Combinatorial Structures
Part 2: Orthogonal Arrays and Codes
CS 858 Notes

Douglas R. Stinson

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada

May 14, 2019

1 Orthogonal Arrays

Suppose t, k, v, λ are positive integers such that $t \leq k$. An *orthogonal array*, denoted $\mathrm{OA}_\lambda(t, k, v)$, is a λv^t by k array on v symbols such that, in any t columns, every t-tuple of symbols occurs exactly λ times. The integer t is called the *strength* of the OA.

Example: We present an $\mathrm{OA}_1(2, 4, 3)$:

$$\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 \\
0 & 2 & 2 & 2 \\
1 & 0 & 1 & 2 \\
1 & 1 & 2 & 0 \\
1 & 2 & 0 & 1 \\
2 & 0 & 2 & 1 \\
2 & 1 & 0 & 2 \\
2 & 2 & 1 & 0
\end{array}$$

A *Latin square of order* n is an n by n array of n symbols such that every symbol occurs exactly once in each row and each column.

Two Latin squares of order n are *orthogonal* if their superposition contains every possible ordered pair of symbols. We will use the notation OLS(n) to denote orthogonal Latin squares of order n.

Example: We present OLS(3) and their superposition:

0	1	2
1	2	0
2	0	1

0	1	2
2	0	1
1	2	0

$0,0$	$1,1$	$2,2$
$1,2$	$2,0$	$0,1$
$2,1$	$0,2$	$1,0$

Euler asked the question "when do there exist OLS(n)?" This was answered by Bose, Shrikhane and Parker, who proved that there exist OLS(n) if and only if $n \geq 1$, $n \neq 2, 6$. The "hard" cases were when $n \equiv 2 \pmod 4$.

A set of k orthogonal Latin squares of order n are *mutually orthogonal* if all $\binom{k}{2}$ pairs chosen from the k squares are orthogonal. We will use the notation k MOLS(n) to denote k mutually orthogonal Latin squares of order n.

Theorem: If $n \geq 2$ and there exist k MOLS(n), then $k \leq n - 1$.

Theorem: There exist $n-1$ MOLS(n) if and only if there exists a projective plane of order n.

Let $N(n)$ denote the maximum integer k such that there exist k MOLS(n). Then $N(n) = n - 1$ for $n = 2, 3, 4, 5, 7, 8, 9, 11$. Further, $N(6) = 1$, $2 \leq N(10) \leq 8$ and $5 \leq N(12) \leq 11$.

There exist k MOLS(n) if and only if there exists an $\mathrm{OA}_1(2, k+2, n)$.

Example: Suppose we start with the $\mathrm{OA}_1(2, 4, 3)$ given above. We construct two Latin squares L_1 and L_2 as follows: for every row x_1, x_2, x_3, x_4 of the OA, define $L_1(x_1, x_2) = x_3$ and $L_2(x_1, x_2) = x_4$. Then L_1 and L_2 are OLS(3). In fact, they are the OLS(3) presented in the example above.

Plackett-Burman Bound: If there exists an $\mathrm{OA}_\lambda(2, k, v)$, then

$$\lambda \geq \frac{k(v-1)+1}{v^2}.$$

When $\lambda = 1$, the bound becomes $k \leq v + 1$.

Rao Bound: Suppose $t \geq 2$ is an even integer. If there exists an $\mathrm{OA}_\lambda(2, k, v)$, then

$$\lambda v^t \geq 1 + \sum_{i=1}^{t/2} \binom{k}{i} (v-1)^i.$$

When $t = 2$, the Rao bound is the same as the Plackett-Burman bound.

2 Codes

An (n, M, d)-v-code consists of a set of M n-tuples defined over a set of v symbols, such that the hamming distance between any two of the n-tuples is at least d. The integer n is the *length* of the code, M is the size of the code, and d is the *distance* of the code. The n-tuples are called *codewords.*

A code having distance d can correct $(d-1)/2$ errors.

Suppose q is a prime power. An $[n, m, d]$-q-linear code is an (n, q^m, d)-q-code, defined over the symbols $\mathbb{F}_q$, that is an m-dimensional subspace of the vector space $(\mathbb{F}_q)^n$. The integer m is the *dimension* of the code.

A *generator matrix* for an $[n, m, d]$-q-linear code is an m by n matrix, say G, whose rows form a basis of the code. Thus we can obtain the entire code by taking all possible linear combinations (over $\mathbb{F}_q$) of the rows of G.

Suppose $\mathcal{C}$ is an $[n, m, d]$-q-linear code. The *dual code*, denoted $\mathcal{C}^\perp$, is the orthogonal complement of $\mathcal{C}$. This means that

$$\mathcal{C}^\perp = \{\mathbf{y} \in (\mathbb{F}_\mathbf{q})^\mathbf{n} : \mathbf{y} \cdot \mathbf{x} = \mathbf{0} \text{ for every } \mathbf{x} \in \mathcal{C}\}.$$

If $\mathcal{C}$ is an $[n, m, d]$-q-linear code, then $\mathcal{C}^\perp$ is an $[n, n-m, d']$-q-linear code for some d'.

Suppose $\mathcal{C}$ is an $[n, m, d]$-q-linear code. A *parity-check matrix* for $\mathcal{C}$, say H, is a generator matrix for the dual code $\mathcal{C}^\perp$. Thus H is an $n-m$ by n matrix with entries from $\mathbb{F}_q$, such that $GH^T = 0$.

G is a generator matrix for a code having distance $\geq d$ if and only if any $d-1$ columns of the parity check matrix are linearly independent.

Suppose $\ell \geq 2$ is an integer. We can construct a *Hamming code* as follows. Consider the $q^\ell - 1$ nonzero column ℓ-tuples over $\mathbb{F}_q$. Two such ℓ-tuples are said to be *equivalent* if they are scalar multiples of each other. It is easy to see that there are $(q^\ell - 1)/(q-1)$ equivalence classes. Take one column from each equivalence class and construct a matrix H having the given columns. This matrix is the parity-check matrix of the Hamming code. Since any two columns of H are linearly independent, the Hamming code has distance 3. Thus the Hamming code is an $[n, n-\ell, 3]$-q-linear code, where $n = (q^\ell - 1)/(q-1)$.

Example: Suppose $q = 2$ and $\ell = 3$. The Hamming code is a $[7, 4, 3]$-2-linear code. A parity-check matrix for this code is

$$H = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Example: Suppose $q = 3$ and $\ell = 2$. The four equivalence classes of column vectors of length 2 are $\binom{1}{0}$ and $\binom{2}{0}$; $\binom{0}{1}$ and $\binom{0}{2}$; $\binom{1}{1}$ and $\binom{2}{2}$; and $\binom{1}{2}$ and $\binom{2}{1}$. Suppose we choose the first column from each equivalence class. Then we obtain the parity-check matrix

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}.$$

The resulting Hamming code is a $[4, 2, 3]$-3-linear code.

A generator matrix for the previous code is

$$G = \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix}.$$

It is easy to verify that

$$GH^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Given an $[n, m, d]$-q-linear code, the dual code is an orthogonal array of strength $d - 1$.

Example: The subspace generated by the matrix H above is an orthogonal array of strength $t = 2$. More specifically, it is an $\mathrm{OA}_1(2, 4, 3)$.

Singleton Bound: If an (n, M, d)-v-code exists, then $M \leq v^{n-d+1}$. This is proven as follows: Consider any $n - d + 1$ coordinates. If there exist two codewords that are identical in these coordinates, then their Hamming distance is at most $d - 1$, which is a contradiction. Therefore the size of the code is at most v^{n-d+1}.

An (n, M, d)-v-code with $M = v^{n-d+1}$ is called a *maximum distance separable* code (or, MDS code).

An MDS code is an orthogonal array of strength $t = n - d + 1$ with $\lambda = 1$. More specifically, an (n, v^{n-d+1}, d)-v-code is an $\mathrm{OA}_1(n - d + 1, n, v)$.

Reed-Solomon codes are MDS codes. Let q be a prime power and let $t \geq 2$. For every polynomial $f(x) \in \mathbb{F}_q[x]$ having degree $\leq t - 1$, evaluate $f(x)$ at all the elements of $\mathbb{F}_q$ and let the resulting q-tuple be a codeword. We claim that this is a $(q, q^t, q - t + 1)$-q-code. To show that the distance $d = q - t + 1$, consider two polynomials $f(x)$ and $g(x)$. The distance between the corresponding codewords is $|\{x \in \mathbb{F}_q : f(x) \neq g(x)\}|$. If this quantity is $\leq q - t$, then $|\{x \in \mathbb{F}_q : f(x) = g(x)\}| \geq t$. But then the polynomial $(f - g)(x)$ of degree $\leq t - 1$ has at least t roots, which impossible unless $f = g$. The code is an MDS code because $t = q - d + 1$. It is also an $\mathrm{OA}_1(t, q, q)$.

Example: Suppose we take $q = 3$, $t = 2$. The associated Reed-Solomon code is a $(3, 9, 3)$-3-code which is also an $\mathrm{OA}_1(2, 3, 3)$. The nine polynomials of degree ≤ 1 are $0, 1, 2, x, x+1, x+2, 2x, 2x+1, 2x+2$. We get the following code by evaluating these polynomials at $0, 1, 2$:

$f(x)$	$x = 0$	1	2
0	0	0	0
1	1	1	1
2	1	1	1
x	0	1	2
$x+1$	1	2	0
$x+2$	2	0	1
$2x$	0	2	1
$2x+1$	1	0	2
$2x+2$	2	1	0

Gilbert-Varshamov Bound: This is an existence result for linear codes, as opposed to a necessary condition. It provides some numerical conditions that ensure that a parity-check matrix can be constructed for a code with desired distance. Recall that an ℓ by q matrix H of elements from $\mathbb{F}_q$ is the parity-check matrix of a code with distance $\geq d$ if any $d-1$ columns of H are linearly independent. The idea is to construct H one column at a time, ensuring that no column is a linear combination of $d-2$ or fewer previous columns. Let's compute the number of "bad" choices for the last (i.e., the nth) column. For $1 \leq i \leq d-2$, the number of linear combinations of previous columns in which precisely i coefficients are nonzero is $\binom{n-1}{i}(q-1)^i$. The last column also cannot be the all-zero column. Since there are q^ℓ choices for the last column, there is a "good" choice for this column if

$$q^\ell > 1 + \sum_{i=1}^{d-2} \binom{n-1}{i} (q-1)^i. \tag{1}$$

Now, this argument applies to the last column, but it is not hard to see that there are fewer restrictions for choosing any of the previous columns. Thus the inequality (1) is sufficient to establish the existence of a $[n, n-\ell, d]$-q-linear code. Unfortunately, there is no efficient (i.e., polynomial-time) to construct the matrix H.