Abstract

If $D$ is a partially filled-in $(0, 1)$-matrix with a unique completion to a $(0, 1)$-matrix $M$ (with prescribed row and column sums), we say that $D$ is a defining set for $M$. Let $A_{2m,m}$ be the set of $(0, 1)$-matrices of dimensions $2m \times 2m$ with uniform row and column sum $m$. It is shown in (Cavenagh, 2013) that the smallest possible size for a defining set of an element of $A_{2m,m}$ is precisely $m^2$. In this note when $m$ is a power of two we construct an element of $A_{2m,m}$ which has no defining set of size less than $2m^2 - o(m^2)$. Given that it is easy to show any $A_{2m,m}$ has a defining set of size at most $2m^2$, this construction is asymptotically optimal. Our construction is based on a array, defined using linear algebra, in which any subarray has asymptotically the same number of 0’s and 1’s.

Keywords: $(0, 1)$-matrix, defining set, frequency square, $F$-square, Gale-Ryser Theorem. MSC2010: 05B20

1 Introduction

Where convenient, we keep notation consistent with [1].
Let \( R = (r_1, r_2, \ldots, r_m) \) and \( S = (s_1, s_2, \ldots, s_n) \) be vectors of non-negative integers such that \( \sum_{i=1}^{m} r_i = \sum_{j=1}^{n} s_j \). Then \( A(R, S) \) is defined to be the set of all \( m \times n \) \((0, 1)\)-matrices with \( r_i \) 1's in row \( i \) and \( s_j \) 1's in column \( j \), where \( 1 \leq i \leq m \) and \( 1 \leq j \leq n \). If \( M \in A(R, S) \), we refer to \( R \) and \( S \) as the row sum and column sum vectors for \( M \), respectively.

With \( R \) and \( S \) as above, we next define \( A'(R, S) \) to be the set of all \( m \times n \) \((0, 1, \star)\)-matrices with:

1. at most \( r_i \) 1’s in row \( i \),
2. at most \( n - r_i \) 0’s in row \( i \),
3. at most \( s_j \) 1’s in column \( j \),
4. at most \( m - s_j \) 0’s in column \( j \).

We call a matrix \( M \in A'(R, S) \) a partial \((0, 1)\)-matrix with row sum vector \( R \) and column sum vector \( S \). Indeed, \( A(R, S) \subseteq A'(R, S) \) and a partial \((0, 1)\)-matrix is a \((0, 1)\)-matrix if and only if it none of its positions are empty (equal to \( \star \)). (Note that our definition of a partial \((0, 1)\)-matrix allows for the possibility of a matrix \( M \in A'(R, S) \) which has no completion to a \((0, 1)\)-matrix in \( A(R, S) \).)

We sometimes consider a partial \((0, 1)\)-matrix \( M \) as a set of triples

\[
M = \{(i, j, M_{ij}) \mid 1 \leq i \leq m, 1 \leq j \leq n, M_{ij} \in \{0, 1\}\}.
\]

Note that we naturally omit here the empty positions of \( M \). It is usually clear by context whether we are considering a partial \((0, 1)\)-matrix as a matrix or as a set of ordered triples. For example, if we say that \( M_1 \subseteq M_2 \) or write \( M_1 \setminus M_2 \) for two partial \((0, 1)\)-matrices \( M_1 \) and \( M_2 \), we are considering \( M_1 \) and \( M_2 \) as sets. Similarly, the size of a partial \((0, 1)\)-matrix \( M \) refers to \(|M|\) where \( M \) is a set (i.e. the number of 0’s and 1’s).

With this in mind, suppose that \( M \in A(R, S) \) and \( D \in A'(R, S) \). We say that \( D \) is a defining set for \( M \) if \( M \) is the unique member of \( A(R, S) \) such that \( D \subseteq M \).

This is analogous to the usual definition of defining sets of Latin squares and other combinatorial designs ([4, 8]).

Given a \((0, 1)\)-matrix \( M \), the size of the smallest defining set in \( M \) is denoted by \( \text{sds}(M) \). For given row column sum vectors \( R \) and \( S \), \( \text{sds}(A(R, S)) \)
is the size of the smallest defining set for all members of the set $\mathcal{A}(R, S)$. More precisely,

$$sds(\mathcal{A}(R, S)) = \min\{sds(M) \mid M \in \mathcal{A}(R, S)\}.$$ 

We also define

$$\maxsds(\mathcal{A}(R, S)) = \max\{sds(M) \mid M \in \mathcal{A}(R, S)\}.$$ 

This last definition is the focus of this paper, in the case when row and column sums are constant. To this end, we define $R_{n,x}$ to be the row sum vector with dimension $n$ and constant row sum $r_1 = r_2 = \cdots = r_n = x \leq n$. The column sum vector $S_{n,x}$ is defined similarly. Then, $\mathcal{A}_{n,x} = \mathcal{A}(R_{n,x}, S_{n,x})$ is the set of $n \times n (0,1)$-matrices with constant row and column sum $x$.

In fact, elements of $\mathcal{A}_{n,x}$ may also be thought of as frequency squares (sometimes $F$-squares). Let $n, \alpha, \lambda_1, \lambda_2, \ldots, \lambda_\alpha \in \mathbb{N}$ and $\sum_{i=1}^n \lambda_i = n$. A frequency square or $F$-square $F(n; \lambda_1, \lambda_2, \ldots, \lambda_\alpha)$ of order $n$ is an $n \times n$ array on symbol set $\{s_1, s_2, \ldots, s_\alpha\}$ such that each cell contains one symbol and symbol $s_i$ occurs precisely $\lambda_i$ times in each row and $\lambda_i$ times in each column. Thus if we let $\alpha = 2, s_1 = 1$ and $s_2 = 0$, the frequency square $F(n; x, n-x)$ is in effect an element of $\mathcal{A}_{n,x}$.

Critical and defining sets of frequency squares have previously been studied in [5]. The following results are directly implied by Theorems 2, 3, 4, 5 of [5].

**Theorem 1.** ([5]) $sds(\mathcal{A}_{n,1}) = n - 1$. $sds(\mathcal{A}_{n,2}) \leq 2n - 3$ and $sds(\mathcal{A}_{n,2}) = 2n - 4$ if $n$ is even. $sds(\mathcal{A}_{n,n}) \leq xn - x^2$ if $x$ divides $n$ and $x < n$. If $x \leq k$ then $sds(\mathcal{A}_{x+k+1,x}) \leq (k-1)x^2 + x(x+1)/2$.

In particular, observe that the above results imply that $sds(\mathcal{A}_{2m,m}) \leq m^2$, for each integer $m$. In [3] some lower bounds for $sds(\mathcal{A}_{n,x})$ are given, showing in particular that $sds(\mathcal{A}_{2m,m})$ is in fact equal to $m^2$.

**Theorem 2.** ([3]) Any defining set $D$ in a matrix from $\mathcal{A}_{n,x}$ has size at least $\min\{x^2, (n-x)^2\}$.

**Corollary 3.** ([3]) $sds(\mathcal{A}_{2m,m}) = m^2$.

In this paper we will ultimately prove the following.

**Theorem 4.** If $m$ is a power of two, $\maxsds(\mathcal{A}_{2m,m}) = 2m^2 - O(m^{7/4})$. 

3
Since taking every occurrence of the symbol 1 in a $(0, 1)$-matrix always forms a defining set, each element of $\mathcal{A}_{2m,m}$ has a defining set of size $2m^2$. In fact, fixing a cell containing 1, it is not hard to show that each 1 in the same row and column can be omitted, retaining the property of being a defining set. Thus $\text{maxsds}(\mathcal{A}_{2m,m}) \leq 2m^2 - 2m + 1$. It follows that, at least when $m$ is a power of 2, $\text{maxsds}(\mathcal{A}_{2m,m}) = 2m^2 - o(m^2)$.

The analogous question has been considered for Latin squares in [7], where it is shown that every Latin square of order $n$ has a defining set of size at most $n^2 - \sqrt{\pi/2} n^{9/6}$ and that for each $n$ there exists a Latin square $L$ with no defining set of size less than $n^2 - (e + o(1))n^{10/6}$. In contrast to the proof in this paper, the latter result is non-constructive.

2 Theory on trades and defining sets in $(0, 1)$-matrices

In this section we develop the theory from [3] which is relevant to our paper. The results on trades in this section are a restatement of theory in [1]; however results on defining sets are new. We define a trade to be a non-empty partial $(0, 1)$-matrix $T$ such that there exists a disjoint mate $T'$ such that:

- $T_{ij}$ is empty if and only if $T'_{ij}$ is empty;
- if $T_{ij}$ is non-empty, then $T_{ij} \neq T'_{ij}$;
- if 1 appears precisely $k$ times in a row $r$ (column $c$) of $T$, then 1 also appears $k$ times in row $r$ (column $c$) of $T'$;
- if 0 appears precisely $k$ times in a row $r$ (column $c$) of $T$, then 0 also appears $k$ times in row $r$ (column $c$) of $T'$;

**Lemma 5.** ([3]) A partial $(0, 1)$-matrix $D$ is a defining set for a $(0, 1)$-matrix $M$ if and only if $D \subseteq M$ and $|D \cap T| \geq 1$ for every trade $T \subseteq M$.

Thus we can study the properties of defining sets of $(0, 1)$-matrices through an analysis of the trade structure of $(0, 1)$-matrices. We say that a trade $T$ is a cycle if each row and each column of $T$ contains either 0 or 2 non-empty positions.

The notions of cycle and intercalate are very similar to the notions of minimal balanced matrix and interchange (respectively) given in [1]; however
in [1] these matrices are formed as $(0 \pm 1)$-matrices (with the 0’s denoting “empty” cells) rather than $(0, 1, \ast)$-matrices. For our purposes it is helpful to define trades as subsets of $(0, 1)$-matrices, hence our choice of definitions.

By similar reasoning to 3.2 of [1] however, the following can be shown.

**Theorem 6.** (Lemma 3.2.1 of [1]) Any trade $T$ in a $(0, 1)$-matrix is a union of disjoint cycles.

**Lemma 7.** [3] Let $P$ be a finite, non-empty partial $(0, 1)$-matrix such that every non-empty row or column contains at least one 0 and at least one 1. Then $P$ contains a trade.

We say that matrix $M \in \mathcal{A}'(R, S)$ is in good form if whenever $(i, j, 0)$, $(i, j', 1) \in M$, then $j < j'$ and whenever $(i, j, 0)$, $(i', j, 1) \in M$, then $i < i'$. Somewhat informally, a partial matrix $M \in \mathcal{A}'(R, S)$ is in good form if and only if a South-East walk $C$ exists with only 1’s below the line and only 0’s above the line.

The following theorem is equivalent to the Gale-Ryser theorem [6, 9] and Theorem 3.2.4 ([2, 9, 10]) in [2].

**Theorem 8.** A matrix $M \in \mathcal{A}(R, S)$ is the unique member of $\mathcal{A}(R, S)$ if and only if its rows and columns can be rearranged so that it is in good form.

In Figure 1, the unique member of $\mathcal{A}((2, 3, 4), (0, 0, 1, 2, 3))$ is given, with the South-East walk $C$ shown as a thick line.

We next give a new classification of defining sets in $(0, 1)$-matrices that will be useful for our purposes.
Theorem 9. The set $D$ is a defining set of a $(0,1)$-matrix $M$ if and only if $D \subset M$ and the rows and columns of $M \setminus D$ can be rearranged so that $M \setminus D$ is in good form.

Proof. Suppose first that $D$ is a subset of $M$ such that the rows and columns of $M \setminus D$ can be arranged so that a South-East walk $C$ exists such that there are only 1’s below $C$ and only 0’s above $C$. If there exists a trade $T$ which is a subset of $M \setminus D$, then $T$ is also a subset of any superset of $M \setminus D$. In particular, $T$ is a subset of the $(0,1)$-matrix $M'$ created by placing a 1 in each cell below $C$ and a 0 in each cell above $C$. But $M'$ has no trades by Theorem 8, a contradiction.

Conversely, let $D$ be a defining set of a $(0,1)$-matrix $M$. Consider $M_0 := M \setminus D$. By Lemma 5, $M_0$ contains no trades. We obtain a non-increasing sequence $M_0, M_1, \ldots$ via an iterative process. Given $M_k$, where $k \geq 0$, rearrange the rows of $M_k$ so that rows containing only 0 are contiguously the first rows and that any rows of $M_k$ containing only 1 are contiguously the last rows. Next, rearrange the columns of $M_k$ so that the columns containing only 0 are contiguously the last columns and that any columns of $M_k$ containing only 1 are contiguously the last rows. Let $M_{k+1}$ be the partial $(0,1)$-matrix obtained by deleting from $M_k$ the above rows and columns (i.e. any rows or columns of $M_k$ that do not contain both 0 and 1).

If $M_{k+1} = M_k$ and is not empty, then every row and column of $M_k$ contains both 0 and 1. Thus $M_k$ contains a trade by Lemma 7, a contradiction. Thus we have a sequence $M_0, M_1, \ldots, M_K$ where $M_K$ is empty. Next, reconstruct $M \setminus D$ via nesting the above matrices in reverse; observe that via this process we have rearranged the rows and columns of $M \setminus D$ so that it is in good form. □

3 A matrix in which 0’s and 1’s are closely balanced

In this section we show the existence of a $(0,1)$-matrix with the property that within any rectangular subarray the difference between the number of 1’s and 0’s is small.

Let $k \geq 2$. Let $V = (v_1, v_2, \ldots, v_{2^k-1})$ be a fixed vector with entries from $\mathbb{Z}_2$. Construct a $(2^k - 1) \times 2^k (0,1)$-matrix $M := M(V)$ as follows.
Let \( W := \{W_1, W_2, \ldots, W_{2^k-1}\} \) be some ordering of the set of all non-zero column vectors of dimension \( k \) over \( \mathbb{Z}_2 \). Label the columns of \( M \) with the elements of \( W \cup \{W_0\} \) where \( W_0 \) is the zero vector. Row \( i \) of \( M \) corresponds to equation over \( \mathbb{Z}_2 \) of the form
\[
(x_1, x_2, \ldots, x_k) \cdot W_i = v_i. \tag{1}
\]
Then, place a 1 in row \( i \) of column \( W_j \) if and only if \( W_j \) is a solution to the equation corresponding to row \( i \); otherwise place a 0.

For each \( j \), let \( y_j \) be the number of 1’s in column \( W_j \) and \( z_j \) be the number of 0’s in column \( W_j \), with \( \Delta_j := y_j - z_j \). Finally, define \( \Delta := 2^{k-1} \sum_{j=0}^{2^k-1} |\Delta_j| \), noting that \( \Delta \) is a function of \( V \).

**Example 10.** Let \( k = 3 \) and \( V = (v_1, v_2, v_3, v_4, v_5, v_6, v_7) = (0, 1, 1, 0, 1, 1) \). Then \( M(V) \) is given below, where the columns are labelled \( W_0, W_1, \ldots, W_7 \) and each row, \( 1 \leq i \leq 7 \), is labelled as in Equation (1) above. Observe that
\[
(\Delta_1, \Delta_2, \ldots, \Delta_7) = (-3, -3, 5, 1, -3, 1, 1)
\]
and \( \Delta = 18 \).

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 = 0 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( x_2 = 1 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_3 = 1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( x_1 + x_2 = 1 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( x_1 + x_3 = 0 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( x_2 + x_3 = 1 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( x_1 + x_2 + x_3 = 1 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Theorem 11.** For each choice of \( V \), \( \Delta(V) \leq m^{3/2} \), where \( m = 2^k \).

**Proof.** Our proof uses elementary probability theory. We fix \( V \) but choose a column \( W_j \) uniformly at random, for each row \( i \), let \( A_i := 1 \) if the entry of \( M \) in row \( i \) and column \( j \) is 1; otherwise \( A_i := -1 \).
Observe that $A_i$ is a random variable with $\Pr\{A_i = 1\} = \Pr\{A_i = -1\} = 1/2$, $E[A_i] = 0$ and $\text{Var}(A_i) = E[A_i^2] = 1$. In fact, for each $i \neq j$, $A_i$ and $A_j$ are independent events. To see this, the equations corresponding to $A_i$ and $A_j$ each have $2^{k-1}$ solutions; whereas the linear system corresponding to $A_i$ and $A_j$ has $2^{k-2}$ solutions (since the equations corresponding to each pair of rows are linearly independent). Thus $\Pr\{A_i = 1, A_j = 1\} = \Pr\{A_i = 1\} \Pr\{A_j = 1\} = 1/4$. As an aside, it is not always true that three or more of these random variables are independent as a subset (in particular if the subset of rows is inconsistent as a linear system); but pairwise independence is enough for our purposes.

Next, $D_j := \sum_{i=1}^{2^k-1} A_i$. From above, $E[D_j] = 0$ and, from pairwise independence, $\text{Var}(D_j) = 2^k - 1$. However, we can also calculate the variance of $D_j$ by considering all possible columns:

$$\text{Var}(D_j) = E[D_j^2] = \sum_{j=1}^{2^k} \Delta_j^2 / 2^k.$$  

Thus

$$\sum_{j=1}^{2^k} \Delta_j^2 = m(m - 1) < m^2.$$  

The result follows. \hfill \Box

Let $m = 2^k$ and $n = 2m$ where $k \geq 2$. We now define an $n \times n$ $(0, 1)$-matrix $B \in A_{2m,m}$ which we will in turn show cannot have a small defining set. Let $Y := \{Y_1, Y_2, \ldots, Y_{2^k+1}\}$ be some ordering of all the vectors of dimension $k+1$ over $\mathbb{Z}_2$. Label the columns of $B$ with the elements of $Y$. The rows of $B$ are labelled with all equations over $\mathbb{Z}_2$ of the form

$$(x_1, x_2, \ldots, x_k) \cdot W' = a,$$

where $W' \in W$ and $a = \{0, 1\}$. This defines $2(2^k - 1) = 2^{k+1} - 2$ rows; the remaining two rows correspond to $x_{2^{k+1}} = 0$ and $x_{2^{k+1}} = 1$. As above, we place a 1 in column $j$ and row $i$ whenever the vector corresponding to column $j$ is a solution to the equation corresponding to row $i$. We immediately have that each row of $B$ has precisely $m$ 0’s and $m$ 1’s and each column of $B$ has precisely $m$ 0’s and $m$ 1’s. It is also immediate that each row of $B$ has a complement row; formed by replacing each entry $e \in \{0, 1\}$ with $1 - e \in \{1, 0\}$.
Example 12. Let $k = 2$. Then $B$ is given below.

\[
\begin{array}{cccccccc}
0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
\end{array}
\]

\[
\begin{array}{cccccccc}
x_1 = 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
x_2 = 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\
x_1 + x_2 = 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
x_3 = 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
x_1 = 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
x_2 = 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
x_1 + x_2 = 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
x_3 = 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
\end{array}
\]

Given any subset $R$ of the rows of $B$ and subset $C$ of the columns, the subarray $B(R, C)$ is the $|R| \times |C|$ array of $B$ induced by $R$ and $C$ and $\delta(B(R, C))$ is given by the total number of 1’s in $B(R, C)$ minus the total number of 0’s in $B(R, C)$.

We are now ready to apply Theorem 11.

Lemma 13. Let $R$ and $C$ be any subsets of the rows and columns, respectively, of $B$. Then $|\delta(B(R, C))| \leq 2m^{3/2} + m$.

Proof. Without loss of generality assume that $\delta(B(R, C)) \geq 0$.

If $|R| > m$, there are at least $|R| - m$ distinct \{row, complement row\} pairs in $R$. For each such pair, delete the row with more 0’s than 1’s (or either row if they each have the same number of 0’s and 1’s). The resultant $B(R', C)$ clearly has the property $\delta(B(R', C)) \geq \delta(B(R, C))$.

Next, if $|R| < m$, there are at least $m - |R|$ distinct \{row, complement row\} pairs not in $R$. For each such pair, add the row with more 1’s than 0’s (or either row if they each have the same number of 1’s and 0’s). The resultant $B(R', C)$ has the property $\delta(B(R', C)) \geq \delta(B(R, C))$.

Thus we may assume that $|R| = m$. If there exists a row in $R$ whose complement row is also in $R$, there exists another \{row, complement\} pair not in $R$. Again, we may replace a row from $R$ with one not in $R$ so that $\delta(B(R, C))$ is not decreased. Repeat this until $R$ intersects every \{row, complement\} pair.

Next, remove the unique row from $B(R, C)$ corresponding to the equation containing $x^{2k+1}$. We now have a matrix which is based on a matrix $M(V)$ for some $V$, where each column of $M(V)$ is included either once, twice, or
not at all. (The vector $V$ is determined by the constant terms in each of the equations corresponding to rows of $R$.) The result then follows from Theorem 11.

Finally, Theorem 4 is implied by the following theorem.

**Theorem 14.** Let $D$ be a defining set in the $n \times n$ $(0,1)$-matrix $B$. Then $|D| \geq n^2/2 - O(n^{7/4})$.

**Proof.** Let $n = 2^{k+1}$. Since we are obtaining a lower bound for the size of $D$, we may assume that $D$ is a minimal defining set. From Theorem 9, the rows and columns of $D$ (and $B$ in correspondence) can be arranged so that a South-East walk $C$ can be drawn in $B \setminus D$ with only 1’s below $C$ and only 0’s above $C$. Indeed since $D$ is minimal, $B \setminus D$ contains every occurrence of 1 from $B$ below $C$ and every occurrence of 0 from $B$ above $C$.

Let $\alpha_0$ and $\alpha_1$ be the number of 0’s and 1’s (respectively) in $B$ below $C$, with $\beta_0$ and $\beta_1$ the number of 0’s and 1’s (respectively) in $B$ above $C$. Then:

$$\alpha_0 + \beta_0 = \alpha_1 + \beta_1 = 2m^2$$

and $|D| = \alpha_0 + \beta_1$.

Our next aim is to find an upper bound for $|\alpha_1 - \alpha_0|$. Let $K = \lceil (k+1)/4 \rceil$. Create partitions $R = \{R_1, R_2, \ldots, R_{2^K}\}$ and $C = \{C_1, C_2, \ldots, C_{2^K}\}$ of the rows and columns so that each subset is contigious and of size $2^{k+1-K} \leq 2^{3K}$. Each $R_i \in R$ and $C_j \in C$ induces a block; that is a subarray created by the intersection of the rows from $R_i$ and the columns from $C_j$. Observe that at most $2^{K+1}$ blocks contain both 0 and 1 within $B \setminus D$.

Let $S$ be the set of blocks which contain only 1 in $B \setminus D$. Our aim is to show that $S$ can be partitioned into at most $2^K$ rectangles (a rectangle here is a set of contigious blocks forming a rectangle shape). We first remove the largest such rectangle possible from $S$ contained in the last $n/2$ rows. Next, we remove the largest such rectangle from the last $n/4$ rows and from rows $n/4 + 1$ to $n/2$. At step $i$, we remove $2^{i-1}$ rectangles, each contained within a set of $n/2^i$ rows, specifically the sets of rows:

$$\left\{ \frac{n}{2^j} + \frac{n}{2^i} + k \mid 1 \leq k \leq \frac{n}{2^i} \right\}$$

where $0 \leq j \leq 2^{i-1} - 1$. At the final step, $i = K$. In total we have removed at most $\sum_{i=1}^{K} 2^{i-1} = 2^K - 1$ rectangles.
Moreover these rectangles include every block strictly below \( C \). From Corollary 13, the difference between the number of 1’s and 0’s in \( B \) within each such rectangle is at most \( 2m^{3/2} + m < n^{3/2} \). Thus over all of the rectangles we have an upper bound for the difference between the number of 1’s and 0’s of \( O(n^{7/4}) \). The difference between the the number of 1’s and 0’s in an individual block in \( B \) is bounded by the size of that block, so for the blocks intersecting \( C \) we have a net upper bound for the difference between the number of 1’s and 0’s of \( 2^{K+1} \times (2^{3K})^2 = O(n^{7/4}) \). Thus

\[
|\alpha_1 - \alpha_0| = |\beta_1 - \beta_0| = O(n^{7/4}).
\]

It follows that

\[
\alpha_0 - 2m^2 = -\beta_0 \geq -\beta_1 - O(n^{7/4})
\]

and \( |D| = \alpha_0 + \beta_1 \geq n^2 - O(n^{7/4}) \).

\[\square\]

References


