matrices and let k be a positive integer. Let \mathcal{A}_k be the set of all $1/k$ -bounded doubly substochastic $m \times n$ matrices, i.e., $\mathcal{A}_k \triangleq \{E \in \mathcal{A} : e_{i,j} \in [0, 1/k], \forall i = 1, 2, \dots, m, j = 1, 2, \dots, n\}$. Denote \mathcal{B}_k as the set of all matrices in \mathcal{A}_k whose entries are either 0 or $1/k$. We prove that \mathcal{A}_k is the convex hull of all matrices in \mathcal{B}_k . In addition, we introduce an application of this result in communication system.

Index Terms—Doubly substochastic matrix, combinatorial structures, convex hull, extreme points.

I. INTRODUCTION

An $n \times n$ matrix $E = (e_{i,j})$ is *doubly stochastic* if it satisfies the following conditions:

$$\begin{cases} e_{i,j} \geq 0, & \forall i, j = 1, 2, \dots, n; \\ \sum_{j=1}^n e_{i,j} = 1, & \forall i = 1, 2, \dots, n; \\ \sum_{i=1}^n e_{i,j} = 1, & \forall j = 1, 2, \dots, n. \end{cases}$$

Birkhoff in [1] shows that the set of all $n \times n$ doubly stochastic matrix is the convex hull of all $n \times n$ permutation matrices. Recall that a permutation matrix is a square $(0, 1)$ matrix of which any line (row or column) has exactly one 1.

An $m \times n$ matrix $E = (e_{i,j})$ is *doubly substochastic* if it satisfies the following conditions:

$$\begin{cases} e_{i,j} \geq 0, & \forall i = 1, 2, \dots, m, j = 1, 2, \dots, n; \\ \sum_{j=1}^n e_{i,j} \leq 1, & \forall i = 1, 2, \dots, m; \\ \sum_{i=1}^m e_{i,j} \leq 1, & \forall j = 1, 2, \dots, n. \end{cases}$$

Mirsky in [2] shows that the set of all $n \times n$ doubly substochastic matrix is the convex hull of all $n \times n$ subpermutation matrices. Recall that a subpermutation matrix is a $(0, 1)$ matrix of which any line (row or column) has at most one 1. It is straightforward to extend Mirsky's result to general $m \times n$ matrices because we can always augment a rectangular doubly substochastic matrix to a square doubly substochastic matrix by adding some zero lines.

Each entry in doubly stochastic/substochastic matrices can be any real number in the interval $[0, 1]$. However, if we impose an upper bound $1/k$ where k is a positive integer for all entries of doubly stochastic/substochastic matrices, the convex-hull characterization becomes different. Watkins and Merris in [3] prove that the set of all $1/k$ -bounded doubly stochastic matrices is the convex hull of all doubly stochastic matrices whose entries are either 0 or $1/k$. In this work, we obtain the counterpart result of [3] for doubly

substochastic matrices. Specifically, we prove that the set of all $1/k$ -bounded doubly substochastic matrices is the convex hull of all doubly substochastic matrices whose entries are either 0 or $1/k$. In addition, we introduce an application of this result in the design of delay-constrained input-queued switch in communication system.

II. OUR RESULT

Denote \mathcal{A} as the set of all doubly substochastic $m \times n$ matrices and let k be a positive integer. Define the set of all $1/k$ -bounded doubly substochastic $m \times n$ matrices as \mathcal{A}_k ¹, i.e.,

$$\mathcal{A}_k \triangleq \{E \in \mathcal{A} : e_{i,j} \in [0, 1/k], \forall i = 1, 2, \dots, m, j = 1, 2, \dots, n\}. \quad (1)$$

Denote \mathcal{B}_k as the set of all matrices in \mathcal{A}_k whose entries are either 0 or $1/k$. We now present our result.

Theorem 1. \mathcal{A}_k is the convex hull of all matrices in \mathcal{B}_k .

Proof. It is straightforward to show that \mathcal{A}_k is a non-empty compact convex set in $\mathbb{R}^{m \times n}$ and any matrix in \mathcal{B}_k is an extreme point of set \mathcal{A}_k . Since any non-empty compact convex set in \mathbb{R}^n is the convex hull of its extreme points [4, Theorem 2.23], to finish the proof, we thus only need to show that any matrix $E \in \mathcal{A}_k \setminus \mathcal{B}_k$ is not an extreme point of set \mathcal{A}_k .

An entry being neither 0 nor $1/k$ is called a *middle* entry, and a row (resp. column) containing at least one middle entry is called a *middle* row (resp. column). A middle line is either a middle row or a middle column. For any matrix $E \in \mathcal{A}_k \setminus \mathcal{B}_k$, it has at least one middle entry. We consider two cases.

Case I: Any middle line has at least two middle entries. We aim at finding a chain of middle entries of matrix E . We begin with any middle entry, denoted as (r_1, c_1) . Then row r_1 is a middle row which has at least two middle entries. Hence, we can find a new middle entry, denoted as (r_1, c_2) . Similarly, column c_2 is a middle column which has at least two middle entries. Thus we can find another new middle entry (r_2, c_2) . Then we can find another new middle entry (r_2, c_3) in middle row r_2 . If $c_3 = c_1$, we find the (latest-visited) old middle entry $(r_3, c_3) = (r_1, c_1)$ and terminate; otherwise, we can find another new middle entry (r_3, c_3) . The process continues by alternating a middle row and a

¹We do not show the matrix dimension in notation \mathcal{A}_k but it should be clear under context.

middle column. In each new middle line, if there are old middle entries, we find the latest-visited old middle entry and terminate; otherwise, we find a new middle entry and continue. Since E only has a finite number of middle entries, the process must terminate by finding an old middle entry, terminating either (i) at a middle entry $(r_{t'}, c_{t'})$ which coincides with an old middle entry (r_t, c_t) or (ii) at a middle entry $(r_{t'}, c_{t'+1})$ which coincides with an old middle entry (r_t, c_{t+1}) for some positive integers t and t' such that $t < t' - 1$. Let us consider terminating-condition (i). The proof for terminating-condition (ii) is similar. In terminating-condition (i), we can find a loop of middle entries:

$$(r_t, c_t) \rightarrow (r_t, c_{t+1}) \rightarrow (r_{t+1}, c_{t+1}) \rightarrow \cdots \rightarrow (r_{t'-1}, c_{t'}) \rightarrow (r_{t'}, c_{t'}) = (r_t, c_t),$$

which has in total $2(t' - t)$ different middle entries. Now we construct the following two matrices E' and E'' , both of which have the same entries of matrix E except

$$\begin{cases} e'_{r_i, c_i} = e_{r_i, c_i} + \epsilon, e''_{r_i, c_i} = e_{r_i, c_i} - \epsilon, \\ \quad \forall i = t, t+1, \dots, t' - 1; \\ e'_{r_i, c_{i+1}} = e_{r_i, c_{i+1}} - \epsilon, e''_{r_i, c_{i+1}} = e_{r_i, c_{i+1}} + \epsilon, \\ \quad \forall i = t, t+1, \dots, t' - 1. \end{cases} \quad (2)$$

Under such construction, any row/column sum of E' and E'' is equal to that of E . We can always find a sufficiently small positive real number ϵ such that all entries in E' and E'' are in the interval $[0, 1/k]$, ensuring $E' \in \mathcal{A}_k$ and $E'' \in \mathcal{A}_k$. Since we have $E = (E' + E'')/2$ and $E' \neq E''$, matrix E is not an extreme point of set \mathcal{A}_k .

Let us illustrate an example for $m = 4$, $n = 3$ and $k = 2$. Consider the following matrix

$$E = \begin{pmatrix} \mathbf{0.2} & \mathbf{0.3} & 0.5 \\ 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ \mathbf{0.3} & \mathbf{0.2} & 0.1 \end{pmatrix}. \quad (3)$$

Clearly, $E \in \mathcal{A}_2 \setminus \mathcal{B}_2$ and any middle line has at least two middle entries. Using the aforementioned procedure, we can find a chain of middle entries of matrix E , marked as bold in (3). By applying $\epsilon = 0.05$ in (2), we can construct two matrices E' and E'' as follows,

$$E' = \begin{pmatrix} \mathbf{0.25} & \mathbf{0.25} & 0.5 \\ 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ \mathbf{0.25} & \mathbf{0.25} & 0.1 \end{pmatrix},$$

and

$$E'' = \begin{pmatrix} \mathbf{0.15} & \mathbf{0.35} & 0.5 \\ 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ \mathbf{0.35} & \mathbf{0.15} & 0.1 \end{pmatrix}.$$

We can check that $E' \neq E''$, $E' \in \mathcal{A}_2$, $E'' \in \mathcal{A}_2$, and $E = (E' + E'')/2$. Therefore, E in (3) cannot be an extreme point of \mathcal{A}_2 .

Case II: There exists at least one middle line containing only one middle entry. We again aim at finding a chain of middle entries. We first find a middle line (a middle row or a middle column) which has only one middle entry. Let us assume that this is a middle column, say column c_1 . The proof for a middle row is similar. Denote the only middle entry in this middle column as (r_1, c_1) . Now in the row r_1 we try to find another new middle entry. If we cannot find such a middle entry, we terminate; otherwise, we denote the new middle entry as (r_1, c_2) and we continue. Similar to *Case I*, the process continues by alternating a middle row and a middle column. In each new middle line, if there are old middle entries, we find the latest-visited old middle entry and terminate; otherwise, we try to find a new middle entry: we continue if indeed we can find a new middle entry but terminate if we cannot find any new middle entry. Since E only has a finite number of middle entries, the process must terminate either (i) when we find an old middle entry or (ii) when we cannot find any new middle entry. If it terminates at condition (i), we can again find a loop of middle entries and use the proof of *Case I* to show that E is not an extreme points. If it terminates at condition (ii), it either terminates at a middle entry (r_t, c_t) or a middle entry (r_t, c_{t+1}) for some $t \geq 1$. We prove for the case (r_t, c_t) . The other case (r_t, c_{t+1}) has a similar proof. If the process terminates at entry (r_t, c_t) , we cannot find an (old or new) middle entry in row r_t , i.e., r_t is a middle row with only one middle entry (r_t, c_t) . We now have a chain of $(2t - 1)$ different middle entries,

$$(r_1, c_1) \rightarrow (r_1, c_2) \rightarrow (r_2, c_2) \rightarrow \cdots \rightarrow (r_{t-1}, c_t) \rightarrow (r_t, c_t).$$

Now we construct the following two matrices E' and E'' , both of which have the same entries of matrix E except

$$\begin{cases} e'_{r_i, c_i} = e_{r_i, c_i} + \epsilon, e''_{r_i, c_i} = e_{r_i, c_i} - \epsilon, \\ \quad \forall i = 1, 2, \dots, t; \\ e'_{r_i, c_{i+1}} = e_{r_i, c_{i+1}} - \epsilon, e''_{r_i, c_{i+1}} = e_{r_i, c_{i+1}} + \epsilon, \\ \quad \forall i = 1, 2, \dots, t - 1. \end{cases} \quad (4)$$

Under such construction, any row/column sum of E' and E'' is equal to that of E , except column c_1 and row r_t . The column- c_1 (resp. row- r_t) sum in E' is equal to the column- c_1 (resp. row- r_t) sum in E plus ϵ and the column- c_1 (resp. row- r_t) sum in E'' is equal to the column- c_1 (resp. row- r_t) sum in E minus ϵ . However, since column c_1 (resp. row r_t) has only one middle entry (r_1, c_1) (resp. (r_t, c_t)) in E , the column- c_1 (resp. row- r_t) sum in E must be in the interval $(0, 1)$. Thus, we can always find a sufficiently small positive real number ϵ such that both the column- c_1 sum and the row- r_t sum in E' and E'' are still in the interval $(0, 1)$ and all entries in E' and E'' are in the interval $[0, 1/k]$, ensuring $E' \in \mathcal{A}_k$ and $E'' \in \mathcal{A}_k$. Since again we have $E = (E' + E'')/2$ and $E' \neq E''$, matrix E is not an extreme point of set \mathcal{A}_k .

Similar to Case I, let us illustrate an example for $m = 4$,

$n = 3$ and $k = 2$. Consider the following matrix

$$E = \begin{pmatrix} 0 & \mathbf{0.2} & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 \\ \mathbf{0.1} & \mathbf{0.3} & 0 \end{pmatrix}. \quad (5)$$

Clearly, $E \in \mathcal{A}_2 \setminus \mathcal{B}_2$ and the first column and the first row have only one middle entry. Using the aforementioned procedure, we can find a chain of middle entries of matrix E , marked as bold in (5). By applying $\epsilon = 0.05$ in (4), we can construct two matrices E' and E'' as follows,

$$E' = \begin{pmatrix} 0 & \mathbf{0.25} & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 \\ \mathbf{0.15} & \mathbf{0.25} & 0 \end{pmatrix},$$

and

$$E'' = \begin{pmatrix} 0 & \mathbf{0.15} & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 \\ \mathbf{0.05} & \mathbf{0.35} & 0 \end{pmatrix}.$$

We can check that $E' \neq E''$, $E' \in \mathcal{A}_2$, $E'' \in \mathcal{A}_2$, and $E = (E' + E'')/2$. Therefore, E in (5) cannot be an extreme point of \mathcal{A}_2 .

Case I and Case II complete the proof. \square

Remark. Our proof is inspired by Watkins and Merris's work [3]. In particular, the proof of *Case I* is similar to that in [3] for doubly stochastic matrices. However, the proof of *Case II* is new for doubly substochastic matrices which could have middle lines with only one middle entry. In fact, the key to proving Theorem 1 is to divide all possibilities into *Case I* and *Case II* with respect to middle lines.

III. AN APPLICATION IN COMMUNICATION SYSTEM

In this section, we illustrate an application in [5], which applies our result to the design of delay-constrained input-queued switch in communication system. Deng et al. in [5] consider an $n \times n$ input-queued switch where each input transmits a delay-constrained traffic flow to each output. Thus, there are in total n^2 delay-constrained flows. The traffic pattern is frame-synchronized in the sense that all flows generate the first packet at slot 1 and then generate a new packet every T slots, and all packets have a hard deadline of T slots—if a packet cannot be delivered to its output in T slots once it is generated, it will expire and be removed from the system. The performance metric of a delay-constrained flow is its timely throughput, which is defined as the per-slot expected number of packets delivered successfully before expiration.

In the input-queued switch, each input can transmit at most one packet per slot and each output can receive at most one packet per slot, known as the crossbar constraints. Thus, the action of the switch in each slot corresponds to a subpermutation matrix $D = (d_{i,j} : i = 1, 2, \dots, n, j = 1, 2, \dots, n)$ where $d_{i,j} = 1$ means that input i is scheduled to transmit a packet to output j at this slot while $d_{i,j} = 0$ means that input i

is not scheduled to transmit a packet to output j at this slot. A scheduling policy determines the action of all slots, possibly in a random fashion.

The capacity region is the set of all achievable timely throughput matrices $R = (r_{i,j} : i = 1, 2, \dots, n, j = 1, 2, \dots, n)$. A timely throughput matrix $R = (r_{i,j} : i = 1, 2, \dots, n, j = 1, 2, \dots, n)$ is achievable in the sense that there exists a scheduling policy such that the timely throughput of the flow from any input i to any output j is at least $r_{i,j}$.

Deng et al. in [5, Theorem 1] prove that the capacity region of the considered delay-constrained input-queued switch is all $1/T$ -bounded doubly substochastic $n \times n$ matrices \mathcal{A}_T ². It is straightforward to show that any achievable timely throughput matrix R belongs to \mathcal{A}_T . Then the key to proving [5, Theorem 1] is to show that any timely throughput matrix $R \in \mathcal{A}_T$ can be achieved by a scheduling policy, which relies on our result, i.e., Theorem 1, in this paper.

Theorem 1 shows that any timely throughput matrix $R \in \mathcal{A}_T$ can be expressed as a convex combination of all (say in total L) doubly substochastic matrices whose entries are either 0 or $1/T$ (i.e., set \mathcal{B}_T). Namely, we can find a set of $\{p_l\}$ such that

$$R = \sum_{l=1}^L p_l B_l, \quad (6)$$

where $p_l \in [0, 1], \forall l = 1, 2, \dots, L, \sum_{l=1}^L p_l = 1$, and B_l is the l -th element in set \mathcal{B}_T . Then, it follows that

$$TR = \sum_{l=1}^L p_l (TB_l). \quad (7)$$

Let $C_l \triangleq TB_l$. Then any entry of matrix C_l is either 0 or 1 and each line (row or column) has at most T ones. According to [6, Theorem 4.4.3], C_l can be expressed as the sum of T subpermutation matrices, i.e.,

$$C_l = \sum_{t=1}^T D_{l,t}, \quad (8)$$

where $D_{l,t}$ is a subpermutation matrix. Thus, it follows that

$$TR = \sum_{l=1}^L p_l (TB_l) = \sum_{l=1}^L p_l C_l = \sum_{l=1}^L p_l \sum_{t=1}^T D_{l,t}. \quad (9)$$

Next a randomized scheduling policy is constructed as follows. Choose the l -th set of permutation matrices $\{D_{l,t} : t = 1, 2, \dots, T\}$ with probability p_l . If the l -th set of permutation matrices $\{D_{l,t} : t = 1, 2, \dots, T\}$ is chosen, the switch schedules the packets according to permutation matrix $D_{l,t}$ at slot $t = 1, 2, \dots, T$ and repeats the actions every T

²The definition of A_k and B_k in Sec. II is for general $m \times n$ matrices. In this section, due to the structure of the switch, the definition of \mathcal{A}_T and \mathcal{B}_T mentioned soon is for $n \times n$ matrices.

slots. Therefore, the timely throughput of the flow from input i to output j is

$$\begin{aligned} \frac{\sum_{l=1}^L p_l \left(\sum_{t=1}^T D_{l,t} \right)_{i,j}}{T} &= \frac{\sum_{l=1}^L p_l (C_l)_{i,j}}{T} \\ &= \frac{(TR)_{i,j}}{T} = \frac{Tr_{i,j}}{T} = r_{i,j}. \end{aligned} \quad (10)$$

Thus, for any timely throughput matrix $R \in \mathcal{A}_T$, there exists a scheduling policy such that R is achievable. This proves that the capacity region of the delay-constrained input-queued switch is indeed \mathcal{A}_T .

As we can see, the key step of the proof above is (6), which is derived from our result, i.e., Theorem 1. This illustrates an application of our combinatorial result, i.e., Theorem 1, in communication system.

IV. CONCLUSION

In this work, we prove that the set of all $1/k$ -bounded doubly substochastic $m \times n$ matrices is the convex hull of all doubly substochastic matrices whose entries are either 0 or $1/k$. This generalizes the work [3] by Watkins and Merris from doubly stochastic matrices to doubly substochastic matrices. We also introduce how to apply this result to characterize the capacity region of a delay-constrained input-queued switch in communication system.

ACKNOWLEDGEMENT

The author would like to thank Qiulin Lin for insightful discussions.

REFERENCES

- [1] G. Birkhoff, "Tres observaciones sobre el algebra lineal," *Universidad Nacional de Tucumán Revista, Serie A*, vol. 5, pp. 147–151, 1946.
- [2] L. Mirsky, "On a convex set of matrices," *Archiv der Mathematik*, vol. 10, no. 1, pp. 88–92, 1959.
- [3] W. Watkins and R. Merris, "Convex sets of doubly stochastic matrices," *Journal of Combinatorial Theory, Series A*, vol. 16, no. 1, pp. 129–130, 1974.
- [4] M. Giaquinta and G. Modica, *Mathematical Analysis: Foundations and Advanced Techniques for Functions of Several Variables*. Springer Science & Business Media, 2011.
- [5] L. Deng, W. S. Wong, P. Chen, Y. S. Han, and H. Hou, "Delay-constrained input-queued switch," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2464–2474, 2018.
- [6] R. A. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory*. Cambridge University Press, 1991.