Constrained $(0, 1)$-Matrix Completion with a Staircase of Fixed Zeros

Wei Chen[a,*], Yanfang Mo[a], Li Qiu[a], Pravin Varaiya[b]

[a] *Department of Electronic & Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*
[b] *Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, CA 94720, USA*

Abstract

We study the $(0, 1)$-matrix completion with prescribed row and column sums wherein the ones are permitted in a set of positions that form a Young diagram. We characterize the solvability of such $(0, 1)$-matrix completion problems via the nonnegativity of a structure tensor which is defined in terms of the problem parameters: the row sums, column sums, and the positions of fixed zeros. This reduces the exponential number of inequalities in a direct characterization yielded by the max-flow min-cut theorem to a polynomial number of inequalities. The result is applied to two engineering problems arising in smart grid and real-time systems, respectively.

Keywords: Constrained $(0, 1)$-matrix completion, Young diagram, fixed zeros, structure tensor

2010 MSC: 05C90, 15A83

1. Introduction

Given are two nonnegative integral vectors $r = \begin{bmatrix} r_1 & r_2 & \dots & r_m \end{bmatrix}'$ and $h = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix}'$. Assume that $\sum_{i=1}^m r_i = \sum_{j=1}^n h_j$. Let $\mathscr{A}(r, h)$ denote the set of $m \times n$ $(0, 1)$-matrices with row sum vector given by r and column sum vector given by h. In other words, a matrix A belongs to $\mathscr{A}(r, h)$ if and only if all its entries are 0's or 1's such that the ith row sum is r_i and the jth column sum is h_j. The completion of such $(0, 1)$-matrices with given row and column sums has been attracting continuing interest since the independent pioneering works by Gale [12] and Ryser [20] in 1957. Let A^* be a $(0, 1)$-matrix with row sum vector given by r such that the 1's are put as far to the left as possible. The column sum vector of A^*, denote by r^*, is referred to as the conjugate vector of r. Note

[*] Corresponding author
Email addresses: wchenust@gmail.com (Wei Chen), ymoaa@connect.ust.hk (Yanfang Mo), eeqiu@ust.hk (Li Qiu), varaiya@berkeley.edu (Pravin Varaiya)

that r^* naturally has its elements ordered non-increasingly. The famous Gale-Ryser Theorem says that $\mathscr{A}(r,h) \neq \emptyset$ if and only if the majorization relation $h \prec r^*$ holds, which by definition means

$$\sum_{j=1}^{k} h_j^{\downarrow} \leq \sum_{j=1}^{k} r_j^*, \text{ for } k = 1, 2, \ldots, n,$$

where $h^{\downarrow}$ stands for the reordered version of h with its elements rearranged in a non-increasing order. For a comprehensive treatment of majorization theory and its applications in combinatorics, one can refer to the book [17].

Gale and Ryser's papers pointed to a number of research opportunities. Efforts have been devoted to investigating various types of constrained $(0,1)$-matrix completion problems. In particular, attention has been paid to scenarios where the 1's are only permitted in a certain set of positions, and the remaining positions are forced to be filled with 0's. The positions wherein the 1's are permitted are called free positions, while the others are called forbidden positions. Studying the constrained $(0,1)$-matrix completion problems, in addition to being theoretically interesting, helps in many application areas such as operation research [12], discrete tomography [14], real-time systems [6], smart grid [8, 13, 19], electoral seat allocation [16], etc. The article [2] and the book [3] and the references therein provide a comprehensive survey. To better understand the state of the art, some pertinent results closely related to this paper are briefly reviewed below.

First, it has been widely recognized that there exists a natural one-one correspondence between a $(0,1)$-matrix and an associated network flow [12, 10, 2, 4]. As such, any $(0,1)$-matrix completion problem, whether constrained or not, can be computationally solved via an associated maximal flow problem in polynomial time. While a numerical solution may serve the purpose in some applications, in many others an analytic characterization for the solvability of the constrained $(0,1)$-matrix completion problems is also of great interest. In general, a direct application of the well-known max-flow min-cut theorem yields a characterization given by an exponential number of inequalities [18]. It is often desirable to obtain a simpler characterization involving fewer inequalities by exploiting the underlying pattern of the fixed zeros.

Indeed, substantial progress has been made in various special cases when the fixed 0's admit certain particular structures. Reference [11] considered a subset of $\mathscr{A}(r,h)$ consisting of square $(0,1)$-matrices with zero trace, i.e., the diagonal elements are fixed to be 0's. To deal with this case, let $A^{\flat}$ be a zero-trace $(0,1)$-matrix with row sum vector given by r such that all the 1's are put as far to the left as possible. Denote by $r^{\flat}$ the column sum vector of $A^{\flat}$. It was shown in [11] that there exists a matrix in $\mathscr{A}(r,h)$ with zero trace if and only if the majorization relation $h \prec r^{\flat}$ holds. This result was then extended in [1, 9] to the case with at most one fixed 0 in each column. Another interesting extension was reported in [16] which considered the case when the fixed 0's correspond to a series of square submatrices on the main diagonal. Moreover, the study of $(0,1)$-matrix completion has also been generalized to the completion of integral matrices with given row and column sums wherein each entry lies between a lower bound and an upper bound [7]. Under certain restrictions, [7] provides a condition for the existence of such integral matrices given again by a majorization type relation.

In addition to the majorization results, an alternative way to address the $(0,1)$-matrix completion problems is via the notion of structure matrix, introduced by Ryser in the investigation of the trace of $(0,1)$-matrices [21]. Specifically, the set $\mathscr{A}(r,h)$ is associated with an $(m+1)\times(n+1)$ structure matrix $W=[w_{k_1k_2}]$ defined as:

$$w_{k_1k_2} = k_1k_2 + \sum_{i=k_1+1}^{m} r_i - \sum_{j=1}^{k_2} h_j, \qquad \text{for } k_1 = 0,1,\ldots,m, \text{ and } k_2 = 0,1,\ldots,n.$$

This matrix is called structure matrix since it is solely determined by the structure information available, namely, the row sum vector r and column sum vector h, and is independent of the specific choice of a matrix A in $\mathscr{A}(r,h)$. It was noted in [2] that the nonnegativity of W in fact gives an equivalent condition for the nonemptyness of $\mathscr{A}(r,h)$. For the constrained $(0,1)$-matrix completion problems as described before, various modified versions of structure matrix have been proposed. It turns out that in many cases, equivalent existence conditions can be established in terms of the nonnegativity of the associated structure matrices [9, 4, 16]. In particular, [4] considered a subset of $\mathscr{A}(r,h)$ in which every matrix can be partitioned into the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix},$$

where 0 is an $(m-p)\times(n-q)$ zero matrix. The authors defined an associated structure matrix $W=[w_{k_1k_2}]$ of dimension $(p+1)\times(q+1)$ as follows:

$$w_{k_1k_2} = k_1k_2 + \sum_{i=k_1+1}^{m} r_i - \sum_{j=1}^{k_2} h_j - \sum_{i=p+1}^{m} (r_i-k_2)^+ - \sum_{j=q+1}^{n} (h_j-k_1)^+, \qquad \text{for } k_1 = 0,1,\ldots,p, \text{ and } k_2 = 0,1,\ldots,q.$$

It was shown that this subset of $\mathscr{A}(r,h)$ is nonempty if and only if W is nonnegative.

In this paper, we study a particular type of constrained $(0,1)$-matrix completion problems. Our study is motivated by two applications arising in smart grid and real-time systems. To be specific, we are interested in the completion of $(0,1)$-matrices wherein the 1's are only permitted in a set of positions that form a Young diagram. In this case, the positions of the fixed 0's correspond to a staircase pattern. This problem, when the fixed 0's form a zero block, specializes to the one considered in [4]. We note that the scenario with a staircase of fixed 0's has also been considered in [5] in a more general setting. Therein the authors generalized several classical results known for $\mathscr{A}(r,h)$ to the constrained case and proposed a simple constructing algorithm. In this paper, we obtain a characterization on r, h, and the positions of fixed 0's for the solvability of the concerned constrained $(0,1)$-matrix completion problems. The characterization is given in terms of the nonnegativity of a certain structure tensor. Such characterization was not present in [5]. We also apply the result to the motivating problems in smart grid and real-time systems.

The notation is more or less standard and will be made clear as we proceed. A tensor W is said to be nonnegative, denoted $W \geq 0$, if all the entries of W are nonnegative.

2. Problem Formulation

We consider a specific type of constrained $(0,1)$-matrix completion problems wherein the free positions form a Young diagram. A Young diagram is a collection of cells, arranged in left-justified rows with non-increasing row sizes. A Young diagram is said to be of shape $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ if its row sizes are given by $\lambda_i, i = 1, 2, \ldots, m$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. As an illustration, a Young diagram of shape $\lambda = \{6, 4, 3\}$ is shown in Figure 1.

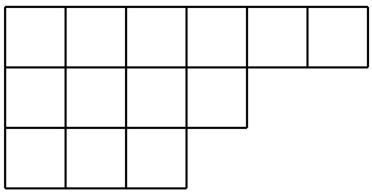

Figure 1: A Young diagram of shape $\lambda = \{6, 4, 3\}$.

Let $\mathscr{A}_\lambda(r,h)$ be a subset of $\mathscr{A}(r,h)$ wherein the free positions form a Young diagram of shape $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$. Without loss of generality, assume that $\lambda_1 = n$. Suppose there are τ distinct values in total among $\lambda_1, \lambda_2, \ldots, \lambda_m$. Every matrix $A \in \mathscr{A}_\lambda(r,h)$ then admits the following block upper anti-triangular form

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1(\tau-1)} & A_{1\tau} \\ A_{21} & \cdots & A_{2(\tau-1)} & 0 \\ \vdots & .\cdot^{\cdot} & .\cdot^{\cdot} & \vdots \\ A_{\tau 1} & 0 & \cdots & 0 \end{bmatrix}, \tag{1}$$

where the forbidden positions form a $(\tau-1)$-step staircase pattern. Let the size of A_{ij} be $p_i \times q_j$. Denote by $\tilde{p}_i$ the ith partial sum of the sequence $p_1, p_2, \ldots, p_\tau$, i.e., $\tilde{p}_i = \sum_{l=1}^{i} p_l, i = 1, 2, \ldots, \tau$. Similarly, denote by $\tilde{q}_j$ the jth partial sum of the sequence $q_1, q_2, \ldots, q_\tau$, i.e., $\tilde{q}_j = \sum_{l=1}^{j} q_l, j = 1, 2, \ldots, \tau$. Let $\tilde{p}_0 = \tilde{q}_0 = 0$.

We wish to characterize when this constrained $(0,1)$-matrix completion problem is solvable, i.e., $\mathscr{A}_\lambda(r,h) \neq \emptyset$. We are also interested in finding algorithms to construct such a matrix $A \in \mathscr{A}_\lambda(r,h)$. Notice that when $\lambda = \{n, n, \ldots, n\}$, $\mathscr{A}_\lambda(r,h)$ coincides with $\mathscr{A}(r,h)$ that has been extensively studied.

As is well known, a direct application of the max-flow min-cut theorem leads to 2^{m+n} inequalities. By exploiting the staircase pattern of fixed 0's in $\mathscr{A}_\lambda(r,h)$, we get a much simpler characterization involving only $(q_1+1)(q_2+1)\ldots(q_\tau+1)$ inequalities. These inequalities can be expressed in a compact form via a certain structure tensor that will be defined later.

The results in this paper extend straightforwardly to the case when the fixed 0's exhibit the staircase pattern as in (1) after re-ordering of rows and columns.

3. Preliminary

Consider a network $G = (V, E)$ with one source node s and one sink node t, where V and E represent the node set and edge set, respectively. Denote by (u, v) a directed edge from node u to node v. Each edge (u, v) has a nonnegative capacity $c(u, v)$. An $s-t$ cut is a partition of the nodes into two disjoint sets S and T such that $s \in S$ and $t \in T$. The capacity of a cut, denoted $c(S, T)$, is the sum of the capacities of the edges leaving the set S.

A celebrated result in network flow theory is the max-flow min-cut theorem.

Lemma 3.1 ([10]). *In any network flow with source s and sink t, the value of the maximum flow is equal to the minimum capacity over all $s-t$ cuts.*

Numerically, the maximum flow problem can be solved via various algorithms in $O(|E|)$ time where $|E|$ is the number of edges, for instance, the well-known Ford-Fulkerson algorithm [10]. Moreover, if all the edges have integer capacities, there exists an integer maximal flow in the network. For more details on network flow theory, one can refer to [10, 15].

Now, let us confine our attention to a bipartite network $G = (V, E)$ where

$$\begin{aligned} V &= \{u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_n\}, \\ E &= \{(v_j, u_i) | (i, j) \text{ is a free position in } \mathscr{A}_\lambda(r,h)\}, \end{aligned}$$

and each edge has capacity equal to 1. Associate each node with a demand:

$$\begin{aligned} d(u_i) &= r_i, \quad\quad i = 1, 2, \ldots, m, \\ d(v_j) &= -h_j, \ j = 1, 2, \ldots, n. \end{aligned}$$

A flow f in such a bipartite network with node demands is said to be feasible if it satisfies the capacity constraint and

$$\begin{aligned} \sum_{v_j : (v_j, u_i) \in E} f(v_j, u_i) &\geq d(u_i), \quad \text{for each } u_i, i = 1, 2, \ldots, m, \\ \sum_{u_i : (v_j, u_i) \in E} f(v_j, u_i) &\leq -d(v_j), \ \text{for each } v_j, j = 1, 2, \ldots, n. \end{aligned}$$

One can establish a one-one correspondence between a feasible integer flow in G and a matrix A in $\mathscr{A}_\lambda(r,h)$: the flow on edge (v_j,u_i) corresponds to the (i,j)th entry of A. Therefore, the constrained $(0,1)$-matrix completion at hand can be translated to a flow feasibility problem, which can be further translated to the maximal flow problem of an augmented network. Specifically, let us add two nodes s and t to the network. Also add an edge from node s to each node $v_j, j=1,2,\ldots,n$, and an edge from each node $u_i, i=1,2,\ldots,m$, to node t. The capacities of the added edges are given by

$$\begin{aligned} c(s,v_j) &= -d(v_j) = h_j,\ j=1,2,\ldots,n,\\ c(u_i,t) &= d(u_i) = r_i, \quad\ \ i=1,2,\ldots,m. \end{aligned}$$

We thus obtain an augmented network $\tilde{G}$ with only one source node s and one sink node t.

Lemma 3.2 ([15]). *The network G has a feasible integer flow f, if and only if the augmented network $\tilde{G}$ has a maximum integer flow that saturates all the edges leaving the source s and all the edges entering the sink t.*

Lemma 3.2 together with the max-flow min-cut theorem indicates that there exists a feasible integer flow in G if and only if the minimum capacity over all cuts of $\tilde{G}$ is equal to $\sum_{i=1}^m r_i$. In other words, all cuts of $\tilde{G}$ should have capacity greater than or equal to $\sum_{i=1}^m r_i$.

4. Main Result

Our main result relies critically on a notion called structure tensor defined as below.

Let r, h, λ, and τ be as before. Assume the following monotonicity on r and h:

$$\begin{aligned} & r_{\tilde{p}_i+1} \ge r_{\tilde{p}_i+2} \ge \cdots \ge r_{\tilde{p}_{i+1}}, i=0,1,\ldots,\tau-1,\\ & h_{\tilde{q}_j+1} \ge h_{\tilde{q}_j+2} \ge \cdots \ge h_{\tilde{q}_{j+1}}, j=0,1,\ldots,\tau-1. \end{aligned} \tag{2}$$

This assumption is without loss of generality since one can always re-order the rows and columns such that the monotonicity as in (2) is satisfied while the Young pattern of the free positions is maintained.

Associated with $\mathscr{A}_\lambda(r,h)$, define a τth order tensor $W(r,h,\lambda)=[w_{k_1k_2\ldots k_\tau}]$ as

$$\begin{aligned} w_{k_1k_2\ldots k_\tau} = & \sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \cdots + \sum_{j=\tilde{q}_{\tau-1}+k_\tau+1}^{\tilde{q}_\tau} h_j\\ & - \sum_{i=1}^{\tilde{p}_1}[r_i-(k_1+k_2+\cdots+k_\tau)]^+ - \sum_{i=\tilde{p}_1+1}^{\tilde{p}_2}[r_i-(k_1+k_2+\cdots+k_{\tau-1})]^+\\ & - \cdots - \sum_{i=\tilde{p}_{\tau-1}+1}^{\tilde{p}_\tau}(r_i-k_1)^+, \end{aligned} \tag{3}$$

where $0 \leq k_1 \leq q_1, 0 \leq k_2 \leq q_2, \ldots, 0 \leq k_\tau \leq q_\tau$. Such a tensor W is called the structure tensor since it is completely determined by the structural information r, h, λ, and has nothing to do with the specific choice of a matrix A in $\mathscr{A}_\lambda(r,h)$.

The main theorem in this paper gives a necessary and sufficient condition for the nonemptyness of $A_\lambda(r,h)$. The proof uses the same idea as the proof of Theorem 2.3 in [4]. Nevertheless, the reduction of inequalities here is more technically involved, leading to a condition given by the structure tensor.

Theorem 4.1. *$\mathscr{A}_\lambda(r,h) \neq \emptyset$ if and only if $W(r,h,\lambda) \geq 0$.*

Proof. For notational convenience, we shall show the result for the representative case when $\tau = 3$. The general case can be shown by a simple analogy.

As remarked before, $\mathscr{A}_\lambda(r,h) \neq \emptyset$ if and only if all the cuts in the augmented network $\tilde{G}$ have capacities greater than or equal to $\sum_{i=1}^m r_i$. Consider an arbitrary cut (S,T) of $\tilde{G}$. Observe that the node set S has the following form:

$$S = \{s\} \cup \{v_j : j \in K\} \cup \{u_i : i \in \overline{L}\},$$

for some subset $K \subseteq \{1,2,\ldots,n\}$ and $L \subseteq \{1,2,\ldots,m\}$, where we define

$$\overline{K} = \{1,2,\ldots,n\}\backslash K, \quad \overline{L} = \{1,2,\ldots,m\}\backslash L.$$

The capacity of such a cut is then given by

$$c(S,T) = \sum_{j\in\overline{K}} h_j + \sum_{i\in\overline{L}} r_i + \sum_{j\in K, i\in L, (v_j,u_i)\in E} c(v_j,u_i). \tag{4}$$

In view of the above analysis, it follows that $\mathscr{A}_\lambda(r,h)$ is nonempty if and only if

$$\sum_{j\in\overline{K}} h_j + \sum_{i\in\overline{L}} r_i + \sum_{j\in K, i\in L, (v_j,u_i)\in E} c(v_j,u_i) \geq \sum_{i=1}^m r_i, \tag{5}$$

for all possible subsets K and L.

Note that condition (5) involves 2^{m+n} inequalities. In the sequel, we shall show that a large part of those inequalities are redundant and, thus, can be excluded. To this end, we divide all the cuts into different classes and attempt to find the cut with minimum capacity within each class. Specifically, let all the cuts satisfying

$$|K\cap\{1,2,\ldots,\tilde{q}_1\}|=k_1, |K\cap\{\tilde{q}_1+1,\tilde{q}_1+2,\ldots,\tilde{q}_2\}|=k_2, |K\cap\{\tilde{q}_2+1,\tilde{q}_2+2,\ldots,\tilde{q}_3\}|=k_3,$$
$$|L\cap\{1,2,\ldots,\tilde{p}_1\}|=l_1, |L\cap\{\tilde{p}_1+1,\tilde{p}_1+2,\ldots,\tilde{p}_2\}|=l_2, |L\cap\{\tilde{p}_2+1,\tilde{p}_2+2,\ldots,\tilde{p}_3\}|=l_3,$$

for certain $0 \leq k_1 \leq q_1$, $0 \leq k_2 \leq q_2$, $0 \leq k_3 \leq q_3$, and $0 \leq l_1 \leq p_1$, $0 \leq l_2 \leq p_2$, $0 \leq l_3 \leq p_3$ form a class. One can see that any $s-t$ cut belongs to one of such classes.

Consider the class of cuts associated with k_1, k_2, k_3 and l_1, l_2, l_3. In view of (4), the cut with minimal capacity within this class corresponds to

$$K = \{1,2,\ldots,k_1\} \cup \{\tilde{q}_1+1,\tilde{q}_1+2,\ldots,\tilde{q}_1+k_2\} \cup \{\tilde{q}_2+1,\tilde{q}_2+2,\ldots,\tilde{q}_2+k_3\},$$
$$L = \{1,2,\ldots,l_1\} \cup \{\tilde{p}_1+1,\tilde{p}_1+2,\ldots,\tilde{p}_1+l_2\} \cup \{\tilde{p}_2+1,\tilde{p}_2+2,\ldots,\tilde{p}_2+l_3\}.$$

This is due to the monotonicity assumption on r and h as in (2) and the fact that the term

$$\sum_{j\in K, i\in L, (v_j,u_i)\in E} c(v_j,u_i)$$

is completely determined by k_1,k_2,k_3 and l_1,l_2,l_3. In fact, we have

$$\sum_{j\in K, i\in L, (v_j,u_i)\in E} c(v_j,u_i) = k_1l_1 + k_1l_2 + k_1l_3 + k_2l_1 + k_2l_2 + k_3l_1,$$

and, thus, the minimal capacity of the cuts within this class is given by

$$\begin{aligned}
&\sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j + \sum_{i=1}^{m} r_i - \sum_{i=1}^{l_1} r_i - \sum_{i=\tilde{p}_1+1}^{\tilde{p}_1+l_2} r_i - \sum_{i=\tilde{p}_2+1}^{\tilde{p}_2+l_3} r_i \\
&\quad + k_1l_1 + k_1l_2 + k_1l_3 + k_2l_1 + k_2l_2 + k_3l_1 \\
&= \sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j + \sum_{i=1}^{m} r_i \\
&\quad + \left[(k_1+k_2+k_3)l_1 - \sum_{i=1}^{l_1} r_i\right] + \left[(k_1+k_2)l_2 - \sum_{i=\tilde{p}_1+1}^{\tilde{p}_1+l_2} r_i\right] + \left[k_1l_3 - \sum_{i=\tilde{p}_2+1}^{\tilde{p}_2+l_3} r_i\right].
\end{aligned}$$

Now, by invoking Lemma 2.1 in [4], we can further find the minimal capacity within all classes of cuts corresponding to the same set of k_1,k_2,k_3, i.e.,

$$\begin{aligned}
&\min_{\substack{0\le l_1\le p_1\\0\le l_2\le p_2\\0\le l_3\le p_3}} \sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j + \sum_{i=1}^{m} r_i \\
&\qquad + \left[(k_1+k_2+k_3)l_1 - \sum_{i=1}^{l_1} r_i\right] + \left[(k_1+k_2)l_2 - \sum_{i=\tilde{p}_1+1}^{\tilde{p}_1+l_2} r_i\right] + \left[k_1l_3 - \sum_{i=\tilde{p}_2+1}^{\tilde{p}_2+l_3} r_i\right] \\
&= \sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j + \sum_{i=1}^{m} r_i \\
&\quad - \sum_{i=1}^{\tilde{p}_1} [r_i-(k_1+k_2+k_3)]^+ - \sum_{i=\tilde{p}_1+1}^{\tilde{p}_2} [r_i-(k_1+k_2)]^+ - \sum_{i=\tilde{p}_2+1}^{\tilde{p}_3} (r_i-k_1)^+ \\
&= w_{k_1k_2k_3} + \sum_{i=1}^{m} r_i.
\end{aligned}$$

From here, one can see that the 2^{m+n} inequalities in (5) hold if and only if $w_{k_1k_2k_3} \ge 0$ for all $0\le k_1\le q_1, 0\le k_2\le q_2, 0\le k_3\le q_3$. □

Theorem 4.1 reduces the 2^{m+n} inequalities in (5) to $(q_1+1)(q_2+1)\ldots(q_\tau+1)$ inequalities. When $\tau = 1$, the result simplifies to the well-known majorization condition for the nonemptyness of $\mathscr{A}(r,h)$. When $\tau = 2$, the structure tensor $W(r,h,\lambda)$ becomes a structure matrix. Note that the case of $\tau = 2$ has also been studied in [4] which deploys a slightly different structure matrix. It can be easily verified that the nonnegativity of both structure matrices is equivalent.

In the sequel, an intuitive interpretation is given for the nonnegativity of the structure tensor W. For illustrative purpose, the case when $\tau = 3$ is considered. Suppose there exists a matrix $A \in \mathscr{A}_\lambda(r,h)$. For given k_1, k_2, k_3, where $0 \le k_1 \le q_1, 0 \le k_2 \le q_2, 0 \le k_3 \le q_3$, define the following sets of positions:

$$
\begin{aligned}
P_1 &= \{(i,j) | 1 \le i \le \tilde{p}_1, k_1 < j \le \tilde{q}_1\},\\
P_2 &= \{(i,j) | 1 \le i \le \tilde{p}_1, \tilde{q}_1 + k_2 < j \le \tilde{q}_2\},\\
P_3 &= \{(i,j) | 1 \le i \le \tilde{p}_1, \tilde{q}_2 + k_3 < j \le \tilde{q}_3\},\\
P_4 &= \{(i,j) | \tilde{p}_1 < i \le \tilde{p}_2, k_1 < j \le \tilde{q}_1\},\\
P_5 &= \{(i,j) | \tilde{p}_1 < i \le \tilde{p}_2, \tilde{q}_1 + k_2 < j \le \tilde{q}_2\},\\
P_6 &= \{(i,j) | \tilde{p}_2 < i \le \tilde{p}_3, k_1 < j \le \tilde{q}_1\}.
\end{aligned}
$$

These sets of positions are depicted in Figure 2, where the grey shaded staircase area represents the forbidden positions. Let $N_1, N_2, \ldots, N_6$ be the number of

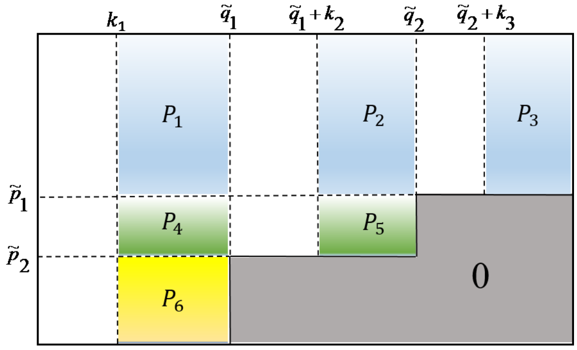

Figure 2: An interpretation of the nonnegativity of W.

1's contained in the position sets $P_1, P_2, \ldots, P_6$, respectively. It is clear that

$$
\begin{aligned}
N_1 + N_4 + N_6 &= \sum_{j=k_1+1}^{\tilde{q}_1} h_j,\\
N_2 + N_5 &= \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j,\\
N_3 &= \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j.
\end{aligned}
$$

On the other hand, one can easily verify that

$$
\begin{aligned}
&N_1+N_2+N_3 \geq \sum_{i=1}^{\tilde{p}_1}[r_i-(k_1+k_2+k_3)]^+,\\
&N_4+N_5 \geq \sum_{i=\tilde{p}_1+1}^{\tilde{p}_2}[r_i-(k_1+k_2)]^+,\\
&N_6 \geq \sum_{i=\tilde{p}_2+1}^{\tilde{p}_3}(r_i-k_1)^+.
\end{aligned}
$$

Combining the above observations, we have

$$
\begin{aligned}
\sum_{i=1}^{6} N_i &= \sum_{j=k_1+1}^{\tilde{q}_1} h_j + \sum_{j=\tilde{q}_1+k_2+1}^{\tilde{q}_2} h_j + \sum_{j=\tilde{q}_2+k_3+1}^{\tilde{q}_3} h_j\\
&\geq \sum_{i=1}^{\tilde{p}_1}[r_i-(k_1+k_2+k_3)]^+ + \sum_{i=\tilde{p}_1+1}^{\tilde{p}_2}[r_i-(k_1+k_2)]^+ + \sum_{i=\tilde{p}_2+1}^{\tilde{p}_3}(r_i-k_1)^+,
\end{aligned}
$$

which leads directly to the nonnegativity of $W(r,h,\lambda)$.

The characterization of nonemptyness of $\mathscr{A}_\lambda(r,h)$ in Theorem 4.1 gives a starting point in the study of such constrained $(0,1)$-matrix completion problems. A natural question is how to construct such a matrix $A \in \mathscr{A}_\lambda(r,h)$. In this regard, an effective algorithm has been proposed in [5] which suggests a sequential allocation of 1's from the last column to the preceding columns with priority given to the rows with larger remaining demands.

5. Applications

We apply the results to two engineering problems arising in smart grid and real-time systems.

5.1. Duration-deadline jointly differentiated energy services

A large penetration of uncertain renewable power such as wind and solar into the electric grid makes it difficult to maintain the balance of demand and supply of power. One way to help achieve balance is called demand response, which attempts to utilize the flexibility in demand to compensate for the uncertainty in supply. One form of demand response is the duration-deadline jointly differentiated energy service proposed in [8].

Suppose that renewable power is delivered over n time slots. The power available in time slot j is $h_j, j=1,2,\ldots,n$. Suppose there are m flexible loads, indexed by $i=1,2,\ldots,m$. Load i requires 1kW of power for a duration of r_i time slots delivered no later than the λ_ith time slot, with $r_i \leq \lambda_i$. The flexibility resides in the fact that any r_i time slots before the deadline λ_i will satisfy the requirement of load i. Thus the energy service is differentiated only by the duration and the deadline of the power delivery.

Without loss of generality, assume that the delivery deadlines satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Let τ be the number of distinct values among $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then there are τ different deadlines in total required by the flexible loads. Denote the demand profile by $r = \begin{bmatrix} r_1 & r_2 & \ldots & r_m \end{bmatrix}'$ and the supply profile by $h = \begin{bmatrix} h_1 & h_2 & \ldots & h_n \end{bmatrix}'$. Without loss of generality, assume the monotonicity of r and h as in (2).

A supply profile h is said to be adequate if there exists an allocation of power such that all the load requirements are satisfied. Further, a supply profile h is said to be exactly adequate if it is adequate and $\sum_{j=1}^{n} h_j = \sum_{i=1}^{m} r_i$, i.e., there is no excess supply after the allocation. A first basic question is to find the condition under which a given supply profile h is adequate. We first consider exact adequacy, the case when $\sum_{j=1}^{n} h_j = \sum_{i=1}^{m} r_i$. Observe that allocating the power to the flexible loads is equivalent to filling a $(0,1)$-matrix wherein the free positions form a Young diagram of shape $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$. Hence the supply profile h is exactly adequate if and only if $\mathscr{A}_\lambda(r,h)$ is nonempty. In view of Theorem 4.1, it follows that h is exactly adequate, if and only if the associated structure tensor $W(r,h,\lambda)$ is nonnegative. Note that the order of the structure tensor is determined by the number of different deadlines.

Now we remove the assumption $\sum_{j=1}^{n} h_j = \sum_{i=1}^{m} r_i$ and consider the general adequacy of the supply profile. Similar to the above analysis, a given supply profile h is adequate if and only if there exists a $(0,1)$-matrix wherein the free positions from a Young diagram of shape λ, the ith row sum is equal to $r_i, i = 1, 2, \ldots, m$, and the jth column sum is at most $h_j, j = 1, 2, \ldots, n$. Using the language of network flow, h is adequate if and only if the corresponding augmented network $\tilde{G}$ has a maximum flow that saturates all the edges into the sink node t but may not saturate all the edges out of the source node s. Since the value of the maximum flow is given by

$$\min_{k_1, k_2, \ldots, k_\tau} w_{k_1 k_2 \ldots k_\tau} + \sum_{i=1}^{m} r_i \tag{6}$$

as indicated in the proof of Theorem 4.1, it follows that the maximum flow saturates all the edges into the sink t if and only if $W(r,h,\lambda)$ is nonnegative.

In summary, a given supply profile h is adequate if and only if the associated structure tensor $W(r,h,\lambda)$ is nonnegative. If, moreover, $\sum_{j=1}^{n} h_j = \sum_{i=1}^{m} r_i$, then h is exactly adequate. For a given adequate supply profile h, one can then construct a feasible power allocation by exploiting the algorithm given in [5].

A natural follow-up question is to find the adequacy gap: In the event that the given supply profile h is inadequate, what is the minimum amount of supplementary purchase required to satisfy all the load requirements?

Let $g = \begin{bmatrix} g_1 & g_2 & \ldots & g_n \end{bmatrix}'$ be a nonnegative integer vector representing the supplementary purchase. Finding the adequacy gap amounts to solving the following optimization problem:

$$\begin{aligned}&\text{minimize } \sum_{j=1}^{n} g_j,\\ &\text{subject to } W(r, h{+}g, \lambda) \geq 0.\end{aligned}$$

Recall that the maximal flow in the augmented network $\tilde{G}$ is given by the quantity (6). In light of the one-one correspondence between a $(0,1)$-matrix and the associated network flow, this quantity also gives the largest number of 1's that can be contained in a $(0,1)$-matrix with row sums upper-bounded by $r_i, i = 1, 2, \ldots, m$, and column sums upper-bounded by $h_j, j = 1, 2, \ldots, n$. With this observation, one can reason that the minimum supplementary purchase required so as to meet all the demands is given by $|\min_{k_1,k_2,\ldots,k_\tau} w_{k_1 k_2 \ldots k_\tau}|$.

We wish to mention that the idea of duration-differentiated energy service was initiated in [19]. The authors therein assumed all the loads require the same delivery deadline. The results discussed above, when reduced to the case of one deadline, are consistent with those obtained in [19].

5.2. Job scheduling of preemptive multi-processor real-time systems

Consider the scheduling of a set of m independent jobs on d homogenous processors, where $d \geq 1$. Assume that all the jobs have synchronous arrival times. Each job J_i is then characterized by two parameters: a computation time r_i and a deadline λ_i, where r_i, λ_i are integers and $r_i \leq \lambda_i$. Assume that a processor can only execute one job at a time. Preemption is allowed in the sense that an executing job can be interrupted by another job at any time. We are interested in finding a feasible job schedule such that all the jobs can be finished before their respective deadlines. For a general introduction of real-time systems, one can refer to the book [6].

Without loss of generality, assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and let τ be the number of distinct values among $\lambda_1, \lambda_2, \ldots, \lambda_m$. Let $n = \lambda_1$. Denote the demand profile by $r = \begin{bmatrix} r_1 & r_2 & \ldots & r_m \end{bmatrix}'$ and supply profile by $h = d \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}'$.

Observe that a feasible job schedule exists if and only if there exists a $(0,1)$-matrix wherein the free positions from a Young diagram of shape $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, the ith row sum is $r_i, i = 1, 2, \ldots, m$, and the jth column sum is at most d for $j = 1, 2, \ldots, n$. Then, by the same argument as in the previous application, we know that a feasible job schedule exists if and only if the associated structure tensor $W(r, h, \lambda)$ is nonnegative leading to $(q_1{+}1)(q_2{+}1)\cdots(q_\tau{+}1)$ number of inequalities. What's more, the following theorem shows that by exploiting the flat nature of the supply profile h, one can further reduce the number of inequalities to simply τ, i.e., the number of distinct deadlines. Before stating the theorem, we introduce some notation. Let $A^\sharp$ be a $(0,1)$-matrix with row sum vector r such that all the 1's are put in the free positions as far to the right as possible. Denote by $r^\sharp$ the column sum vector of $A^\sharp$. It is easy to see that the entries of $r^\sharp$ satisfy the following monotonicity:

$$r^\sharp_{\tilde{q}_i+1} \le r^\sharp_{\tilde{q}_i+2} \le \cdots \le r^\sharp_{\tilde{q}_{i+1}}, i = 0, 1, 2, \ldots, \tau - 1.$$

Theorem 5.1. *There exists a feasible job schedule if and only if*

$$ld \ge \sum_{j=1}^{l} r^\sharp_j \tag{7}$$

holds for each $l = \tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_\tau$.

Proof. From the above analysis, we know that a feasible job schedule exists if and only if $W(r, h, \lambda) \ge 0$. By some straightforward yet tedious derivation, one can verify that

$$\min_{\substack{k_1+k_2+\cdots+k_\tau=n-l \\ 0\le k_1\le q_1, 0\le k_2\le q_2, \ldots, 0\le k_\tau \le q_\tau}} w_{k_1 k_2 \ldots k_\tau} = ld - \sum_{j=1}^{l} r^\sharp_j$$

holds for $l = 0, 1, 2, \ldots, n$. This yields that a feasible job schedule exists if and only if the inequality (7) holds for $l = 1, 2, \ldots, n$. Comparing with the result stated in the theorem, it remains to show that the inequality (7) holds for $l = 1, 2, \ldots, n$ if it holds for $l = \tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_\tau$.

First, it is rather straightforward to verify that if the inequality (7) holds for $l = \tilde{q}_1$, then it holds for $l = 1, 2, \ldots, \tilde{q}_1$, since the supply profile is pure flat and $r^\sharp_1 \le r^\sharp_2 \le \cdots \le r^\sharp_{\tilde{q}_1}$.

In what follows, we shall show that if the inequality (7) holds for $l = \tilde{q}_1$ and $l = \tilde{q}_2$, then it holds for $l = \tilde{q}_1 + 1, \tilde{q}_1 + 2, \ldots, \tilde{q}_2$. We show this by contradiction. Assume that $ld < \sum_{j=1}^{l} r^\sharp_j$ for some l, where $\tilde{q}_1 + 1 \le l < \tilde{q}_2$. Then one of the following two scenarios will occur.

Scenario 1: $(\tilde{q}_2 - l)d \le \sum_{j=l+1}^{\tilde{q}_2} r^\sharp_j$.

Scenario 2: $(\tilde{q}_2 - l)d > \sum_{j=l+1}^{\tilde{q}_2} r^\sharp_j$.

In scenario 1, we have $\tilde{q}_2 d < \sum_{j=1}^{\tilde{q}_2} r^\sharp_j$ which results in a contradiction. In scenario 2, since $(\tilde{q}_2 - l)d > \sum_{j=l+1}^{\tilde{q}_2} r^\sharp_j$ and $r^\sharp_{\tilde{q}_1+1} \le r^\sharp_{\tilde{q}_1+2} \le \cdots \le r^\sharp_{\tilde{q}_2}$, it follows that $(l - \tilde{q}_1)d > \sum_{j=\tilde{q}_1+1}^{l} r^\sharp_j$ yielding $\tilde{q}_1 d < \sum_{j=1}^{\tilde{q}_1} r^\sharp_j$ which also results in a contradiction.

Finally, by a simple analogy, one can show that if the inequality (7) holds for $l = \tilde{q}_i$ and $l = \tilde{q}_{i+1}$, then it holds for $l = \tilde{q}_i + 1, \tilde{q}_i + 2, \ldots, \tilde{q}_{i+1}$, where $i = 2, 3, \ldots, \tau - 1$. □

Note that when the condition in Theorem 5.1 holds, a feasible job schedule can be found by applying the algorithm proposed in [5].

6. Conclusion

In this paper, we study a type of constrained $(0,1)$-matrix completion problems wherein the free positions form a Young diagram. The main contribution lies in a nice and simple solvability characterization in terms of the nonnegativity of a structure tensor. To illustrate the effectiveness of our results, we apply them to two engineering problems arising in smart grid and real-time systems, respectively.

Acknowledgment

The work in this paper was partially supported by National Science Foundation under Grant 1331692, Research Grants Council of Hong Kong Special Administrative Region, China, under the Theme-Based Research Scheme T23-701/14-N and Hong Kong PhD Fellowship. The authors also wish to thank Prof. Kameshwar Poolla of University of California at Berkeley, Prof. Matias N. Pincetic of Pontificia Universidad Catolica de Chile, Dr. He Hao of Pacific Northwest National Laboratory, Prof. Karl Johansson of The Royal Institute of Technology (KTH), and Prof. Tryphon Georgiou, Dr. Seizhen Khong, and Dr. Yongxin Chen of University of Minnesota for valuable discussions.

References

[1] R. P. Anstee, "Properties of a class of $(0,1)$-matrices covering a given matrix", *Canad. J. Math.*, vol. 34, pp. 438-453, 1982.

[2] R.A. Brualdi, "Matrices of zeros and ones with fixed row and column sum vectors", *Linear Algebra Appl.*, vol. 33, pp. 159-231, 1980.

[3] R. A. Brualdi, *Combinatorial Matrix Classes*, Cambridge: Cambridge University Press, 2006.

[4] R. A. Brualdi and G. Dahl, "Matrices of zeros and ones with given line sums and a zero block", *Linear Algebra Appl.*, vol. 371, pp. 191-207, 2003.

[5] R. A. Brualdi and G. Dahl, Constructing $(0,1)$-matrices with given line sums and certain fixed zeros. in G. T. Herman and A. Kuba (Eds.) *Advances in Discrete Tomography and Its Applications*, Boston: Birkhaüser, 2007.

[6] G. C. Buttazzo, *Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications*, 3rd edition, Springer, 2011.

[7] W. Y. C. Chen, "Integral matrices with given row and column sums", *J. Combinatorial Theory, Series A*, vol. 61, pp. 153-172, 1992.

[8] W. Chen, L. Qiu, and P. Varaiya, "Duration-deadline jointly differentiated energy services", in *Proc. 54th IEEE Conf. Decision Contr.*, pp. 7220-7225, Osaka, Japan, 2015.

[9] Y. Chen, "Simple existence conditions for zero-one matrices with at most one structural zero in each row and column", *Discrete Mathematics*, vol. 306, pp. 2870-2877, 2006.

[10] L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton: Princeton University Press, 1962.

[11] D. R. Fulkerson, "Zero-one matrices with zero trace", *Pacific J. Math.*, vol. 10, pp. 831-836, 1960.

[12] D. Gale, "A theorem on flows in networks", *Pacific J. Math.*, vol. 7, pp. 1073-1082, 1957.

[13] H. Hao and W. Chen, "Characterizing flexibility of an aggregation of deferrable loads", in *Proc. 53rd IEEE Conf. Decision Contr.*, pp. 4059-4064, Los Angeles, CA, 2014.

[14] G. T. Herman, A. Kuba (Eds.), *Discrete Tomography. Foundations, Algorithms, and Applications*, Birkhäuser, Boston, MA, 1999.

[15] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, 2nd edition, Springer, 2002.

[16] I. Lari, F. Ricca, and A. Scozzari, "Bidimensional allocation of seats via zero-one matrices with given line sums", *Ann. Oper. Res.*, vol. 215, pp. 165-181, 2014.

[17] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theorey of Majorization and Its Applications*, 2nd ed., Springer, New York, 2011.

[18] L. Mirsky, "Combinatorial theorems and integral matrices", *J. Combinatorial Theory*, vol. 5, pp. 30-44, 1968.

[19] A. Nayyar, M. N.-Pincetic, K. Poolla and P. Varaiya, "Duration-differentiated energy services with a continuum of loads", *IEEE Trans. Control Network Syst.*, vol. 3, no. 2, pp. 182-191, 2016.

[20] H. J. Ryser, "Combinatorial properties of matrices of zeros and ones", *Pacific J. Math.*, vol. 7, pp. 1073-1082, 1957.

[21] H. J. Ryser, "Traces of matrices of zeros and ones", *Canad. J. Math.*, vol. 12, pp. 463-476, 1960.