SQUARES AND DIFFERENCE SETS IN FINITE FIELDS

C. Bachoc[1]
Univ Bordeaux, Institut de Mathématiques de Bordeaux, Talence, France
bachoc@math.u-bordeaux1.fr

M. Matolcsi[2]
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary
matolcsi.mate@renyi.mta.hu

I. Z. Ruzsa[3]
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary
ruzsa.z.imre@mta.renyi.hu

Received: 4/29/13, Accepted: 11/17/13, Published: 11/27/13

Abstract

For infinitely many primes $p = 4k + 1$ we give a slightly improved upper bound for the maximal cardinality of a set $B \subset \mathbb{Z}_p$ such that the difference set $B - B$ contains only quadratic residues. Namely, instead of the "trivial" bound $|B| \leq \sqrt{p}$ we prove $|B| \leq \sqrt{p} - 1$, under suitable conditions on p. The new bound is valid for approximately three quarters of the primes $p = 4k + 1$.

1. Introduction

Let q be a prime-power, say $q = p^k$. We will be interested in estimating the maximal cardinality $s(q)$ of a set $B \subset \mathbb{F}_q$ such that the difference set $B - B$ contains only squares. While our main interest is in the case $k = 1$, we find it instructive to compare the situation for different values of k.

This problem makes sense only if -1 is a square; to ensure this we assume $q \equiv 1 \pmod 4$. The universal upper bound $s(q) \leq \sqrt{q}$ can be proved by a pigeonhole argument or by simple Fourier anlysis, and it has been re-discovered several times (see [8, Theorem 3.9], [12, Problem 13.13], [4, Proposition 4.7], [3, Chapter XIII, Theorem 14], [11, Theorem 31.3], [10, Proposition 4.5], [7, Section 2.8] for various

[1]This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Programme IdEx Bordeaux (ANR-10-IDEX-03-02).
[2]Supported by ERC-AdG 228005, and OTKA Grants No. K81658 and the Bolyai Scholarship.
[3]Supported by ERC-AdG 228005, and OTKA Grants No. K81658. All authors supported by the CNRS-MTA "Haagacant" project.

proofs). For even k we have equality, since $\mathbb{F}_{p^k}$ can be constructed as a quadratic extension of $\mathbb{F}_{p^{k/2}}$, and then every element of the embedded field $\mathbb{F}_{p^{k/2}}$ will be a square. It is known that every case of equality can be obtained by a linear transformation from this one, [2].

Such problems and results are often formulated in terms of the Paley graph P_q, which is the graph with vertex set $\mathbb{F}_q$ and an edge between x and y if and only if $x-y=a^2$ for some non-zero $a \in \mathbb{F}_q$. Paley graphs are self-complementary, vertex and edge transitive, and $(q,(q-1)/2,(q-5)/4,(q-1)/4)$-strongly regular (see [3] for these and other basic properties of P_q). Paley graphs have received considerable attention over the past decades because they exhibit many properties of random graphs $G(q,1/2)$ where each edge is present with probability $1/2$. Indeed, P_q form a family of *quasi-random* graphs, as shown in [5].

With this terminology $s(q)$ is the *clique number* of P_q. The general lower bound $s(q) \geq (\frac{1}{2}+o(1))\log_2 q$ is established in [6], while it is proved in [9] that $s(p) \geq c\log p \log\log\log p$ for infinitely many primes p. The "trivial" upper bound $s(p) \leq \sqrt{p}$ is notoriously difficult to improve, and it is mentioned explicitly in the selected list of problems [7]. The only improvement we are aware of concerns the special case $p=n^2+1$ for which it is proved in [13] that $s(p) \leq n-1$ (the same result was proved independently by T. Sanders – unpublished, personal communication). It is more likely, heuristically, that the lower bound is closer to the truth than the upper bound. Numerical data [16, 15] up to $p<10000$ suggest (very tentatively) that the correct order of magnitude for the clique number of P_p is $c\log^2 p$ (see the discussion and the plot of the function $s(p)$ at [17]).

In this note we prove the slightly improved upper bound $s(p) \leq \sqrt{p}-1$ for the *majority* of the primes $p=4k+1$ (we will often suppress the dependence on p, and just write s instead of $s(p)$).

We will denote the set of nonzero quadratic residues by Q, and that of nonzero non-residues by NQ. Note that $0 \notin Q$ and $0 \notin NQ$.

2. The Improved Upper Bound

Theorem 2.1. *Let q be a prime-power, $q=p^k$, and assume that k is odd and $q \equiv 1 \pmod 4$. Let $s=s(q)$ be the maximal cardinality of a set $B \subset \mathbb{F}_q$ such that the difference set $B-B$ contains only squares.*
(i) If $[\sqrt{q}]$ is even then $s^2+s-1 \leq q$; *(ii) if $[\sqrt{q}]$ is odd then $s^2+2s-2 \leq q$.*

Proof. The claims hold if $s<[\sqrt{q}]$. Hence we may assume that $s \geq [\sqrt{q}]$.

Lemma 2.2. *Let $D \subset \mathbb{F}_q$ be a set such that $D \subset NQ$, $D-D \subset Q \cup \{0\}$. With $r=|D|$ we have*

$$s(q) \leq 1+\frac{q-1}{2r}. \tag{1}$$

Proof. Let B be a maximal set such that $B - B \subset Q \cup \{0\}$, $|B| = s(q) = s$. Consider the equation $b_1 - b_2 = zd$, $b_1, b_2 \in B$, $d \in D$, $z \in NQ$. This equation has exactly $s(s-1)r$ solutions; indeed, every pair of distinct $b_1, b_2 \in B$ and a $d \in D$ determines z uniquely. On the other hand, given b_1 and z, there can be at most one pair b_2 and d to form a solution. Indeed, if there were another pair b_2', d', then by substracting the equations $b_1 - b_2 = zd$, $b_1 - b_2' = zd'$ we get $(b_2' - b_2) = z(d - d')$, a contradiction, as the left hand side is a square and the right hand side is not. This gives $s(s-1)r \le s(q-1)/2$ as wanted. □

We try to construct such a set D in the form $D = (B - t) \cap NQ$ with a suitable t. The required property then follows from $D - D \subset B - B$.

Let χ denote the quadratic multiplicative character, i.e., $\chi(t) = 1$ according to whether $t \in Q$ or $t \in NQ$ (and $\chi(0) = 0$). Let

$$\varphi(t) = \sum_{b \in B} \chi(b - t). \tag{2}$$

Clearly $\varphi(t) = |(B - t) \cap Q| - |(B - t) \cap NQ|$, and hence for $t \notin B$ we have

$$|(B - t) \cap NQ| = \frac{s - \varphi(t)}{2}.$$

To find a large set in this form we need to find a negative value of φ.

We list some properties of this function. For $t \in B$ we have $\varphi(t) = s - 1$, and otherwise $\varphi(t) \le s - 2$, $\varphi(t) \equiv s \pmod 2$ (the inequality expresses the maximality of B). Furthermore, $\sum_t \varphi(t) = 0$, and, since translations of the quadratic character have the quasi-orthogonality property

$$\sum_t \chi(t + a)\chi(t + b) = -1$$

for $a \ne b$, we conclude that

$$\sum_t \varphi(t)^2 = s(q-1) - s(s-1) = s(q-s).$$

By subtracting the contribution of $t \in B$ we obtain

$$\sum_{t \notin B} \varphi(t) = -s(s-1); \qquad \sum_{t \notin B} \varphi(t)^2 = s(q-s) - s(s-1)^2 = s(q - s^2 + s - 1).$$

These formulas assume an even nicer form by introducing the function $\varphi_1(t) = \varphi(t) + 1$:

$$\sum_{t \notin B} \varphi_1(t) = q - s^2, \tag{3}$$

$$\sum_{t \notin B} \varphi_1(t)^2 = (s+1)(q - s^2). \tag{4}$$

As a byproduct, the second equation shows the familiar estimate $s \leq \sqrt{q}$, so we have $s = [\sqrt{q}] < \sqrt{q}$ (recall that we assume that $s \geq [\sqrt{q}]$, the theorem being trivial otherwise).

Now we consider separately the cases of odd and even s. If s is even, then, since $\sum_{t \notin B} \varphi(t) < 0$ and each summand is even, we can find a t with $\varphi(t) \leq -2$. This gives us an r with $r \geq (s+2)/2$, and on substituting this into (1) we obtain the first case of the theorem.

If s is odd, we claim that there is a t with $\varphi(t) \leq -3$. Otherwise we have $\varphi(t) \geq -1$, that is, $\varphi_1(t) \geq 0$ for all $t \notin B$. We also know $\varphi(t) \leq s-2$, $\varphi_1(t) \leq s-1$ for $t \notin B$. Consequently

$$\sum_{t \notin B} \varphi_1(t)^2 \leq (s-1) \sum_{t \notin B} \varphi_1(t) = (s-1)(q-s^2),$$

a contradiction to (4). (Observe that to reach a contradiction we need that $q - s^2$ is strictly positive. In case of an even k it can happen that $q = s^2$ and the function φ_1 vanishes outside B.)

This t provides us with a set D with $r \geq (s+3)/2$, and on substituting this into (1) we obtain the second case of the theorem. □

Remark 2.3. An alternative proof for the case $q = p$ and s being odd is as follows. Assume by contradiction that φ_1 is even-valued and nonnegative. Then by (3) it must be 0 for at least

$$q - |B| - \frac{q - s^2}{2} = \frac{q + s^2 - 2s}{2}$$

values of t. Let $\tilde{\chi}, \tilde{\varphi}, \tilde{\varphi}_1$ denote the images of χ, φ, φ_1 in $\mathbb{F}_q$ (i.e., the functions are evaluated $\bmod p$). By the previous observation $\tilde{\varphi}_1$ has at least $(q + s^2 - 2s)/2$ zeroes. On the other hand, we have $\tilde{\chi}(x) = x^{\frac{q-1}{2}}$, and hence $\tilde{\varphi}_1$ is a polynomial of degree $(q-1)/2$; its leading coefficient is $s = [\sqrt{q}] \neq 0 \bmod p$ (This last fact may fail if $q = p^k$, even if k is odd. Therefore this proof is restricted in its generality. Nevertheless we include it here, because we believe that it has the potential to lead to stronger results if $q = p$.) Consequently $\tilde{\varphi}_1$ can have at most $(q-1)/2$ zeros, a contradiction. In the case of even k we can have $s = \sqrt{q} \equiv 0 \pmod p$ and so the polynomial $\tilde{\varphi}_1$ can vanish, as it indeed does when B is a subfield.

Remark 2.4. It is clear from (1) that any improved lower bound on r will lead to an improved upper bound on s. If one thinks of elements of $\mathbb{Z}_p$ as being quadratic residues randomly with probability 1/2, then we expect that $r \geq \frac{s}{2} + c\sqrt{s}$. This would lead to an estimate $s \leq \sqrt{p} - cp^{1/4}$. This seems to be the limit of this method. In order to get an improved lower bound on r one can try to prove non-trivial upper bounds on the third moment $\sum_{t \in \mathbb{Z}_p} \varphi^3(t)$. To do this, we would need that the distribution of numbers $\frac{b_1 - b_2}{b_1 - b_3}$ is approximately uniform on Q as b_1, b_2, b_3 ranges over B. This is plausible because if $s \approx \sqrt{p}$ then the distribution of $B - B$

must be close to uniform on NQ. However, we could not prove anything rigorous in this direction.

Remark 2.5. Theorem 2.1 gives the bound $s \leq [\sqrt{p}] - 1$ for about three quarters of the primes $p = 4k + 1$. Indeed, part *(ii)* gives this bound for almost all p such that $n = [\sqrt{p}]$ is odd, with the only exception when $p = (n+1)^2 - 3$. Part *(i)* gives the improved bound $s \leq n - 1$ if $n^2 + n - 1 > p$. This happens for about half of the primes $p = 4k + 1$ for which n is even. To make these statements rigorous we note that $\sqrt{p}/2$ is uniformly distributed modulo one, when p ranges over primes of the form $p = 4k + 1$: this is a special case of a result of Balog, [1, Theorem 1].

Acknowledgment The authors are grateful to Péter Csikvári for insightful comments regarding the prime-power case.

References

[1] A. Balog, *On the distribution of p^θ mod 1*, Acta Math. Hungar. 45 (1985), no. 1-2, 179-199.

[2] A. Blokhuis, *On subsets of $GF(q^2)$ with square differences*, Indag. Math., vol. 87, no. 4, pp. 369-372, (1984).

[3] B. Bollobás, Random Graphs, (second ed.), Cambridge University Press, Cambridge, 2001.

[4] P. J. Cameron, *Automorphism groups in graphs*, in: R. J. Wilson, L. W. Beineke (Eds.), Selected Topics in Graph Theory, vol. 2, Academic Press, NewYork, (1983), pp. 89-127.

[5] F. R. K. Chung, R. L. Graham, R. M. Wilson, *Quasi-random graphs*, Combinatorica, Volume 9, Issue 4, (1989), pp 345-362.

[6] S. D. Cohen, *Clique numbers of Paley graphs*, Quaest. Math. **11**, (2) (1988), 225-231.

[7] E. Croot, V. Lev, *Open problems in additive combinatorics*, Additive combinatorics CRM Proc. Lecture Notes Amer. Math. Soc., Providence, RI, 43, (2007), 207-233.

[8] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Res. Rep. Suppl. 10 (1973).

[9] S. Graham, C. Ringrose, *Lower bounds for least quadratic non-residues*, Analytic Number Theory (Allterton Park, IL, 1989), 269-309.

[10] M. Krivelevich, B. Sudakov, *Pseudo-random graphs*, in: More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer, (2006), 199-262.

[11] J. H. van Lint, R. M. Wilson, *A Course in Combinatorics*, Cambridge University Press, Cambridge, 1992 (2nd edition in 2001).

[12] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, 1979 (2nd edition in 1993).

[13] E. Maistrelli, D. B. Penman, *Some colouring problems for Paley graphs*, Discrete Math. **306** (2006) 99-106.

[14] M. Matolcsi, I. Z. Ruzsa, *Difference sets and positive exponential sums I. General properties*, J. Fourier Anal. Appl., to appear.

[15] Web-page of Geoffrey Exoo with clique numbers of Paley graphs for $7000 < p < 10000$, http://ginger.indstate.edu/ge/PALEY/

[16] Web-page of J. B. Shearer with clique numbers of Paley graphs for $p < 7000$, http://www.research.ibm.com/people/s/shearer/indpal.html

[17] Web-page discussion of clique numbers and plot of the function $s(p)$ for $p < 10000$, http://mathoverflow.net/questions/48591/cliques-paley-graphs-and-quadratic-residues