Perfect single error-correcting codes in the Johnson scheme

Daniel M. Gordon

***Abstract*— Delsarte conjectured in 1973 that there are no nontrivial pefect codes in the Johnson scheme. Etzion and Schwartz recently showed that perfect codes must be k-regular for large k, and used this to show that there are no perfect codes correcting single errors in $J(n, w)$ for $n \leq 50{,}000$. In this paper we show that there are no perfect single error-correcting codes for $n \leq 2^{250}$.**

I. Introduction

The Johnson graph $J(n, w)$ has vertices corresponding to V_w^n, the w-subsets of the set $\mathcal{N} = \{1, 2, \ldots, n\}$, with two vertices adjacent if their intersection has size $w - 1$.

The distance between two w-sets is half the size of their symmetric difference. The e-sphere of a point, the set of all w-sets within distance e, has cardinality

$$\Phi_e(n, w) = \sum_{i=0}^{e} \binom{w}{i} \binom{n-w}{i}.$$

A code $\mathcal{C} \subset J(n, w)$ is called *e-perfect* if the e-spheres of all the codewords of $\mathcal{C}$ form a partition of V_w^n. Delsarte [2] conjectured that no nontrivial perfect codes exist in $J(n, w)$.

Etzion and Schwartz [3] introduced the concept of k-regular codes. In this paper we use their results to improve the lower bound on the size of a 1-perfect code. The method of proof will be to look at the factors of $\Phi_1(w, a)$. We show that $\Phi_1(w, a)$ is squarefree, and for each prime $p_i | \Phi_1(w, a)$, there is an integer α_i such that $p_i^{\alpha_i}$ must be close to $n - w$. Then we will show that the α_i's are distinct and pairwise coprime, and the sum of their reciprocals is close to two. A computer search for perfect powers in short intervals then shows that no such codes exist with $n < 2^{250}$.

For the rest of this paper we will deal with the case $e = 1$, and write $n = 2w + a$. This may be done without loss of generality, since the complement of an e-perfect code in $J(n, w)$ is e-perfect in $J(n, n - w)$. Also, to simplify the statements of theorems, we will assume throughout the paper that $\mathcal{C}$ is a nontrivial 1-perfect code in $J(n, w)$.

II. Regularity of 1-Perfect codes

In this section we summarize the results of Etzion and Schwartz [3] that we will need. Let $\mathcal{A}$ be a k-subset of $\mathcal{N} = \{1, 2, \ldots, n\}$. For all $0 \leq i \leq k$, define

$$\mathcal{C}_{\mathcal{A}}(i) = |\{c \in \mathcal{C} : |c \cap \mathcal{A}| = i\}|,$$

and for each $\mathcal{I} \subseteq \mathcal{A}$, define

$$\mathcal{C}_{\mathcal{A}}(\mathcal{I}) = |\{c \in \mathcal{C} : c \cap \mathcal{A} = \mathcal{I}\}|.$$

$\mathcal{C}$ is *k-regular* if:

1) There exist numbers $\alpha(0), \alpha(1), \ldots, \alpha(k)$ such that for any k-set $\mathcal{A}$ in $\mathcal{N}$, $\mathcal{C}_{\mathcal{A}}(i) = \alpha(i)$, for $i = 0, 1, \ldots, k$.
2) For any k-set $\mathcal{A}$ in $\mathcal{N}$, there exist numbers $\beta_{\mathcal{A}}(0), \beta_{\mathcal{A}}(1), \ldots, \beta_{\mathcal{A}}(k)$ such that if $\mathcal{I} \subseteq \mathcal{A}$, then $\mathcal{C}_{\mathcal{A}}(\mathcal{I}) = \beta_{\mathcal{A}}(|\mathcal{I}|)$.

Etzion and Schwarz give a necessary condition for a code to be regular:

Theorem 1: If $\mathcal{C}$ is k-regular, then

$$\Phi_1(w, a) = 1 + w(w + a) \left| \binom{2w + a - i}{w + a} \right. \quad (1)$$

for $i = 0, \ldots, k$.

They then show that 1-perfect codes must be highly regular.

Theorem 2: $\mathcal{C}$ is k-regular if the polynomial

$$\sigma_1(w, a, m) = m^2 - (2w + a + 1)m + w(w + a) + 1 \quad (2)$$

has no integer roots for $1 \leq m \leq k$.

Let

$$L(w, a) = \frac{2w + a + 1 - \sqrt{(a + 1)^2 + 4(w - 1)}}{2}.$$

The smallest root of (2) is $L(w, a)$, so

Theorem 3: $\mathcal{C}$ is k-regular for any $k < L(w, a)$.

This means that we can rule out 1-perfect codes by showing that there is some i with $0 \leq i \leq L(w, a)$ such that (1) is not satisfied. $L(w, a)$ is an increasing function of a, so

Lemma 1: $L(w, a) \geq L(w, 0) > w - \lceil \sqrt{w} \rceil$.

Lemma 2: We have

$$0 < a < w/2.$$

Proof: Theorem 13 in [3], which is a strengthening of a theorem of Roos [7], gives $a < w - 3$. If $a = 0$ then $\mathcal{C}$ is a trivial code.

If $a \geq w/2$, then

$$L(w, a) > L\left(w, \frac{w - 7}{2}\right) = w - 2,$$

so $\mathcal{C}$ is $(w - 2)$-regular. $\mathcal{C}$ is also $(w - 1)$-regular, since $\sigma_1(w, a, w - 1) = a - (w - 3) \neq 0$ for $a < w - 3$.

Since $\mathcal{C}$ corrects single errors, any two codewords are at least distance 3 apart in $J(n, w)$. Let $\mathcal{A}$ be a $(w - 1)$-set contained in some codeword c_1. Remove any element of $\mathcal{A}$ and add one not in c_1 to get a new $(w - 1)$-set $\mathcal{A}'$. Since $\mathcal{C}$ is $(w - 1)$-regular, there is a codeword c_2 containing $\mathcal{A}'$, but c_1 and c_2 have distance 2 in $J(n, w)$, a contradiction. ∎

III. Divisors of $\Phi_1(w, a)$

We will derive necessary conditions for 1-perfect codes by looking at possible prime divisors of $\Phi_1(w, a)$. One tool will be:

Lemma 3: (Kummer) Let p be a prime. The number of times p appears in the factorization of $\binom{a}{b}$ equals the number of carries when adding b to $a - b$ in base p.

Theorem 3 and Lemmas 1 and 3 imply

Corollary 1: If p is a prime with $p^k | \Phi_1(w, a)$, then there are at least k carries when adding $w + a$ to $j = w - i$ for $j = \lceil \sqrt{w} \, \rceil + 1, \lceil \sqrt{w} \, \rceil + 2, \ldots, w$.

The author is with the IDA Center for Communications Research, San Diego, CA 92121-1969 (email: gordon@ccrwest.org)

Let

$$w+a=(r_m,r_{m-1},\ldots,r_1,r_0)_p \tag{3}$$

be the base p representation of $w+a$, with $r_m \geq 1$. Let $l=\lfloor m/2\rfloor$.

Lemma 4: $r_i=p-1$ for $i=l+1,l+2,\ldots,m$.

Proof: For any i with $\lceil\sqrt{w}\rceil+1\leq p^i\leq w$, adding p^i to $w+a$ must have a carry by Corollary 1, so the lemma follows for $i=l+1,\ldots,m-1$. To complete the proof, we need to show that $w\geq p^m$. We have

$$w+a\geq p^m+(p-1)p^{m-1}\geq \frac{3}{2}p^m.$$

Since $a<w/2$ by Lemma 2, this implies $w>p^m$. ■

Theorem 4: $\Phi_1(w,a)$ must be squarefree.

Proof: Adding p^m to $w+a$ has only one carry, so by Corollary 1 only one power of p divides $\Phi_1(w,a)$. ■

Theorem 5: For any prime p dividing $\Phi_1(w,a)$, let $\alpha=m+1=\lfloor\log_p(w+a)\rfloor+1$. Then

$$p^\alpha-\lceil\sqrt{w}\rceil-1\leq w+a<p^\alpha \tag{4}$$

Proof: We have $w+a<p^\alpha$ from (3). By Lemma 4, we must have $r_i=p-1$ for $i=l+1,l+2,\ldots,m$. Let

$$(t_l,t_{l-1},\ldots,t_0)_p$$

be the base p representation of $\lceil\sqrt{w}\,\rceil$. The left inequality of (4) is equivalent to

$$\begin{aligned}p^\alpha-1-(w+a) &= (p-1-r_l,\ldots,p-1-r_0)_p\\ &\leq (t_l,t_{l-1},\ldots,t_0)_p=\lceil\sqrt{w}\rceil.\end{aligned}$$

If this is not satisfied, let i be the largest integer such that $p-1-r_i>t_i$. The number $(t_l,t_{l-1},\ldots,t_{i+1},t_i+1,0,\ldots,0)_p$ is greater than $\lceil\sqrt{w}\rceil$ and has no carries when when added to $w+a$ in base p, which contradicts Corollary 1. ■

Thus we have that p^α is in a short interval around $w+a$. We will use this result in the following form:

Corollary 2: For a prime p dividing $\Phi_1(w,a)$, we have

$$0<\log_{w+a}p-\frac{1}{\alpha}<\frac{1}{\alpha}\left(\frac{1}{\sqrt{w+a}}+\frac{4}{(w+a)}\right). \tag{5}$$

Proof: From (4), we have

$$\begin{aligned}p^\alpha>w+a &\geq p^\alpha\left(1-\frac{\lceil\sqrt{w}\rceil+1}{p^\alpha}\right)\\ &> p^\alpha\left(1-\frac{1}{\sqrt{w+a}}-\frac{2}{w+a}\right)\end{aligned}$$

using $\lceil\sqrt{w}\rceil+1<\sqrt{w+a}+2$. Taking the log base $w+a$, we have

$$\alpha\log_{w+a}p>1>\alpha\log_{w+a}p+\log_{w+a}\left(1-\frac{1}{\sqrt{w+a}}-\frac{2}{w+a}\right)$$

Using the bound $-\log(1-x)<x+x^2$ for $x<1/2$ gives the corollary. ■

IV. Powers in Short Intervals

Theorem 5 shows that for a 1-perfect code to exist, several prime powers must be close to $w+a$. Having a large number of prime powers in a short interval seems unlikely. Loxton [6] showed (a gap in the proof was later fixed by Bernstein [1]) that the number of perfect powers in $[w,w+\sqrt{w}\,]$ is at most

$$\exp(40\sqrt{\log\log w\log\log\log w}).$$

Loxton conjectured that the number of perfect powers in such an interval is bounded by a constant, but a proof seems very far off.

For the rest of this paper, take

$$p_1p_2\ldots p_r=\Phi_1(w,a)=1+w(w+a). \tag{6}$$

Taking the log of (6) gives

$$\sum_{i=1}^r\log_{w+a}p_i=\log_{w+a}(w(w+a)+1),$$

so

$$\begin{aligned}0 &< \sum_{i=1}^r\log_{w+a}p_i-(1+\log_{w+a}w)\\ &= \log_{w+a}(1+\frac{1}{w(w+a)})\\ &\leq \frac{1}{w(w+a)}.\end{aligned} \tag{7}$$

Theorem 6:

$$\left|\sum_{i=1}^r\frac{1}{\alpha_i}-(1+\log_{w+a}w)\right|<\frac{4}{\sqrt{w+a}}.$$

Proof: If $\sum_{i=1}^r\frac{1}{\alpha_i}-(1+\log_{w+a}w)\geq 0$, then the theorem follows immediately from (7) and Corollary 2. Otherwise, summing (5) we have

$$\begin{aligned}0 &< (1+\log_{w+a}w)-\sum_{i=1}^r\frac{1}{\alpha_i}\\ &< \sum_{i=1}^r\log_{w+a}p_i-\sum_{i=1}^r\frac{1}{\alpha_i}\\ &< \sum_{i=1}^r\frac{1}{\alpha_i}\left(\frac{1}{\sqrt{w+a}}+\frac{4}{w+a}\right)\\ &< 2\frac{2}{\sqrt{w+a}}.\end{aligned}$$

■

Clearly the constant 4 in Theorem 6 can be strengthened, but this will be enough for our purposes.

For $0<a<w/2$, we have $w+a<3w/2$, so

$$1-\log_{w+a}3/2<\log_{w+a}w<1$$

and Theorem 6 says that we have an Egyptian fraction representing a number close to 2. Etzion and Schwartz showed that there are no 1-perfect codes with $n\leq 50000$, and so

$$\frac{1}{\alpha_1}+\frac{1}{\alpha_2}+\ldots\frac{1}{\alpha_r}\in[1.934,2.026]. \tag{8}$$

Lemma 5: The α_i's are distinct and pairwise coprime.

Proof: We cannot have $\alpha_i = \alpha_j = 1$, since then $p_i, p_j > (w+a)$ implies $p_i p_j > 1+w(w+a) = \Phi_1(w,a)$, contradicting (6).

Suppose we have α_i, α_j with $\gcd(\alpha_i, \alpha_j) = g > 1$. Then by Theorem 5, $p_i^{\alpha_i}$ and $p_j^{\alpha_j}$ are two g^{th} powers in an interval around $w+a$ of length $\sqrt{w+a}$, which is impossible. ■

For an integer k, let $p^-(k)$ denote the smallest prime factor of k.

Corollary 3: Some α_i has $p^-(\alpha_i) \geq 7$.

Proof: If there are more than four α's, clearly one of them must have a prime factor bigger than 5. For four α's, the set $\{1, 2, 3, 5\}$ has sum of reciprocals 2.033, which by (8) is too big, and an easy computation finds that any set of powers of these numbers has a sum of reciprocals that is too small. The largest is $\{1, 2, 3, 25\}$, with sum 1.8733. ■

Let $\gamma(n)$ denote the largest squarefree divisor of n. The abc conjecture asserts that, for any $\epsilon > 0$ there are only finitely many integers a, b and c such that $a + b = c$ and

$$\max\{a, b, c\} \leq C_\epsilon \gamma(abc)^{1+\epsilon}.$$

See [4] for information and references about the abc conjecture

For any choice of α's satisfying (8), Masser-Oesterlé's abc conjecture implies there are only a finite number of solutions. For example, take $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$, and $\alpha_4 = 7$. Let $a = p_3^3$, $c = p_4^7$ and b be their difference, which is at most $\max\{p_3^{3/2}, p_4^{7/2}\}$ by Theorem 5. Then

$$\begin{aligned} \max\{a, b, c\} \approx w + a &\leq C_\epsilon p_3 p_4 c \\ &< (w+a)^{(1+\epsilon)(1/3+1/7+1/2)} \\ &< (w+a)^{0.98} \end{aligned}$$

for all but finitely many w's.

V. A New Lower Bound for n

While we cannot show that there are no perfect codes, Theorem 5 gives us an efficient way to search for possible codes, by searching for powers in short intervals.

To show a bound of 2^C for n, we need to check for primes $a, b \geq 2$ and integers $3 \leq p, q < C$ with

$$0 < a^p - b^q < \sqrt{a^p}.$$

It suffices to consider prime values of p and q, since any kth power is also a $p^-(k)$th power. It is possible to run through the possibilities efficiently. Let $\{p_1 = 3, p_2 = 5, \ldots, p_k\}$ be the odd primes up to C. The following procedure will find all pairs i, j and integers b_i, b_j for which $b_i^{p_i}$ and $b_j^{p_j}$ are close:

1) Start with $b_1 = b_2 \cdots = b_k = 2$. Compute powers $c_i = b_i^{p_i}$ for $i = 1, 2, \ldots, k$.
2) Let c_i be the smallest power, and c_j the second smallest. Compare them to see if they are close enough.
3) Increment the base b_i, recompute c_i, and continue.
4) Stop when all powers are larger than 2^C.

If two powers less than 2^C are in a short interval, they will eventually be the two smallest powers in the list, and will be found. A heap (see, for example, [5]) is an efficient data structure to maintain the powers in, requiring only one comparison to find the two smallest powers, and $\leq \log_2 k$ steps to reorder the heap after changing c_i.

$p_1^{\alpha_1}$	$p_2^{\alpha_2}$	difference
2^7	5^3	3
13^3	3^7	10
3251^3	32^7	83883
33^7	3493^3	178820
1965781^3	498^7	1539250669

TABLE I

Pairs of Higher Powers in Short Intervals up to 2^{109}

Note that the above algorithm looks for any integers b_i and b_j with powers in a short interval, not just primes. Only considering primes would reduce the number of comparisons, but complicate the rule for stepping the bases b_i.

In five hours on a 2.6 GHz Opteron, an implementation of this algorithm eliminated everything up to 2^{109}. It found 60 powers higher than squares in short intervals, most of which involved a cube and fifth power. By Corollary 3, we may discount these. The only higher powers are given in Table I.

Only the first two pairs are powers of primes, and they are in the range already eliminated by Etzion and Schwartz's result. The larger ones all involve at least one composite, so they do not result in a 1-perfect code. Therefore we have

Theorem 7: There are no 1-perfect codes in $J(n, w)$ for all $n < 2^{109}$.

Finally, we may bootstrap this result to a stronger one. Using this larger bound in Theorem 6, we can tighten (8) to

$$\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \ldots \frac{1}{\alpha_r} \in [1.99, 2.001].$$

No set of four α_i's have a sum of reciprocals in this interval, and the only sets of five that do are $\{1, 2, 3, 7, k\}$, where $k \in [41, 71]$ with $\gcd(k, 2 \cdot 3 \cdot 7) = 1$. Any set of six α_i's clearly have two α's with a factor ≥ 7, so we have

Corollary 4: At least two α_i's have $p^-(\alpha_i) \geq 7$.

Therefore we may do a search as above, but starting with $p_1 = 7$ instead of 3. The search work is proportional to $2^{C/p_1}$, so this greatly reduces the search time. A search for seventh and higher powers up to 2^{250} in a short interval took four hours and found none, so

Theorem 8: There are no 1-perfect codes in $J(n, w)$ for all $n < 2^{250}$.

Acknowledgments. The author would like to thank the anonymous referee, who suggested changes which greatly improved the presentation of this paper, and pointed out Lemma 2.

References

[1] D. J. Bernstein. Detecting perfect powers in essentially linear time. *Math. Comp.*, 67:1253–1283, 1998.

[2] P. Delsarte. An algebraic approach to association schemes of coding theory. *Philips J. Res.*, 10:1–97, 1973.

[3] T. Etzion and M. Schwartz. Perfect constant-weight codes. *IEEE Trans. Info. Th.*, 50:2156–2165, 2004.

[4] A. Granville and T. Tucker. It's as easy as abc. *Notices AMS*, 49:1224–1231, 2002.

[5] Donald E. Knuth. *Sorting and Searching*, volume 3 of *The Art of Computer Programming*. Addison-Wesley, Reading, Massachusetts, second edition, 1998.

[6] J. H. Loxton. Some problems involving powers of integers. *Acta Arith.*, 46:113–123, 1986.

[7] C. Roos. A note on the existence of perfect constant weight codes. *Discr. Math.*, 47:121–123, 1983.