Exponential Length of Certain Sequences of Subsets

Kyriakos Kalorkoti[1]

§1. Introduction. Let n be a non-negative integer and $A = \{1, 2, \ldots, n\}$. For a sequence $S_1, S_2, \ldots, S_k$ of subsets of A define $\mathrm{Odd}(S_1, S_2, \ldots, S_k)$ to be the set of elements of A that are members of an odd number of sets in the sequence. We are interested in the largest length of sequences $A_1, A_2, \ldots, A_m$ of subsets of A such that:

$$A_i \not\subseteq A_j \cup \mathrm{Odd}(A_{i+1}, A_{i+2}, \ldots, A_{j-1}) \tag{$\dagger$}$$

for all i, j with $1 \le i < j \le m$. We will show that there are such sequences with

$$m = 2^{\lfloor n/2 \rfloor}.$$

§2. Encoding. We will use the familiar bit vector representation of subsets of A. Thus a subset A_i is represented by the vector

$$(\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{in})$$

where

$$\epsilon_{ij} = \begin{cases} 0 & \text{if } j \notin A_i; \\ 1 & \text{if } j \in A_i. \end{cases}$$

We will work over $\mathbb{Z}_2$ and regard the ϵ_{ij} as elements of this field. A sequence of sets $A_1, A_2, \ldots, A_m$ is encoded as a matrix

$$\begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \ldots, & \epsilon_{1n} \\ \epsilon_{21} & \epsilon_{22} & \ldots, & \epsilon_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{m1} & \epsilon_{m2} & \ldots & \epsilon_{mn} \end{pmatrix}.$$

Such a matrix corresponds to a valid sequence of subsets of A (i.e., a sequence with the property ($\dagger$)) if and only if there are column indices c_{rs} such that the following equations have a solution over $\mathbb{Z}_2$:

$$\begin{array}{rlrlrl}
\epsilon_{1c_{11}} & = 1, & \epsilon_{2c_{11}} & = 0, & & \\
\epsilon_{1c_{12}} & = 1, & \epsilon_{2c_{12}} & = 0, & \epsilon_{3c_{12}} & = 0, \\
\epsilon_{1c_{13}} & = 1, & \epsilon_{2c_{13}} + \epsilon_{3c_{13}} & = 0, & \epsilon_{4c_{13}} & = 0, \\
\vdots & & \vdots & & \vdots & \\
\epsilon_{1c_{1,m-1}} & = 1, & \epsilon_{2c_{1,m-1}} + \epsilon_{3c_{1,m-1}} + \cdots + \epsilon_{m-1,c_{1,m-1}} & = 0, & \epsilon_{mc_{13}} & = 0, \\
\hline
\epsilon_{2c_{21}} & = 1, & \epsilon_{3c_{21}} & = 0, & & \\
\epsilon_{2c_{22}} & = 1, & \epsilon_{3c_{22}} & = 0, & \epsilon_{4c_{22}} & = 0, \\
\epsilon_{2c_{23}} & = 1, & \epsilon_{3c_{23}} + \epsilon_{4c_{23}} & = 0, & \epsilon_{5c_{23}} & = 0, \\
\vdots & & \vdots & & \vdots & \\
\epsilon_{2c_{2,m-1}} & = 1, & \epsilon_{3c_{2,m-1}} + \epsilon_{4c_{2,m-1}} + \cdots + \epsilon_{m-1,c_{2,m-1}} & = 0, & \epsilon_{mc_{23}} & = 0, \\
\hline
\vdots & & \vdots & & \vdots & \\
\hline
\epsilon_{m-2,c_{m-2,1}} & = 1, & \epsilon_{m-1,c_{m-2,1}} & = 0, & & \\
\epsilon_{m-2,c_{m-2,2}} & = 1, & \epsilon_{m-1,c_{m-2,2}} & = 0, & \epsilon_{mc_{m-2,2}} & = 0, \\
\hline
\epsilon_{m-1,c_{m-1,1}} & = 1, & \epsilon_{m,c_{m-1,1}} & = 0. & &
\end{array}$$

[1]School of Informatics, University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, Scotland. Email: kk@inf.ed.ac.uk

§3. Rephrasing the problem. The problem states that we are given n, the number of elements, and wish to find m, the length of the longest sequence of subsets satisfying the condition (†). We can reverse this so that we are given m and wish to find the smallest n for which a sequence exists. To do this we restate the equations of the previous section by dropping the second subscript of each ϵ_{ij}. We now obtain the following system:

$$
\begin{array}{l r r r}
\text{Row Conditions 1:} & \epsilon_1 = 1, & \epsilon_2 = 0; & \\
 & \epsilon_1 = 1, & \epsilon_2 = 0, & \epsilon_3 = 0; \\
 & \epsilon_1 = 1, & \epsilon_2 + \epsilon_3 = 0, & \epsilon_4 = 0; \\
 & \vdots & \vdots & \vdots \\
 & \epsilon_1 = 1, & \epsilon_2 + \epsilon_3 + \cdots + \epsilon_{m-1} = 0, & \epsilon_m = 0; \\
\hline
\text{Row Conditions 2:} & \epsilon_2 = 1, & \epsilon_3 = 0; & \\
 & \epsilon_2 = 1, & \epsilon_3 = 0, & \epsilon_4 = 0; \\
 & \epsilon_2 = 1, & \epsilon_3 + \epsilon_4 = 0, & \epsilon_5 = 0; \\
 & \vdots & \vdots & \vdots \\
 & \epsilon_2 = 1, & \epsilon_3 + \epsilon_4 + \cdots + \epsilon_{m-1} = 0, & \epsilon_m = 0; \\
\hline
 & \vdots & \vdots & \vdots \\
\hline
\text{Row Conditions } m-2: & \epsilon_{m-2} = 1, & \epsilon_{m-1} = 0; & \\
 & \epsilon_{m-2} = 1, & \epsilon_{m-1} = 0, & \epsilon_m = 0; \\
\hline
\text{Row Conditions } m-1: & \epsilon_{m-1} = 1, & \epsilon_m = 0; & \\
\hline
\text{Row Conditions } m: & & &
\end{array}
\qquad (S1)
$$

We call each line of (two or three) equations a *Condition*. In the system the label *Row Conditions i* refers to all conditions with $\epsilon_i = 1$ as the first equation. Note that Row Conditions m is empty; this is introduced for simplicity of notation later on. In general Row Conditions r consists of $m - r$ conditions. We define the length of $(S1)$ to be m.

The original problem can now be rephrased as follows: given the system $(S1)$ we define a partial solution to be an assignment to the variables that satisfies at least one condition. We now seek the following: find a partial solution S_1 and remove from the system all conditions that are satisfied. Now find another partial solution S_2 and remove once again all those conditions that are satisfied. Carry on till there are no conditions left, after n steps. We call such a sequence a solution sequence to the system. Clearly any such solution sequence $S_1, S_2, \ldots, S_n$ yields a sequence of m subsets of $A = \{1, 2, \ldots, n\}$ satisfying the condition (†). The partial solutions give us the successive columns of the matrix encoding the sets as described in the preceding section.

THEOREM 3.1 *For each $m \geq 1$ there is a solution sequence $S_1, S_2, \ldots, S_n$ to the system $(S1)$ with $n = 2\lceil \lg m \rceil$.*

PROOF. We use induction on m. If $m = 1$ there are no conditions and so we can take the empty set for our sequence.

Suppose now that $m > 1$. We will show that after two partial solutions the remaining system of conditions consists of two disjoint systems of size at most $\lceil m/2 \rceil$. Once this is established the result follows since by induction we can solve the two smaller systems simultaneously in at most $2\lg\lceil m/2 \rceil$ steps. This means that the overall system can be solved in at most $2 + 2\lg\lceil m/2 \rceil$ steps. The result follows since $\lg\lceil m/2 \rceil \leq \lceil \lg m \rceil - 1$.

For the first partial solution we take

$$S_1 = \{\, \epsilon_1 = 1, \ldots, \epsilon_{\lceil m/2 \rceil} = 1, \epsilon_{\lceil m/2 \rceil + 1} = 0, \ldots, \epsilon_m = 0 \,\}.$$

We claim that the last $\lfloor m/2 \rfloor$ conditions of Row Conditions $\lceil m/2 \rceil, \lceil m/2 \rceil - 2, \lceil m/2 \rceil - 4, \ldots$ are satisfied. To see this let $r = \lceil m/2 \rceil - 2s$. The last $\lfloor m/2 \rfloor$ conditions of Row Conditions r are

$$\begin{array}{rcr} \epsilon_r = 1, & \epsilon_{r+1} + \epsilon_{r+2} + \cdots + \epsilon_{\lceil m/2 \rceil} = 0, & \epsilon_{\lceil m/2 \rceil + 1} = 0; \\ \vdots & \vdots & \vdots \\ \epsilon_r = 1, & \epsilon_{r+1} + \epsilon_{r+2} + \cdots + \epsilon_{m-2} = 0, & \epsilon_{m-1} = 0; \\ \epsilon_r = 1, & \epsilon_{r+1} + \epsilon_{r+2} + \cdots + \epsilon_{m-1} = 0, & \epsilon_m = 0; \end{array}$$

Clearly S_1 satisfies the first and third equation of each of these conditions. The middle equations all reduce to the single equation

$$\epsilon_{r+1} + \epsilon_{r+2} + \cdots + \epsilon_{\lceil m/2 \rceil} = 0.$$

Each of the variables in this equation is set to 1 and as the number of them is $\lceil m/2 \rceil - r = 2s$ it follows that the equation holds.

We now remove the satisfied conditions from ($S1$) to obtain a new (sub)system ($S2$). Note that as a consequence Row Conditions $\lceil m/2 \rceil, \lceil m/2 \rceil - 2, \lceil m/2 \rceil - 4, \ldots$ now contain $0, 2, 4, \ldots$ conditions none of which uses the variables $e_{\lceil m/2 \rceil + 1}, \ldots, e_m$. Indeed in ($S2$) Row Conditions $\lceil m/2 \rceil - 2s$ consists of the first $2s$ conditions of Row Conditions $\lceil m/2 \rceil - 2s$ of ($S1$). The crucial point is that these form part of the system corresponding to ($S1$) but of length $\lceil m/2 \rceil$. For a partial solution to ($S2$) we take

$$S_2 = \{\, \epsilon_1 = 1, \ldots, \epsilon_{\lceil m/2 \rceil - 1} = 1, \epsilon_{\lceil m/2 \rceil} = 0, \ldots, \epsilon_m = 0 \,\}.$$

We claim that the last $\lceil m/2 \rceil$ conditions of Row Conditions $\lceil m/2 \rceil - 1, \lceil m/2 \rceil - 3, \lceil m/2 \rceil - 5, \ldots$ are satisfied. The proof is similar to the above, so we omit it.

We now remove the satisfied conditions from the system ($S2$) to obtain system ($S3$) and note that as a consequence Row Conditions $\lceil m/2 \rceil - 1, \lceil m/2 \rceil - 3, \lceil m/2 \rceil - 5, \ldots$ contain $0, 2, 4, \ldots$ conditions none of which uses the variables $\epsilon_{\lceil m/2 \rceil}, \ldots, \epsilon_m$. As above, the crucial point is that these form part of the system corresponding to ($S1$) but of size $\lceil m/2 \rceil$. It follows that in the system ($S3$) Row Conditions $\lceil m/2 \rceil, \lceil m/2 \rceil - 1, \lceil m/2 \rceil - 2$ form a subsystem ($S4$) of the system corresponding to ($S1$) but of length $\lceil m/2 \rceil$ in the variables $\epsilon_1, \ldots, \epsilon_{\lceil m/2 \rceil}$.

Furthermore the last $\lfloor m/2 \rfloor$ Row Conditions of ($S3$) are the same as those of the original system ($S2$) and are in the variables $\epsilon_{\lceil m/2 \rceil + 1}, \ldots, \epsilon_m$. Thus this is a system of size $\lfloor m/2 \rfloor$ which is *disjoint* from the subsystem ($S4$). The result follows since a solution sequence to the subsystem ($S4$) is no longer than a solution sequence to a complete system. □

THEOREM 3.2 *For each $n \geq 0$ there is a sequence of subsets $A_1, A_2, \ldots, A_m$ of $A = \{\, 1, 2, \ldots, n \,\}$ satisfying condition (†) with $m = 2^{\lfloor n/2 \rfloor}$.*

PROOF. Let $m = 2^{\lfloor n/2 \rfloor}$. The preceding Theorem shows that there is a sequence of the required type with subsets from a base set of cardinality at most $2\lceil \lg m \rceil = 2\lfloor n/2 \rfloor \leq n$. □