

t -Covering Arrays Generated by a Tiling Probability Model

Michael S. Donders

Department of Mathematics and Computer Science
McDaniel College

Anant P. Godbole

Department of Mathematics and Statistics
East Tennessee State University

November 2, 2010

Abstract

A $t - \alpha$ covering array is an $m \times n$ matrix, with entries from an alphabet of size α , such that for any choice of t rows, and any ordered string of t letters of the alphabet, there exists a column such that the “values” of the rows in that column match those of the string of letters. We use the Lovász Local Lemma in conjunction with a new tiling-based probability model to improve the upper bound on the smallest number of columns $N = N(m, t, \alpha)$ of a $t - \alpha$ covering array.

1 Introduction

Consider an $m \times n$ matrix with entries from the “alphabet” $A = \{1, 2, \dots, \alpha\}$. Let the (i, j) th entry be represented by $r_{i,j}$. We say that this matrix is a $t - \alpha$ -covering matrix or a $t - \alpha$ -covering array if given any t rows, p_1, p_2, \dots, p_t of the matrix, and any vector $\langle v_1, v_2, \dots, v_t \rangle$, with $v_i \in A$, there exists a column q such that

$$\langle v_1, v_2, \dots, v_t \rangle = \langle r_{p_1,q}, r_{p_2,q}, \dots, r_{p_t,q} \rangle.$$

Extensive surveys of covering arrays may be found in the papers of Sloane [5] and Colbourn [3]. Given t, m and the alphabet size $|A|$, we wish to find the minimum number of columns, n , such that there exists an $m \times n$ matrix that is t -covering. We will define $N = N(m, t, \alpha)$ as the smallest positive integer n such that there exists a covering array of dimensions $m \times n$. At the Coimbra Zero-One Matrix Conference, the second author talked about the need to introduce new probability models to improve upper bounds on $N(m, t, \alpha)$ and the corresponding numbers for *partial covering arrays* [2]. In this paper we propose a specific way of doing so, once again using the Lovász local lemma as an auxiliary tool.

Lemma 1 *The Lovász Local Lemma ([1]): Let C_1, C_2, \dots, C_K be the events in arbitrary probability space. Suppose that each event C_i is mutually independent of a set of all the other events C_k but at most d , and that $P(C_i) \leq p$ for all $1 \leq i \leq K$. If $ep(d+1) \leq 1$ then $P(\bigcap_{k=1}^K C'_k) > 0$.*

Let R be the index set of all sets of t rows; $|R| = \binom{m}{t}$. For $r \in R$, let C_r be the event that the r^{th} row set does not contain some vector $\langle v_1, v_2, \dots, v_t \rangle$ in any of its columns. We wish to prove that $P(\bigcap_{r \in R} C'_r) > 0$ if $n \geq N_0$, proving that $N(m, t, \alpha) \leq N_0$. Now in [4] a general upper bound was provided on the size of covering arrays; this was

$$N(m, t, \alpha) \leq N_0 := (t-1) \frac{\log_2(m)}{\log_2 \frac{\alpha^t}{\alpha^t - 1}} \{1 + o(1)\}. \quad (1)$$

The proof used an elementary probability model that consisted of placing one letter of the alphabet independently in each of the mn positions with probability $\frac{1}{\alpha}$, i.e. by letting $P(r_{i,j} = x) = \frac{1}{\alpha} \forall x \in A$. In the same paper, a special probability model was used, but only for the case $\alpha = 2, t = 3$. Here the authors of [4], following the approach used in the doctoral thesis of Roux (see, e.g. [5]), used a probability model that independently places an equal number of zeros and ones in the rows of the matrix (the so-called “fixed weight rows” model.) Unfortunately this method becomes quite intractable in general, and it is our intent in this paper to explore a probability model that is, in some sense, intermediate between the general technique in [4] and the special method used there for $\alpha = 2, t = 3$: Specifically, we seek to improve the general bound (1) using the method of placing *consecutive and equally weighted tiles* along the rows. We use tiles of dimension $1 \times k\alpha$, such that there are exactly k x 's in each tile for each $x \in A$. By way of

comparison, the general method used 1×1 tiles that led to a loss of control over the numbers of letters of each type in any row, while Roux's method used a *single long tile* in each row, i.e., corresponded to $k = n/2$ (n even).

We consider two cases, when (i) $k = 1$ which yields an elementary equation relating $N(m, t, \alpha)$ and the variables m , t and α , and when (ii) $k > 1$, which yields better bounds as k increases, but which generates increasingly more complicated solutions.

(i) We start with the case $k = 1$, and fill in our matrix using tiles that contain one randomly placed copy of each letter of the alphabet, assuming that $\alpha|n$. Note that there are a total of α^t possible vectors, and by the symmetry of our construction, all are equally likely to occur in the selected rows. Thus $P(C_r) \leq \lambda \alpha^t$ where λ is the probability that a specific vector $z^* = \langle z_1, z_2, \dots, z_t \rangle$ is missing in the set r of selected rows. Select an arbitrary set of t rows in the matrix. Consider the columns in any vertically aligned set of tiles. For each z_i , there is exactly one value in any tile equal to z_i , and α places it can be; moreover z^* cannot occur in more than one column of the vertically stacked tiles in the selected rows. Therefore, the probability that z^* is somewhere in these tiles is $\alpha \cdot \left(\frac{1}{\alpha}\right)^t = \left(\frac{1}{\alpha}\right)^{t-1}$. Since there are $\frac{n}{\alpha}$ tiles in any row of the $m \times n$ matrix, and the composition of these is determined independently, we have

$$\lambda = \left(1 - \left(\frac{1}{\alpha}\right)^{t-1}\right)^{\frac{n}{\alpha}} = \left(\left(1 - \left(\frac{1}{\alpha}\right)^{t-1}\right)^{\frac{1}{\alpha}}\right)^n,$$

and thus,

$$P(C_r) \leq \alpha^t \left(\left(1 - \left(\frac{1}{\alpha}\right)^{t-1}\right)^{\frac{1}{\alpha}}\right)^n. \quad (2)$$

We can improve this bound slightly by using a technique found in [2], where the vectors $z_i = \langle i, i, \dots, i \rangle; 1 \leq i \leq \alpha$ can be achieved for all sets r by including columns consisting of all i 's. There are α of these vectors; thus this reduces the number of z^* 's from α^t to $\alpha^t - \alpha$. We can ignore these vectors in our calculation of $P(C_r)$ so long as we remember to add α columns to our value $N(m, t, \alpha)$. So (2) may be improved as follows:

$$P(C_r) \leq (\alpha^t - \alpha) \left(\left(1 - \left(\frac{1}{\alpha}\right)^{t-1}\right)^{\frac{1}{\alpha}}\right)^n. \quad (3)$$

Our next step is to calculate d . For any set r of rows, there will be a dependency only on sets $r_0 \in R$ such that $r \cap r_0 \neq \emptyset$. We will bound the number of such r_0 's by choosing one row from r , and then choosing an arbitrary $t-1$ rows from the $m-1$ other rows in the matrix. Thus $d \leq t \binom{m-1}{t-1}$, so $d+1 \leq \frac{tm^{t-1}}{(t-1)!}$. Substituting this into the Lovász local lemma we get

$$ep(d+1) = \frac{etm^{t-1}}{(t-1)!} (\alpha^t - \alpha) \left(\left(1 - \left(\frac{1}{\alpha} \right)^{t-1} \right)^{\frac{1}{\alpha}} \right)^n \leq 1$$

if

$$n \geq \frac{(t-1) \log_2(m)}{\log_2 \left(\left(\frac{\alpha^{t-1}}{\alpha^{t-1}-1} \right)^{\frac{1}{\alpha}} \right)} \left\{ 1 + \frac{\log_2(\alpha^t - \alpha)}{(t-1) \log_2(m)} + \frac{\log_2(et)}{(t-1) \log_2(m)} - \frac{\log_2((t-1)!)}{(t-1) \log_2(m)} \right\},$$

i.e., if

$$n \geq \frac{\alpha(t-1) \log_2(m)}{\log_2 \left(\frac{\alpha^{t-1}}{\alpha^{t-1}-1} \right)} \{1 + o(1)\} \quad m \rightarrow \infty.$$

It follows that

$$N(m, t, \alpha) \leq \frac{\alpha(t-1) \log_2(m)}{\log_2 \left(\frac{\alpha^{t-1}}{\alpha^{t-1}-1} \right)} \{1 + o(1)\}. \quad (4)$$

since adding back, into (4), the α columns we removed earlier only changes the $o(1)$ term. Notice that the above process gives us both a precise and an asymptotic bound for $n(m, t, \alpha)$. Note too that (4) gives an improvement over the previous best bound (1) due to the fact that

$$\left(1 - \frac{1}{\alpha^{t-1}} \right)^{\frac{1}{\alpha}} \leq 1 - \frac{1}{\alpha^t}.$$

(ii) We now consider the case $k > 1$; recall that the size of our tiles is $1 \times k\alpha$. First note that the size of the tile does not change d , and thus $d+1 \leq \frac{tm^{t-1}}{(t-1)!}$ as before. We next reconsider $P(C_r)$, and compute it using inclusion exclusion. Let

$$\begin{aligned} \gamma_k &= \frac{\sum_{i=1}^k (-1)^{(i+1)} \binom{\alpha k}{i} \binom{\alpha k - i}{k-i, k, \dots, k}^t}{\binom{\alpha k}{k, k, \dots, k}^t} \\ &= \frac{\sum_{i=1}^k (-1)^{(i+1)} \binom{\alpha k}{i} \binom{\alpha k - i}{k-i}^t}{\binom{\alpha k}{k}^t} \end{aligned} \quad (5)$$

be the probability that a given vector z^* is in a given vertical array of t tiles. This yields $\lambda = \lambda_k = (1 - \gamma_k)^{\frac{n}{k\alpha}}$ and hence

$$P(C_r) \leq (\alpha^t - \alpha)(1 - \gamma_k)^{\frac{n}{k\alpha}},$$

so that the Lovász local lemma yields $P(\cap C'_r) > 0$ if,

$$(\alpha^t - \alpha)(1 - \gamma_k)^{\frac{n}{k\alpha}} \frac{etm^{t-1}}{(t-1)!} \leq 1$$

i.e., if ,

$$n \geq \frac{(t-1) \log_2(m)}{\log_2 \left(\left(\frac{1}{1-\gamma_k} \right)^{\frac{1}{k\alpha}} \right)} \left\{ 1 + \frac{\log_2(\alpha^t - \alpha)}{(t-1) \log_2(m)} + \frac{\log_2(et)}{(t-1) \log_2(m)} - \frac{\log_2((t-1)!)}{(t-1) \log_2(m)} \right\},$$

or

$$n \geq \frac{k\alpha(t-1) \log_2(m)}{\log_2 \left(\frac{1}{1-\gamma_k} \right)} \{1 + o(1)\} \quad m \rightarrow \infty.$$

It follows that

$$N(m, t, \alpha) \leq \frac{k\alpha(t-1) \log_2(m)}{\log_2 \left(\frac{1}{1-\gamma_k} \right)} \{1 + o(1)\}. \quad (6)$$

Comments It is clear that as we increase k from 1 to $\frac{n}{\alpha}$, the bound on $N(m, t, \alpha)$ becomes better and better, while the equation to solve for it becomes more and more convoluted. Take for example, the case when $t = 3$, $\alpha = 2$. The previous best known bound (1) for a general $N(m, t, \alpha)$ yields the solution $N(m, 3, 2) \leq 10.38 \log_2(m) \{1 + o(1)\}$, while the best known solution for this specific case (Roux [4]) yields $N(m, 3, 2) \leq 7.56 \log_2(m) \{1 + o(1)\}$, a result obtained by equally weighing all the rows to have the same number of 1's and 0's. The solution obtained via tiling yields $N(m, 3, 2) \leq 9.64 \log_2(m) \{1 + o(1)\}$ when $k = 1$. With this we can see that even the simplest case of the tiling solution, $k = 1$, offers a fairly significant improvement in the bounds, while more complex solution will provide the better bounds for the size of a covering array. A few values of $N(m, t, \alpha)$ as given by (4) and (6) may be found in the following table; $k = 0$ refers to the bound in (1):

α	t	k	$N(m, t, \alpha)/\log_2 m$
2	3	0	10.38
2	3	1	9.64
2	3	2	8.68
2	3	3	8.31
2	4	0	32.22
2	4	1	31.15
2	4	2	29.55
2	4	3	28.85
3	3	0	36.73
3	3	1	35.31
3	3	3	33.28
3	3	5	32.79
3	4	0	167.39
3	4	1	165.3
3	4	3	161.57
3	4	5	160.47
4	3	0	88.03
4	3	2	83.97
4	3	4	82.72
4	3	6	82.27
4	4	0	531.3
4	4	2	524.75
4	4	4	521.98
4	4	6	520.90
5	4	0	1298.61
5	4	2	1290.12
5	4	4	1286.46
5	4	6	1285.01
5	5	0	8662.95
5	5	2	8651.13
5	5	4	8644.67
5	5	6	8641.86

2 Open Problems

Perhaps the overarching open problem is that of using alternative probability models in order to tease out better and better bounds on the size of minimal covering arrays. Markov models and others involving global dependence are one option. A method more relevant to the central problem addressed at the Coimbra conference, would, however, be to work with zero-one or alphabet based matrices with fixed row and column totals (in this paper we fix just the row totals!). Last but not least, can we let k go to infinity (at a relatively slow rate) and analyze the sum in (5)? Can we conduct the analysis with $k = n/\alpha$?

3 Acknowledgment

The research of both authors was supported by NSF Grant 1004624.

References

- [1] N. Alon and J. Spencer, *The Probabilistic Method*, 2nd Edition, John Wiley, New York, 2000.
- [2] P. Carey and A. Godbole (2010), “Partial covering arrays and a generalized Erdős-Ko-Rado property,” *J. Combinatorial Designs* **18**, 155-166.
- [3] C. Colbourn (2004), “Combinatorial aspects of covering arrays,” *Le Matematiche (Catania)* **58**, 121–167.
- [4] A. Godbole, D. Skipper, and R. Sunley (1996), “ t -covering arrays: upper bounds and Poisson approximations,” *Combinatorics, Probability and Computing* **5**, 105–118.
- [5] N. J. A. Sloane (1993), “Covering arrays and intersecting codes,” *Journal of Combinatorial Designs* **1**, 51–63.