How many weights can a linear code have?

Minjia Shi, Hongwei Zhu, Patrick Sole, Gerard. D. Cohen

Des. Codes Cryptogr. (2019) 87:87–95
https://doi.org/10.1007/s10623-018-0488-z

How many weights can a linear code have?

Minjia Shi[1,2] · **Hongwei Zhu**[2] · **Patrick Solé**[3] · **Gérard D. Cohen**[4]

Received: 29 January 2018 / Revised: 27 March 2018 / Accepted: 17 April 2018 /
Published online: 5 May 2018

Abstract We study the combinatorial function $L(k, q)$, the maximum number of nonzero weights a linear code of dimension k over $\mathbb{F}_q$ can have. We determine it completely for $q = 2$, and for $k = 2$, and provide upper and lower bounds in the general case when both k and q are ≥ 3. A refinement $L(n, k, q)$, as well as nonlinear analogues $N(M, q)$ and $N(n, M, q)$, are also introduced and studied.

Keywords Linear codes · Hamming weight · Perfect difference sets

Mathematics Subject Classification 94B05 · 05B10

1 Introduction

There are several problems in extremal combinatorics on distances in codes. For instance, the famous paper [5] derives an upper bound on the size of a code C over $\mathbb{F}_q$ with exactly s distinct distances:

Communicated by J. Jedwab.

This research is supported by National Natural Science Foundation of China (61672036), Excellent Youth Foundation of Natural Science Foundation of Anhui Province (1808085J20), Technology Foundation for Selected Overseas Chinese Scholar, Ministry of Personnel of China (05015133) and Key projects of support program for outstanding young talents in Colleges and Universities (gxyqZD2016008).

✉ Minjia Shi
mjshi@ahu.edu.cn

1 Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui University, No. 3 Feixi Road, Hefei 230039, Anhui, People's Republic of China

2 School of Mathematical Sciences, Anhui University, Hefei 230601, People's Republic of China

3 CNRS/LAGA, University of Paris 8, 2 rue de la Liberté, 93 526 Saint-Denis, France

4 TelecomParisTech, 46 rue Barrault, 75 013 Paris, France

$$|C| \leq \sum_{j=0}^{s} \binom{n}{j}(q-1)^j. \qquad (1)$$

In the same spirit, other authors have given upper bounds on the size of codes with one or several forbidden distances [6].

In this note, we tackle a related but distinctly different problem: how many distinct weights can a linear code of given dimension over a given finite field have? In other words, we study the combinatorial function $L(k, q)$, the maximum number of nonzero weights a code of dimension k over $\mathbb{F}_q$ may have. While an upper bound is easy to prove (Proposition 2), its tightness is nontrivial[1] and we only manage to establish it in some special cases like $k = 2$ or $q = 2$ (Cf. Theorems 1 and 2). Numerical experiments with very long random codes suggest it is tight for all k's and q's. We leave the question as an open problem. We can also study the more refined function $L(n, k, q)$, the maximum number of nonzero weights an $[n, k]_q$ code may have. This latter function is related to both $L(k, q)$ and Eq. (1) above. The nonlinear counterpart of $L(k, q)$ denoted by $N(M, q)$, can be determined explicitly (Theorem 6). The nonlinear counterpart of $L(n, k, q)$ denoted by $N(n, M, q)$, can also be studied. The rate of convergence of $N(n, M, q)$ towards $N(M, q)$ requires perfect difference sets [3] and primes in short intervals [2] for its careful study.

The material is organized as follows. Section 2 collects the necessary notations and definitions. Section 3 studies upper bounds in the linear code case. Section 4 derives lower bounds in that situation. Section 5 introduces and investigates the function $L(n, k, q)$. Section 6 tackles the nonlinear analogues of $L(k, q)$ and $L(n, k, q)$, denoted by $N(M, q)$, and $N(n, M, q)$, respectively. Section 7 concludes the article. An appendix collects some numerical values, which comfort the Conjecture that Proposition 2 is tight.

2 Definitions and notation

Let q be a prime power, and $\mathbb{F}_q$ denote the finite field of order q. By a **code** of length n over $\mathbb{F}_q$, we shall mean a proper subset of $\mathbb{F}_q^n$. This code is **linear** if it is a $\mathbb{F}_q$-vector subspace of $\mathbb{F}_q^n$. The **dimension** of a code, denoted by k, is equal to its dimension as a vector space. The parameters of such a code are written compactly as $[n, k]_q$. The **Hamming weight** of $x \in \mathbb{F}_q^n$, denoted by $w(x)$, is the number of indices i where $x_i \neq 0$. The **Hamming distance** between $x \in \mathbb{F}_q^n$, and $y \in \mathbb{F}_q^n$, denoted by $d(x, y)$, is defined by $d(x, y) = w(x - y)$. For a given prime power q and given values of k, let $L(k, q)$ denote the largest possible number of nonzero weights a q-ary code can have. If $C(n)$ is a family of codes of parameters $[n, k_n]_q$, the **rate** R is defined as

$$R = \limsup_{n\to\infty} \frac{k_n}{n}.$$

Recall that the q-ary **entropy function** $H_q(.)$ is defined for $0 < y < 1$, by

$$H_q(y) = y \log_q(q-1) - y \log_q(y) - (1-y) \log_q(1-y).$$

3 Upper bounds

The following monotonicity properties of $L(k, q)$ are given without proof.

[1] After submission of this article, a proof was found in [1].

Proposition 1 *For all nonnegative integers k, m and all prime powers q we have:*

$$L(k, q) \leq L(k+1, q),$$
$$L(k, q) \leq L(k, q^m).$$

The next result is trivial but crucial.

Proposition 2 *For all prime powers q, and all integers $k \geq 1$, we have*

$$L(k, q) \leq \frac{q^k - 1}{q - 1}.$$

Proof The total numbers of nonzero codewords of a code of dimension k over $\mathbb{F}_q$ is $q^k - 1$, and all the nonzero multiples of a given codeword share the same weight. □

This bound is met with equality if $q = 2$.

Theorem 1 *For all integers $k \geq 1$, we have*

$$L(k, 2) = 2^k - 1.$$

Proof Denote by G_k the generator matrix of an $[n, k]_q$ with $L(k, 2)$ weights $w_1 < w_2 < \cdots < w_{L(k,2)}$. Define H_{k+1} a matrix obtained from G_k by adding a k by t block of zeros, and by G_{k+1} the matrix obtained by H_{k+1} by adding an a row with first n coordinates zero and last t coordinates $= 1$. The code spanned by the rows of G_{k+1} has all these weights plus the $L(k, 2) + 1$ new weights $t < t + w_1 < \cdots < t + w_{L(k,2)}$. The two sets of weights will have void intersection if $w_{L(k,2)} < t$. This makes $2L(k, 2) + 1$ weights altogether. Note that the rank of G_{k+1} is $k + 1$. Thus we have proved that $L(k + 1, 2) \geq 2L(k, 2) + 1$, which implies by induction, starting from $L(1, 2) = 1$, the lower bound $L(k, 2) \geq 2^k - 1$. The result follows. □

Remark We are now ready to given an alternative proof of Theorem 1. we can exhibit a linear code C with dimension k over $\mathbb{F}_2$ with $2^k - 1$ nonzero weights. Let the generator matrix of C be

$$\left(\begin{array}{c|cc|cccc|cccccccc|c|ccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & \cdots & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0
\end{array}\right),$$

$$\underbrace{\quad}_{a_1}\ \underbrace{\qquad}_{a_2}\ \underbrace{\qquad\qquad}_{a_3}\ \underbrace{\qquad\qquad\qquad\qquad}_{a_4}\ \cdots\ \underbrace{\qquad\qquad\qquad}_{a_k}$$

where $a_1 = 1$, $a_2 = 2$, $a_3 = 2^2$, $a_4 = 2^3$, …, $a_k = 2^{k-1}$. Since $a_{j_1} + a_{j_2} + \ldots + a_{j_t} = \underbrace{(\ldots 010 \ldots 010 \ldots 010 \ldots)_2}_{k}$ in base 2, and the coordinates of $1'$s are $j_1, j_2, \ldots, j_t$, respectively. Thus, we obtain all integers of k bits as possible weights that is the set $\{1, 2, 3, \ldots, 2^k - 1\}$ of cardinality $2^k - 1$ in all.

The bound in Proposition 2 is also tight when $k = 2$.

Theorem 2 *For all prime powers q, we have $L(2, q) = q + 1$.*

Proof Let $\{u, v\}$ be a basis of a code C candidate to have $q + 1$ weights. Denote by S, T the supports of u, v respectively. Let $|S \setminus T| = a$, $|T \setminus S| = b$. On the intersection $S \bigcap T$ assume v is the all-one vector. Denote by ω a primitive root of $\mathbb{F}_q$. Assume $|S \bigcap T| = \binom{q}{2}$ and that u restricted to $S \bigcap T$ is

$$(1, \omega, \omega, \omega^2, \omega^2, \omega^2, \dots, \omega^{q-2}, \dots, \omega^{q-2}),$$

where ω^i occurs $i + 1$ times. With these conventions, we see that the weights of C are

- $w(u) = a + \binom{q}{2}$,
- $w(v) = b + \binom{q}{2}$,
- $w(u - xv) = a + b + \binom{q}{2} - i$ if $x = \omega^{i-1}$ for $i = 1, 2, \dots, q - 1$.

Assume $a < b$. The above weights will be pairwise different if $a+b+\binom{q}{2}-(q-1) > b+\binom{q}{2}$, that is if $a \geq q$. Thus, under these conditions, C counts $2 + q - 1 = q + 1$ nonzero weights. □

Remark The shortest $[n, 2]_q$ code with $L(2, q)$ nonzero weights obtained by this construction has $n = \binom{q}{2} + 2q + 1$.

4 Lower bounds

The easiest lower bound is

Proposition 3 *For all prime powers q, and all integers $k \geq 1$, we have $L(k, q) \geq k$.*

Proof Consider the code $\mathbb{F}_q^k$, of length and dimension k. □

This can be improved to a bound that is exponential in k.

Proposition 4 *For all prime powers q, and all integers $k \geq 1$, we have*

$$L(k + 1, q) \geq 2L(k, q) + 1.$$

In particular, for all integers $k \geq 2$, we have

$$L(k, q) \geq 2^{k-2}q + 2^{k-2} + 1.$$

Proof Same argument as in the first proof of Theorem 1. The second assertion follows by iterating this bound starting from $L(2, q) = q + 1$. □

An asymptotic version of the preceding results is as follows. Define

$$\lambda(q) = \limsup_{n \to \infty} \frac{1}{k} \log_q(L(k, q)).$$

Theorem 3 *For all prime powers q we have*

$$\log_q 2 \leq \lambda(q) \leq 1.$$

In particular $\lambda(2) = 1$.

Proof The first inequality comes from Proposition 4. The second one comes Proposition 2. □

Remark Since we conjecture that the bound of Proposition 2 is tight, it is natural to conjecture that $\lambda(q) = 1$ for all prime powers q.

5 Refinements and asymptotics

A more complex function is $L(n, k, q)$ the largest number of nonzero weights an $[n, k]_q$-code can have. This function is related to $L(k, q)$ in several ways. The following monotonicity properties of $L(n, k, q)$ are given without proof.

Proposition 5 *For all nonegative integers k, m and all prime powers q we have:*

$$L(n, k, q) \leq L(n, k+1, q),$$
$$L(n, k, q) \leq L(n, k, q^m).$$

The three following lemmas are useful for the proof of Theorem 4.

Lemma 1 *For all prime powers q, and all nonnegative integers n, k we have $L(n, k, q) \leq L(k, q)$.*

Proof Immediate from the definitions. □

The new function is also monotone in n.

Lemma 2 *For all prime powers q, and all nonnegative integers n, k we have $L(n, k, q) \leq L(n+1, k, q)$.*

Proof If C is an $[n, k]_q$ code with $L(n, k, q)$ nonzero weights, then C extended by a constant zero coordinate is an $[n+1, k]_q$-code with the same number of nonzero weights. □

Lemma 3 *For all prime powers q, and all nonnegative integers n, k we have $L(n, k, q) \leq n$.*

Proof Note that, by definition of the Hamming weight, a code of length n can have at most n distinct weights. □

We now connect the new function $L(n, k, q)$ with $L(k, q)$.

Theorem 4 *For all prime powers q, and all nonnegative integers k we have*

$$\lim_{n\to\infty} L(n, k, q) = L(k, q).$$

More precisely, there is an integer $n_0 \geq L(k, q)$, such that for all $n \geq n_0$ we have $L(n, k, q) = L(k, q)$.

Proof By Lemmas 1 and 2, the sequence $n \mapsto L(n, k, q)$ is increasing and bounded. Hence, being integral, it converges stably to a limit which can be no other than $L(k, q)$. Let n_0 be such that $L(n_0, k, q) = L(k, q)$. By Lemma 3, we see that $n_0 \geq L(k, q)$. □

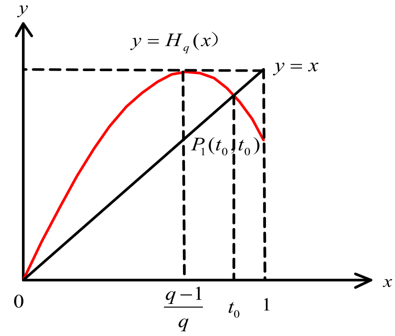
Remark The computations of the Appendix suggest that such an n_0 can be very large. If Proposition 2 is tight then, by Theorem 4 $n_0 \geq \frac{q^k-1}{q-1}$. In the special case $q = 2$, the second proof of Theorem 1 shows that $n_0 = 2^k - 1$.

There is a link to Delsarte's bound (Eq. (1)) quoted in the Introduction.

Proposition 6 *For all prime powers q, and all integers $n \geq k \geq 1$, we have*

$$q^k \leq \sum_{i=0}^{L(n,k,q)} \binom{n}{i} (q-1)^i.$$

Fig. 1 Definition of $t(q)$



Further

$$L(k,q) \leq \frac{\sum_{i=0}^{L(n,k,q)} \binom{n}{i}(q-1)^i - 1}{q-1}.$$

Proof The first assertion is a direct application of Eq. (1) in Introduction ([5, Theorem 4.1]) with $|C| = q^k$, and $s = L(n,k,q)$. Combining this result with Proposition 2 gives the second assertion. □

We give an asymptotic version of the preceding results. Let

$$\mathcal{L}(R) = \limsup_{n\to\infty} \frac{1}{n} \log_q(L(n, \lfloor Rn \rfloor, q)).$$

Theorem 5 *If C_n is a family of codes of rate R then*

$$\mathcal{L}(R) \leq R\lambda(q) \leq H_q(\mathcal{L}(R)).$$

In particular $\mathcal{L}(R) \leq t(q)$, where $t(q)$ is the unique solution in the range $(0, \frac{q-1}{q})$ of $H_q(x) = x$. See Fig. 1.

Proof The first inequality follows by Lemma 1, upon observing that

$$\limsup_{n\to\infty} \frac{1}{n}(L(k,q)) = R\lambda(q).$$

The second inequality comes from the second assertion of Proposition 6, after using standard entropic estimates [7]. The second assertion is obtained by combining the first and second inequality. □

Define the domain $\mathcal{D}$ as the set of points in the plane $(R, \mathcal{L})$ that are realized by a family of codes. By the preceding result, this domain is contained in the domain of boundaries given by, counterclockwise, in Fig. 2 by

1. the straight line $\mathcal{L} = R$ from $R = 0$ till $R = t(q)$,
2. the horizontal line $\mathcal{L} = t(q)$ from $R = t(q)$ till $R = 1$,
3. the vertical line $R = 1$ from $\mathcal{L} = t(q)$ till $\mathcal{L} = \frac{q-1}{q}$,
4. the curve $\mathcal{L} = H_q^{-1}(R)$, from $R = 1$ till $R = 0$.

Determining the domain $\mathcal{D}$ explicitly, in the same way as the domain of packing and covering codes in [4] is a challenging open problem.

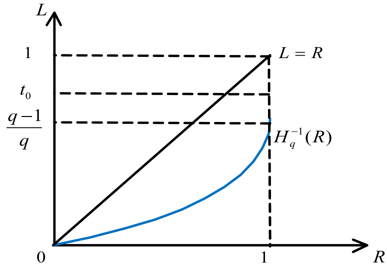


Fig. 2 Boundaries of domain $\mathcal{D}$

6 Nonlinear codes

Warning In this section only q is an arbitrary integer > 1.

The nonlinear analogue of the function $L(k, q)$ is the function $N(M, q)$ which is the largest number of distances between two codewords of an unrestricted code of size M over some finite alphabet A_q of size q. This function is completely determined in the following Theorem.

Theorem 6 *For all integers $M \geq 2$, we have*

$$N(M, q) = \binom{M}{2}.$$

Proof By definition we have immediately $N(M, q) \leq \binom{M}{2}$. By an inductive process, we construct a code C_M with $\binom{M}{2}$ distances. To simplify matters take $q = 2$. We search for codes in a special form where nonzero codewords are of the form $(1, 1, \ldots, 1, 0, \ldots, 0)$, that is a run of ones followed by a run of zeros. Thus the distance between two such codewords is equal to the difference of their weights. For $M = 2$, we may take the length 1 code $\{0, 1\}$. Assume C_M is constructed with codewords of successive weights $w_0 = 0 < w_1 < \cdots < w_{M-1}$. We construct a code C_{M+1} by adding a tail of zeros to C_M on the right, of length to be specified later, and by adding a new codeword of weight w_M. The new distances are M in number, given by $w_M, w_M - w_1, \ldots, w_M - w_{M-1}$. These distances are pairwise distinct because $(w_M - w_i) - (w_M - w_j) = w_j - w_i$. To make sure they are distinct from the distances in C_M, we must check that

$$(w_M - w_i) \neq w_j - w_k,$$

with i, j, k distinct nonnegative integers $\leq M - 1$. This is enforced if we take w_M large enough. This condition on w_M, in turn, will determine how long the tail must be. Since $\binom{M+1}{2} - \binom{M}{2} = M$, we are done. □

The nonlinear analogue of the function $L(n, k, q)$ is the function $N(n, M, q)$ which is the largest number of distances between two codewords of an unrestricted code of size M and length n over some alphabet A_q, of size q.

The analogue of Theorem 4 in this context is as follows. The proof is similar and omitted.

Theorem 7 *For all integers $q > 1$, and all nonnegative integers M we have*

$$\lim_{n\to\infty} N(n, M, q) = N(M, q).$$

More precisely, there is an integer $n_0 \geq N(M, q)$, such that for all $n \geq n_0$ we have $N(n, M, q) = N(M, q)$.

Denote by $N_0(M, q)$ the smallest integer n such that $N(n, M, q) = N(M, q)$.

Proposition 7 *If $M - 1$ is a power of a prime, then $N_0(M, q) \leq 2N(M, q) + 1$.*

Proof Assume $M = s + 1$, where s is a power of a prime. We know there is a Singer difference set [9] $S = \{v_0, v_1, \ldots, v_{s+1}\}$, with parameters $(s^2 + s + s, s + 1, 1)$. Consider the $s + 1$ by $s^2 + s + 1$ matrix with rows g_i, when g_i contains v_i consecutive ones to the left and zeros elsewhere. The Hamming distance from g_i to g_j is $|v_i - v_j|$. The code formed by the M rows of this matrix has length $s^2 + s + 1 = M^2 - M + 1 = 2\binom{M}{2} + 1$ and $\binom{M}{2}$ distances, by the design property. Hence, in this case, $n_0 \leq 2\binom{M}{2} + 1$. For instance, if $s = 2$, we have $S = \{1, 2, 4\}$, and the code is $\{1000000, 1100000, 1111000\}$. See [3, p. 264] for details on, and examples of Singer difference sets. □

Denote, for any integer t, by $pp(t)$ the smallest prime power $\geq t$.

Corollary 1 *For all integers $M > 1$, we have*

$$N_0(M, q) \leq 2N(pp(M - 1) + 1, q) + 1 \leq 2N(2M, q) \sim 8\binom{M}{2}.$$

Proof We claim that $N_0(M, q)$ is a nondecreasing function of M. The first inequality will follow by the previous theorem, since $M \leq pp(M - 1) + 1$. To prove the claim note that, if we have a set of $M + 1$ vectors of length $N_0(M + 1, q)$, with $\binom{M+1}{2}$ distances, removing any vector will result into a set of M vectors with $\binom{M+1}{2} - M = \binom{M}{2}$ distances. Hence $N_0(M, q) \leq N_0(M + 1, q)$. The second inequality follows by the crude bound $pp(x) \leq 2x$, valid for any positive integer x. □

Remark It is possible to reduce the upper bound on $pp(x)$ to $pp(x) \leq x + x^a$, with $a = 0.525$, building on recent estimates on the existence of primes in short intervals [2]. This sharpens the upper bound on $N_0(M, q)$ to $2N(M + O(M^a), q) + 1 \sim 2\binom{M}{2}$, for $M \to \infty$.

7 Conclusion and open problems

In this note, we have studied a problem of extremal combinatorics: maximizing the number of distinct nonzero weights a linear code can have. We conjecture, based on extensive numerical calculations on very long codes, that the bound of Proposition 2 is tight but cannot prove it. A proof was found later in [1]. A recursive approach in the manner of the proof of Theorem 6 would require to produce q^k new weights to go from $L(k, q)$ to $L(k + 1, q)$. But a code achieving $L(k, q)$ has only $\frac{q^k-1}{q-1} < q^k$ distinct weights. Thus establishing the tightness of Proposition 1 is the main open problem of this note. Sharpening the upper bound on $N_0(M, q)$ of Corollary 1 is also a challenging question. Determining explicitly the domain $\mathcal{D}$ of Sect. 5 seems to require better lower bounds on $L(n, kq)$ that those at our disposal.

Appendix: numerical examples

We provide lower bounds on $L(k, q)$ by computing the number of weights in long random codes produced by the computer package Magma [8]. We give some numerical examples in Table 1 about the lower bound of Proposition 4.

When n is in the millions, we can find linear $[n, k]_q$-codes that meet the upper bound in Proposition 2: see Table 2.

Table 1 Proposition 4

k	3	4	4	6	6	10	10	12	12	12
q	3	5	8	9	13	16	25	29	49	121
$L(k, q) \geq$	11	29	41	177	241	4609	6913	31, 745	52, 225	125, 953

Table 2 $n = 6, 000, 000$

k	3	3	3	3	3	3	3	4	4	4	5	5
q	3	4	5	7	8	9	11	3	4	5	3	4
$L(k, q) =$	13	21	31	57	73	91	133	40	85	156	121	341

References

1. Alberson T., Neri A.: Maximum weight spectrum codes. https://arxiv.org/pdf/1803.04020.pdf.
2. Baker R.C., Harman G., Pintz J.: The difference between consecutive primes, II. Proc. Lond. Math. Soc. **83**(3), 532–562 (2001).
3. Beth T., Jungnickel D., Lenz H.: Design theory. BI-Institut, Mannheim, Wien, Zurich (1985).
4. Cohen G., Honkala I., Litsyn S., Solé P.: Long packing and covering codes. IEEE Trans. Inf. Theory **43**, 1617–1619 (1997).
5. Delsarte P.: Four fundamentals parameters of a code and their combinatorial significance. Inf. Control **23**, 407–438 (1973).
6. Enomoto H., Frankl P., Ito N., Nomura K.: Codes with given distances. Graph Comb. **3**, 25–38 (1987).
7. Huffman W.C., Pless V.: Fundamentals of Error Correcting Codes. Cambridge University Press, Cambridge (2003).
8. http://magma.maths.usyd.edu.au/magma/. Accessed 28 Jan 2018
9. Singer J.: A theorem in finite geometry and some applications to number theory. Trans. Am. Math Soc. **43**, 377–385 (1938).