

Error correcting codes from quasi-Hadamard matrices

V. Álvarez, J.A. Armario, M.D. Frau, E. Martín, and A. Osuna*

Dpto. Matemática Aplicada I, Universidad de Sevilla, Avda. Reina Mercedes s/n
41012 Sevilla, Spain,

{valvarez,armario,mdfrau,emartin,aosuna}@us.es

Abstract. Levenshtein described in [5] a method for constructing error correcting codes which meet the Plotkin bounds, provided suitable Hadamard matrices exist. Uncertainty about the existence of Hadamard matrices on all orders multiple of 4 is a source of difficulties for the practical application of this method. Here we extend the method to the case of quasi-Hadamard matrices. Since efficient algorithms for constructing quasi-Hadamard matrices are potentially available from the literature (e.g. [7]), good error correcting codes may be constructed in practise. We illustrate the method with some examples.

Keywords: Error correcting code, Hadamard matrix, Hadamard code.

1 Introduction

One of the main goals in Coding Theory is the design of optimal error correcting codes. For given length n and minimum distance d , the term optimal means a code which consists of a set of code words as large as possible. For (not necessarily linear) binary codes (n, M, d) , Plotkin found out in [10] the following bounds for the number M of codewords:

$$M \leq 2 \lfloor \frac{d}{2d-n} \rfloor \quad \text{if } d \text{ is even and } d \leq n < 2d, \quad (1)$$

$$M \leq 2n \quad \text{if } d \text{ is even and } n = 2d, \quad (2)$$

$$M \leq 2 \lfloor \frac{d+1}{2d+1-n} \rfloor \quad \text{if } d \text{ is odd and } d \leq n < 2d+1, \quad (3)$$

$$M \leq 2n+2 \quad \text{if } d \text{ is odd and } n = 2d+1. \quad (4)$$

Levenshtein proved in [5] that the Plotkin bounds are tight, in the sense that there exist binary codes which meet these bounds, provided that enough Hadamard matrices exist. Unfortunately, the Hadamard Conjecture about the

* All authors are partially supported by the PAICYT research project FQM-296 from Junta de Andalucía (Spain).

existence of Hadamard matrices in all orders multiple of 4 remains still open. Moreover, there are infinite orders for which no Hadamard matrices have been found. This means that, though theoretically correct, Levenshtein's method could not be useful in practise.

In the sequel a matrix for which the inner product of rows two by two is mostly zero is called a quasi-Hadamard matrix. We will use Levenshtein's method in this paper to show that "good" error-correcting codes may be analogously constructed from quasi-Hadamard matrices. Here the term "good" refers to a code formed from a significantly large number of code words, for given length and minimum distance. We must emphasize that quasi-Hadamard matrices may be straightforwardly obtained in all orders multiple of 4, so that the associated error-correcting codes may be constructed in practise.

We organize the paper as follows.

In Section 2 we introduce the notion of quasi-Hadamard matrices, and some processes to construct them, which are available in the literature. Section 3 is devoted to explain how to construct good error-correcting codes from suitable quasi-Hadamard matrices. Some examples are discussed in Section 4.

2 Quasi-Hadamard matrices

A Hadamard matrix H of order n is an $n \times n$ matrix of $+1$'s and -1 's entries such that $HH^T = nI$. That is, the inner product of any two distinct rows of H is zero.

We now generalize this notion.

We define a quasi-Hadamard matrix of order n as an $n \times n$ matrix M of $+1$'s and -1 's entries such that the inner product of rows two by two is mostly zero.

Sometimes it is necessary to precise the largest number q of rows in M which are orthogonal one to each other. The larger q is, the closer M is from being a Hadamard matrix. In these circumstances, M is termed a quasi-Hadamard matrix of depth q .

In some sense, a quasi-Hadamard matrix could be thought as a Hadamard matrix in which some rows have been substituted, so that the Hadamard character is generally lost in turn.

Constructing Hadamard matrices is hard. How about constructing quasi-Hadamard matrices?

We now attend to another characterization of Hadamard matrices, in terms of cliques of graphs (that is, a collection of n vertices and $\frac{n(n-1)}{2}$ edges of a graph G which form a complete subgraph K_n of G).

Consider the graph G_{4t} whose vertices are all the tuples of length $4t$ formed from $2t$ ones and $2t$ minus ones, with the restriction that precisely t ones have to appear within the first $2t$ positions (by analogy, precisely t ones appear within the last $2t$ positions). There is an edge between two vertices if and only if the inner product of the correspondent tuples is zero. A Hadamard matrix of order $4t$ exists if and only if G_{4t} contains a clique of size $4t - 2$. Furthermore, the vertices

of such a clique and the normalized rows $\overbrace{(1, \dots, 1)}^{4t}$ and $\overbrace{(1, \dots, 1, -1, \dots, -1)}^{2t}$ form a Hadamard matrix. This is a particular type of Hadamard Graph, as defined in [8, 9].

Unfortunately, the problem of finding out the maximum clique in a graph has been proven to be NP-hard [4]. Moreover, even its approximations within a constant factor are NP-hard [2, 3]. So one should expect that finding out Hadamard matrices from G , or even quasi-Hadamard matrices for large depths close to $4t$, are to be hard problems. In fact, they are.

Hopefully, heuristic methods for the maximum clique problem can be found in the literature, which output pretty large cliques [7]. These methods can be used in turn to construct quasi-Hadamard matrices of large depth as well.

3 Quasi-Hadamard codes

We firstly recall Levenshtein's method [5] for constructing optimal error correcting codes from suitable Hadamard matrices.

Starting from a normalized (i.e. the first row and column formed all of 1's) Hadamard matrix H of order $4t$, some codes (which are termed Hadamard codes) may be constructed (see [6], for instance). More concretely, consider the matrix A_{4t} related to H_{4t} , which consists in replacing the +1's by 0's and the -1's by 1's. Since the rows of H_{4t} are orthogonal, any two rows of A_{4t} agree in $2t$ places and differ in $2t$ places, and so have Hamming distance $2t$ apart. In these circumstances, one may construct:

1. An $(4t - 1, 4t, 2t)$ code, \mathcal{A}_{4t} , consisting of the rows of A_{4t} with the first column deleted. This is optimal for the Plotkin bound (1).
2. An $(4t - 1, 8t, 2t - 1)$ code, \mathcal{B}_{4t} , consisting of \mathcal{A}_{4t} together with the complements of all its codewords. This is optimal for the Plotkin bound (4).
3. An $(4t, 8t, 2t)$ code, \mathcal{C}_{4t} , consisting of the rows of A_{4t} and their complements. This is optimal for the Plotkin bound (2).
4. An $(4t - 2, 2t, 2t)$ code, \mathcal{D}_{4t} , formed from the codewords in \mathcal{A}_{4t} which begin with 0, with the initial zero deleted. This is optimal for the Plotkin bound (1).

Furthermore, as explained in [6], for any $d \leq n < 2d$, an optimal code attending to the Plotkin bound (1) may be obtained from a suitable combination of codes of the above type.

More concretely, given d even so that $2d > n \geq d$, define $k = \lfloor \frac{d}{2d - n} \rfloor$ and

$$a = d(2k + 1) - n(k + 1), \quad b = kn - d(2k - 1).$$

Then a and b are nonnegative integers satisfying that $n = (2k - 1)a + (2k + 1)b$ and $d = ka + (k + 1)b$. Moreover, if n is even then so are a and b . Analogously, if n is odd and k even, then b is even. Finally, if both of n and k are odd, then a is even.

Depending on the parity of n and k , define the code \mathcal{C} to be:

- If n is even, $\mathcal{C} = \frac{a}{2}\mathcal{D}_{4k} \oplus \frac{b}{2}\mathcal{D}_{4k+4}$.
- If n is odd and k even, $\mathcal{C} = a\mathcal{A}_{2k} \oplus \frac{b}{2}\mathcal{D}_{4k+4}$.
- If n and k are odd, $\mathcal{C} = \frac{a}{2}\mathcal{D}_{4k} \oplus b\mathcal{A}_{2k+2}$.

Here \oplus denotes the following “summation” of codes. Suppose that \mathcal{C}_1 and \mathcal{C}_2 are (n_1, M_1, d_1) and (n_2, M_2, d_2) codes, respectively. Assume, for instance, that $M_2 \geq M_1$. For nonnegative integers a, b , the code $a\mathcal{C}_1 \oplus b\mathcal{C}_2$ consists in pasting a copies of \mathcal{C}_1 , side by side, followed by b copies of the code obtained from \mathcal{C}_2 by omitting the last $M_2 - M_1$ codewords. By construction, $a\mathcal{C}_1 + b\mathcal{C}_2$ is shown to be an $(an_1 + bn_2, M_1, d)$ code, for $d \geq ad_1 + bd_2$.

This way, the code \mathcal{C} defined above meets the Plotkin bound (1), since it has length n , minimum distance d , and contains $2k = 2\lfloor \frac{d}{2d-n} \rfloor$ codewords.

We now extend Levenshtein’s method for constructing optimal error correcting codes from Hadamard matrices to the case of quasi-Hadamard matrices. The codes so obtained are termed quasi-Hadamard codes.

Consider a normalized quasi-Hadamard matrix M_{4t} of order $4t$ and depth q . We define the matrix A'_{4t} related to M_{4t} in the following way: select a q -set of rows of M_{4t} which are orthogonal one to each other (notice that there is no larger set with this property, since q is the depth of M_{4t}), and replace the $+1$ ’s by 0 ’s and the -1 ’s by 1 ’s.

Theorem 1. *In the circumstances above, the following quasi-Hadamard codes may be constructed:*

1. An $(4t-1, q, 2t)$ code, \mathcal{A}'_{4t} , consisting of the rows of A'_{4t} with the first column deleted.
2. An $(4t-1, 2q, 2t-1)$ code, \mathcal{B}'_{4t} , consisting of A'_{4t} together with the complements of all its codewords.
3. An $(4t, 2q, 2t)$ code, \mathcal{C}'_{4t} , consisting of the rows of A'_{4t} and their complements.
4. An $(4t-2, h, 2t)$ code, \mathcal{D}'_{4t} , formed from the h codewords in A'_{4t} which begin with 0 , with the initial zero deleted (we only know that $h \leq q$).

Proof.

It is a straightforward extension of the case of usual Hadamard codes coming from Hadamard matrices, since:

- A'_{4t} consists of q rows.
- Any two rows of A'_{4t} agree in $2t$ places and differ in $2t$ places (since they are pairwise orthogonal), and so have Hamming distance $2t$ apart.

The result follows. □

Remark 1. Obviously, the closer q is from $4t$, the better codes \mathcal{A}'_{4t} , \mathcal{B}'_{4t} , \mathcal{C}'_{4t} and \mathcal{D}'_{4t} are. In the sense that the number of codewords is very close to the optimal value indicated in the Plotkin bound.

Theorem 2. For d even so that $2d > n \geq d$, define $k = \lfloor \frac{d}{2d-n} \rfloor$ and

$$a = d(2k+1) - n(k+1), \quad b = kn - d(2k-1).$$

as before. A good error correcting code \mathcal{C}' of length n and minimum distance d may be obtained, from suitable quasi-Hadamard matrices. More concretely, depending on the parity of n and k , define the code \mathcal{C}' to be:

- If n is even, $\mathcal{C}' = \frac{a}{2}\mathcal{D}'_{4k} \oplus \frac{b}{2}\mathcal{D}'_{4k+4}$.
- If n is odd and k even, $\mathcal{C}' = a\mathcal{A}'_{2k} \oplus \frac{b}{2}\mathcal{D}'_{4k+4}$.
- If n and k are odd, $\mathcal{C}' = \frac{a}{2}\mathcal{D}'_{4k} \oplus b\mathcal{A}'_{2k+2}$.

Proof.

From Levenshtein's method [5] described before, it is readily checked that \mathcal{C}' consists of codewords of length n . Furthermore:

- If n is even, select a normalized quasi-Hadamard matrix ${}^1M_{4k}$ of order $4k$ and depth q_1 , and a normalized quasi-Hadamard matrix ${}^2M_{4k+4}$ of order $4k+4$ and depth q_2 . Denote ${}^i\mathcal{A}'$ a q_i -set of pairwise orthogonal rows in iM with their first entry dropped, and where the $+1$'s and the -1 's have been replaced by 0 's and 1 's, respectively. Denote ${}^i\mathcal{D}'$ the h_i -set of rows in ${}^i\mathcal{A}'$ which begin with 0 , for $0 \leq h_i \leq q_i$. In these circumstances, $\mathcal{C}' = \frac{a}{2}({}^1\mathcal{D}'_{4k}) \oplus \frac{b}{2}({}^2\mathcal{D}'_{4k+4})$ consists in a $(n, \min\{h_1, h_2\}, d)$ -code.
- If n is odd and k even, select a normalized quasi-Hadamard matrix ${}^1M_{2k}$ of order $2k$ and depth q_1 , and a normalized quasi-Hadamard matrix ${}^2M_{4k+4}$ of order $4k+4$ and depth q_2 . Denote ${}^i\mathcal{A}'$ a q_i -set of pairwise orthogonal rows in iM with their first entry dropped, and where the $+1$'s and the -1 's have been replaced by 0 's and 1 's, respectively. Denote ${}^2\mathcal{D}'$ the h_2 -set of rows in ${}^2\mathcal{A}'$ which begin with 0 , for $0 \leq h_2 \leq q_2$. In these circumstances, $\mathcal{C}' = a({}^1\mathcal{A}'_{2k}) \oplus \frac{b}{2}({}^2\mathcal{D}'_{4k+4})$ consists in a $(n, \min\{q_1, h_2\}, d)$ -code.
- If n and k are odd, select a normalized quasi-Hadamard matrix ${}^1M_{4k}$ of order $4k$ and depth q_1 , and a normalized quasi-Hadamard matrix ${}^2M_{2k+2}$ of order $2k+2$ and depth q_2 . Denote ${}^i\mathcal{A}'$ a q_i -set of pairwise orthogonal rows in iM with their first entry dropped, and where the $+1$'s and the -1 's have been replaced by 0 's and 1 's, respectively. Denote ${}^1\mathcal{D}'$ the h_1 -set of rows in ${}^1\mathcal{A}'$ which begin with 0 , for $0 \leq h_1 \leq q_1$. In these circumstances, $\mathcal{C}' = \frac{a}{2}({}^1\mathcal{D}'_{4k}) \oplus b({}^2\mathcal{A}'_{2k+2})$ consists in a $(n, \min\{h_1, q_2\}, d)$ -code.

The ‘‘goodness’’ of the code \mathcal{C}' depends on the choices of q_i and h_i , so that the number of codewords is not far from the Plotkin bound (1).

□

4.2 Example 2: a good (non optimal) quasi-Hadamard code.

The section “Finding out a liar” in ([1], chap. 17) has provided inspiration for this example.

Suppose that someone thinks of a number between 1 and 10, and that you are supposed to guess which number it is. The rules of the game let you to ask 8 questions (with “yes” or “no” answers), and no more than one lie is allowed.

In order to win, it suffices to get a code capable of correcting up to 1 error, formed from at least 10 codewords (one for every number in the given range). Writing 1 for “yes” and 0 for “no”, now choose the questions so that the binary tuple that the answers generates in each case coincides with the corresponding codeword. This requires that the length of the code should coincide with the number of questions. Summing up, you need a (n, M, d) code so that $n = 8$, $M \geq 10$ and d allows to correct at least 1 error.

Your elementary background on the Theory of Codes indicates that in order to correct e errors you need to use a code of minimum distance d such that $\lfloor \frac{d-1}{2} \rfloor \geq e$. Since $e = 1$, you need $d \geq 3$.

Assume that $d = 4$. Taking into account the Plotkin bound (2), it follows that $n = 2d = 8$ and the number M of codewords is always $M \leq 2n = 16$.

Attending to Levenshtein’s method, the code \mathcal{C}_8

$$\mathcal{C}_8 = \begin{pmatrix} 00000000 \\ 01001011 \\ 01100101 \\ 01110010 \\ 01011100 \\ 00111001 \\ 00101110 \\ 00010111 \\ 11111111 \\ 10110100 \\ 10011010 \\ 10001101 \\ 10100011 \\ 11000110 \\ 11010001 \\ 11101000 \end{pmatrix}$$

related to the matrix H_8 above is optimal for given length 8 and minimum distance 4. Since \mathcal{C}_8 consists of 16 codewords, \mathcal{C}_8 may be used to solve the game.

In spite of this fact, a smaller $(8, M, 4)$ code may be used as well, provided $M \geq 10$.

Consider the quasi-Hadamard matrix M_8 obtained from H_8 by randomly changing the entries located at the 1st, 7th and 8th rows,

$$M_8 = \begin{pmatrix} * & * & * & * & * & * & * & * \\ 1 & - & 1 & 1 & - & 1 & - & - \\ 1 & - & - & 1 & 1 & - & 1 & - \\ 1 & - & - & - & 1 & 1 & - & 1 \\ 1 & - & 1 & - & - & - & 1 & 1 \\ 1 & 1 & - & - & - & 1 & 1 & - \\ * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * \end{pmatrix}$$

Taking into account Theorem 1, we may construct the $(8,10,4)$ code \mathcal{C}'_8 ,

$$\mathcal{C}'_8 = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

related to the matrix M_8 above.

Map every integer i in the range $[1, 10]$ to the i -th codeword c_i in \mathcal{C}'_8 .

Now you should ask the following questions:

1. Is the number greater than 5?
2. Is it less or equal to 4 modulo 10?
3. Is it in the set $\{2, 3, 5, 6, 9\}$?
4. Is it in the range $[3, 7]$?
5. Is it in the set $\{1, 4, 5, 7, 8\}$?
6. Is it even?
7. Is it in the set $\{1, 3, 7, 9, 10\}$?
8. Is it in the set $\{1, 2, 5, 8, 9\}$?

Assume that the vector of answers is $a = (a_1, \dots, a_8)$. Select the unique codeword c_i in \mathcal{C}'_8 whose summation with a modulo 2 produces a tuple with at most one non zero entry. Then the correct number is i , and the player lied precisely when he answered the question which corresponds to the column with the non zero entry.

□

Remark 2. Notice that any quasi-Hadamard matrix of order 8 and depth 5 could have been used as well in order to solve the game. The only variation is the questions to ask. In fact, the questions should be formulated so that if the number to guess is i , then the answer to the j -th question is the i -th entry of the j -th codeword of the code.

Summarizing, depending on the needs of the user, suitable quasi-Hadamard matrices have to be constructed in order to perform the desired error correcting code. Notice that working with quasi-Hadamard matrices and codes instead of Hadamard ones does not mean that functionality is lost (see example 1, for instance). In fact, it often occurs that not all codewords of a given code are actually used for transmissions in practise (see example 2 above). So quasi-Hadamard matrices and quasi-Hadamard codes may suffice to perform transmissions at the entire satisfaction of users, including optimal detection and correction affairs.

Acknowledgments

The authors want to express their gratitude to the referees for their suggestions, which have led to a number of improvements of the paper.

References

1. P.J. Cameron. *Combinatorics: topics, techniques, algorithms*. Cambridge University Press, (1994).
2. U. Feige, S. Goldwasser, S. Safra, L. Lovász and M. Szegedy. Approximating clique is almost NP-complete. *Proceedings 32nd Annual Symposium on the Foundations of Computer Science, FOCS*, 2–12, (1991).
3. J. Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. *Proceedings 37th Annual IEEE Symposium on the Foundations of Computer Science, FOCS*, 627–636, (1996).
4. R.M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, 85–103, 1972.
5. V.I. Levenshtein. Application of the Hadamard matrices to a problem in coding. *Problems of Cybernetics*, **5**, 166–184, 1964.
6. F.J. MacWilliams and N.J.A. Sloane N.J.A. The theory of error-correcting codes. North Holland, New York, 1977.
7. E. Marchiori. Genetic, Iterated and Multistart Local Search for the Maximum Clique Problem. *Proceedings EvoWorkshops 2002*. Eds. S. Cagnoni et al. *LNCS 2279*, Springer-Verlag, Berlin Heidelberg, 112–121, (2002).
8. I. Naboru. Hadamard Graphs I. *Graphs Combin.* **1**, vol. 1, 57–64, 1985.
9. I. Naboru. Hadamard Graphs II. *Graphs Combin.* **1**, vol. 4, 331–337, 1985.
10. M. Plotkin. Binary codes with specified minimum distances. *IEEE Trans. Information Theory*, **6**, 445–450, 1960.