SUMS OF DILATES IN GROUPS OF PRIME ORDER

ALAIN PLAGNE

À la mémoire du mathématicien du désert, Yahya ould Hamidoune, en témoignage d'admiration

Abstract. We obtain a first non-trivial estimate for the sum of dilates problem in the case of groups of prime order, by showing that if t is an integer different from $0, 1$ or -1 and if $\mathcal{A} \subset \mathbb{Z}/p\mathbb{Z}$ is not too large (with respect to p), then $|\mathcal{A}+t\cdot\mathcal{A}| > (2+\vartheta_t)|\mathcal{A}|-w(t)$ for some constant $w(t)$ depending only on t and for some explicit real number $\vartheta_t > 0$ (except in the case $|t| = 3$). In the important case $|t| = 2$, we may for instance take $\vartheta_2 = 0.08$.

1. Introduction

It is well known that for two finite sets of integers $\mathcal{A}$ and $\mathcal{B}$, their sumset has a cardinality at least

$$|\mathcal{A}+\mathcal{B}| \geq |\mathcal{A}| + |\mathcal{B}| - 1 \tag{1}$$

and that, except if $\mathcal{A}$ or $\mathcal{B}$ has only one element, the equality case in (1) implies that $\mathcal{A}$ and $\mathcal{B}$ are arithmetic progressions with the same difference; in particular, if $|\mathcal{A}| = |\mathcal{B}|$, this implies that $\mathcal{A} = \mathcal{B}$ (up to a translation). Freiman's theory [6] then explains that the farther we are from such a situation ($\mathcal{A}$ and $\mathcal{B}$ close to a given arithmetic progression), the bigger the sumset $(\mathcal{A}+\mathcal{B})$ has to be.

The problem of investigating sumsets of dilates takes place in this context. We are considering a simplified (one variable) problem where we are given a set $\mathcal{A}$ and where, for $\mathcal{B}$, we take a dilation of $\mathcal{A}$. In some sense such a set $\mathcal{B}$ is close to $\mathcal{A}$ since it is constructed from $\mathcal{A}$, but it is also not that close since a dilation for instance does not keep the difference of an arithmetic progression unchanged. The problem is then to understand what is going on precisely in this situation, by which we mean to measure how far from the trivial lower bound (that is, $2|\mathcal{A}| - 1$) we are in (1).

Looking for such an improved bound, in the framework of sets of integers, goes back at least to the beginning of the 2000's. In [7], when dealing with the $3k-3$ theorem of Freiman, the simple estimate

$$|\mathcal{A}+t\cdot\mathcal{A}| \geq 3|\mathcal{A}| - 2, \tag{2}$$

for any integer $t \neq -1, 0, 1$, was already (and incidentally) proved. Then, Nathanson [12] refined this estimate to

$$|\mathcal{A}+t\cdot\mathcal{A}| \geq \left\lceil \frac{7}{2}|\mathcal{A}| - \frac{5}{2} \right\rceil,$$

as soon as $|t| \geq 3$. The case $t = 3$ was finally completely solved [1, 4] (the authors of [4] even describe all the cases where equality holds) :

$$|\mathcal{A}+3\cdot\mathcal{A}| \geq 4|\mathcal{A}| - 4.$$

In [1], Bukh generalizes this result to an arbitrary number of summands and proves the nice general lower bound

$$|t_1 \cdot \mathcal{A} + \cdots + t_k \cdot \mathcal{A}| \geq (|t_1| + \cdots + |t_k|)|\mathcal{A}| - o(|\mathcal{A}|) \tag{3}$$

as soon as $\gcd(t_1, \ldots, t_k) = 1$, an assumption which is not restrictive since an affine transformation of $\mathcal{A}$ (by dilations and translations) does not change the cardinality of the sumset.

Returning to the case of two summands which is probably the only one where it is reasonable to investigate for precise bounds, in [3], the authors obtained

$$|\mathcal{A} + t \cdot \mathcal{A}| \geq (1+t)|\mathcal{A}| - \left\lceil \frac{t(t+2)}{4} \right\rceil,$$

if t is prime and $|\mathcal{A}|$ large enough compared to t. This was extended to the case where t is either a prime power or the product of two primes in [5]. We shall use these results under the form of a universal lower bound (that is, valid for any $\mathcal{A} \neq \emptyset$)

$$|\mathcal{A} + t \cdot \mathcal{A}| \geq (1+t)|\mathcal{A}| - w(t) \tag{4}$$

for some constant $w(t)$. We add that we may take $w(2) = 2, w(3) = 4$ and $w(4) = 10$ for instance (using (2) and [3, 5]).

The case where the first coefficient is equal to 2 attracted also some energy and the authors of [8] could prove that, if t is an odd prime,

$$|2 \cdot \mathcal{A} + t \cdot \mathcal{A}| \geq (2+t)|\mathcal{A}| - t^2 - t + 2,$$

if $|\mathcal{A}|$ is large enough compared to t. This was very recently extended to the case when t is a prime-power in [11].

In this paper we investigate dilates in $\mathbb{Z}/p\mathbb{Z}$ (p a prime) – a subject which, seemingly, was not yet investigated – and show that a similar phenomenon happens.

2. The case of $\mathbb{Z}/p\mathbb{Z}$

While inequality (1) in the integers is immediate, the Cauchy-Davenport theorem [2] itself is more sophisticated. In this note, similarly, we are interested in a counterpart to the results for integers that were mentioned in the Introduction, in the framework of cyclic groups of prime order. If one is optimistic, one can hope for a result similar to Bukh's lower bound (3), even with the $o(|\mathcal{A}|)$ term replaced by a constant depending only on t. For the sake of simplicity, and to avoid too many technicalities, we shall restrict ourselves to the case of two summands.

Conjecture 1. *Let p be a prime and t be an integer different from 0, 1 or -1. Then there exists a constant $c(t)$ such that for any set of integers $\mathcal{A}$, the following estimate holds*

$$|\mathcal{A} + t \cdot \mathcal{A}| \geq \min((t+1)|\mathcal{A}| - c(t), p).$$

Clearly the restriction that only one of the dilation coefficients is not equal to 1 is not restrictive since the cardinality of the sumset is invariant by invertible dilation in $\mathbb{Z}/p\mathbb{Z}$ so that $|\alpha \cdot \mathcal{A} + \beta \cdot \mathcal{A}| = |\mathcal{A} + (\alpha^{-1}\beta) \cdot \mathcal{A}|$ when α is non-zero modulo p. In the same way, it would not be restrictive to impose $|t| < p/2$ (in fact, in this form, the conjecture is empty if t is not uniformly bounded with respect to p).

Notice that, if such a conjecture is true, it implies the result in the integers by a cyclification argument (from an additive point of view, any set of integers can be seen as a set of elements of a cyclic group of large enough prime order). Therefore the proof of such a result has to contain the proof in the case of integers.

In this paper, we shall not be able to prove Conjecture 1 but we will rather make a step towards it. To state our result we need to introduce a family of auxiliary functions f_t defined as follows (t is an integral parameter satisfying $t \geq 2$).

In the case $t = 2$, we first introduce

$$c_2^{(0)} = \frac{1 - 2^{3/2}/3}{2} = 0.028595\ldots$$

and then define, for any $0 \leq c \leq 1$, the function f_2 as

$$f_2(c) = \begin{cases} \text{the unique solution } x \geq 2 \text{ to the equation } 3(1-cx) = x^{3/2}, & \\ & \text{if } 0 \leq c \leq c_2^{(0)}, \\ 2, \quad \text{otherwise.} & \end{cases}$$

This is a decreasing function which assumes the extremal values $f_2(0) = 3^{2/3} = 2.080083\ldots$ and $f_2(c_2^{(0)}) = 2$.

In the case $t \geq 3$, we define

$$c_t^{(0)} = \frac{1}{2}\left(1 - \frac{2^{3/2}}{(|t|+1)\sin\left(\frac{\pi}{|t|+1}\right)}\right).$$

Notice that $c_t^{(0)} > 0$ (except for $t = 3$ where $c_3^{(0)} = 0$) and then define, for any $0 \leq c \leq 1$, the function f_t as

$$f_t(c) = \begin{cases} \text{the unique solution } x \geq 2 \text{ to the equation} & \\ (|t|+1)\sin\left(\frac{\pi}{|t|+1}\right)(1-cx) = x^{3/2}\sin(\frac{\pi}{x}), \quad \text{if } 0 \leq c \leq c_t^{(0)}, \\ 2, \quad \text{otherwise.} & \end{cases}$$

All the f_t's functions are again decreasing functions.

We underline the fact that when t tends to infinity, the quantity $c_t^{(0)}$ tends towards $(1/2 - \sqrt{2}/\pi) = 0.04984\ldots$ while $f_t(0)$ tends towards the unique solution of

$$x^{3/2}\sin\left(\frac{\pi}{x}\right) = \pi,$$

that is $2.15409\ldots$ These two numerical values are of a certain significance explained below.

We are now ready to formulate our main result, which is much more modest than Conjecture 1 but at least supports this statement (except for $|t| = 3$).

Theorem 1. *Let p be a prime and $\mathcal{A}$ be a non-empty subset of $\mathbb{Z}/p\mathbb{Z}$. If t is a prime-power or a product of two primes, different from $0, 1$ or -1, then*

$$|\mathcal{A} + t\cdot\mathcal{A}| \geq \min\left(f_{|t|}(c)|\mathcal{A}| - w(t), p\right),$$

where $c = |\mathcal{A}|/p$, and $w(t)$ the constant depending only on t defined in (4).

Notice that if we do not impose to t to be a prime-power or a product of two primes, then we can still obtain a non-trivial result of the form obtained by Bukh.

Theorem 2. *Let $\varepsilon > 0$. There is an integer p_0 depending only on ε such that if $p \geq p_0$ is a prime, $\mathcal{A}$ a non-empty subset of $\mathbb{Z}/p\mathbb{Z}$ and t is an integer different from $0, 1$ or -1, then*

$$|\mathcal{A} + t\cdot\mathcal{A}| \geq \min((f_{|t|}(c) - \varepsilon)|\mathcal{A}|, p),$$

where $c = |\mathcal{A}|/p$.

The proof is the same, mutatis mutandis (that is, essentially replacing (4) by (3)), as the one of Theorem 1.

As an important special case of Theorem 1, we obtain for example the following corollary.

Corollary 1. *Let p be a prime and $\mathcal{A}$ be a non-empty subset of $\mathbb{Z}/p\mathbb{Z}$ such that $|\mathcal{A}| < p/35000$. Let σ be equal to either 1 or -1. Then, we have*

$$|\mathcal{A}+2\sigma\cdot\mathcal{A}|\geq(2+\vartheta_2)|\mathcal{A}|-2,$$

with $\vartheta_2=0.08$.

Notice that an equivalent form of the corollary is

$$|\mathcal{A}+2\cdot\mathcal{A}|\geq\min((2+\vartheta_2)|\mathcal{A}|-2,p/17500-1)$$

In view of our earlier computation, for large values of t we would get a $\vartheta_t\sim 0.15409\ldots$ as t tends to infinity. It is therefore a limitation of the method that we cannot hope any lower bound coefficient better than 2.16, say.

This result is clearly demanding for improvements. It should be possible to extend it to $|t|=3$, with a larger value for ϑ_t and larger sets $\mathcal{A}$, that is, with a larger ratio $|\mathcal{A}|/p$.

The proof relies on a so-called rectification argument, first used in Freiman's original proof of his famous theorem [6], and now standard. It also uses a result of Lev [9] and its improvement [10]. This proof is presented in the following three sections: in Section 3 we start with the exponential sums argument, then we proceed with the application of Lev's results in Section 4 and finally, in Section 5, we conclude using the rectification trick.

3. Proof of Theorem 1: the exponential sums method

Let $\mathcal{A}\subset\mathbb{Z}/p\mathbb{Z}$ and define $c=|\mathcal{A}|/p$. We define $\mathcal{S}=\mathcal{A}+t\cdot\mathcal{A}$ and define the real number x by writing

$$|\mathcal{S}|=x|\mathcal{A}|-w(t).$$

If $c\geq c_t^{(0)}$ then we simply apply the Cauchy-Davenport theorem and get

$$|\mathcal{S}|\geq\min(2|\mathcal{A}|-1,p)\geq\min(f_t(c)|\mathcal{A}|-w(t),p)$$

and we are done. From now on and until the end of the proof, assume $c\leq c_t^{(0)}$. In this case, we have $f_t(c)|\mathcal{A}|-w(t)<p$ so that we are led to prove $|\mathcal{S}|\geq f_t(c)|\mathcal{A}|-w(t)$.

If $1_{\mathcal{U}}$ denotes the characteristic function of a subset $\mathcal{U}$ of $\mathbb{Z}/p\mathbb{Z}$, we write

$$\widehat{1_{\mathcal{U}}}(x)=\sum_{u\in\mathcal{U}}\exp(2\pi iux/p),$$

for the discrete Fourier transform of $1_{\mathcal{U}}$ and use the exponential sums counting method, which yields

$$p|\mathcal{A}|^2=\sum_{r=0}^{p-1}\widehat{1_{\mathcal{A}}}(r)\widehat{1_{t\cdot\mathcal{A}}}(r)\overline{\widehat{1_{\mathcal{S}}(r)}}.$$

Then, we proceed with the classical calculations (using Cauchy-Schwartz inequality, Parseval identity,...):

$$\begin{aligned}
p|\mathcal{A}|^2 &= |\mathcal{A}|^2|\mathcal{S}| + \sum_{r=1}^{p-1} \widehat{1_{\mathcal{A}}}(r)\widehat{1_{t\cdot\mathcal{A}}}(r)\overline{\widehat{1_{\mathcal{S}}}(r)} \\
&\leq |\mathcal{A}|^2|\mathcal{S}| + \max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)| \sum_{r=1}^{p-1} |\widehat{1_{t\cdot\mathcal{A}}}(r)||\widehat{1_{\mathcal{S}}}(r)| \\
&\leq |\mathcal{A}|^2|\mathcal{S}| + \max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)| \left(\sum_{r=1}^{p-1} |\widehat{1_{t\cdot\mathcal{A}}}(r)|^2\right)^{1/2} \left(\sum_{r=1}^{p-1} |\widehat{1_{\mathcal{S}}}(r)|^2\right)^{1/2} \\
&\leq |\mathcal{A}|^2|\mathcal{S}| + \max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)| \left(\sum_{r=0}^{p-1} |\widehat{1_{\mathcal{A}}}(tr)|^2\right)^{1/2} \left(\sum_{r=0}^{p-1} |\widehat{1_{\mathcal{S}}}(r)|^2\right)^{1/2} \\
&\leq |\mathcal{A}|^2|\mathcal{S}| + \max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)| \sqrt{p|\mathcal{A}|}\sqrt{p|\mathcal{S}|} \\
&= x|\mathcal{A}|^3 + p\sqrt{|\mathcal{A}||\mathcal{S}|} \max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)|
\end{aligned}$$

from which it follows the following lower bound for the Fourier bias of $\mathcal{A}$,

$$\max_{1\leq r\leq p-1} |\widehat{1_{\mathcal{A}}}(r)| \geq \frac{p - x|\mathcal{A}|}{p\sqrt{x}}|\mathcal{A}| = \frac{1 - xc}{\sqrt{x}}|\mathcal{A}|.$$

Again, since the cardinality of a set is independent of the invertible dilation in $\mathbb{Z}/p\mathbb{Z}$, we may assume without loss of generality that this maximum is attained for $r = 1$.

4. Proof of Theorem 1 continued: a lemma by Lev

We now use (a special case of) Theorem 1 of [9] and Corollary 1 of [10] which we state here as a single unified lemma for our purpose. We shall make use of the function defined on $(0, \pi]$

$$g(u) = \frac{\sin u}{u}$$

and of its reciprocal function g^{-1} which is decreasing on $[0, 1)$.

Lemma 1. *With the preceding notation, define*

$$\eta = \frac{\left|\widehat{1_{\mathcal{A}}}(1)\right|}{|\mathcal{A}|}.$$

Then, for any $\beta \in (0, 1/2]$, there exists an interval I of length βp in $\mathbb{Z}/p\mathbb{Z}$ such that

$$|\mathcal{A} \cap I| \geq M(\beta, \eta)|\mathcal{A}|$$

where

$$M(\beta, \eta) = \max\left(\frac{\eta + 1 - 2\cos\pi\beta}{2(1 - \cos\pi\beta)}, \frac{\pi\beta}{g^{-1}(\eta g(\pi\beta))}\right).$$

In view of what will be needed later in the proof, we now state an important remark.

Remark 1. *Assume that $0 \leq \eta \leq 1/\sqrt{2}$ and $0 \leq \beta \leq 1/3$. If*

$$\beta^{-1}\left(\frac{\eta + 1 - 2\cos\pi\beta}{2(1 - \cos\pi\beta)}\right) \geq 2$$

then $\beta \geq 1/4$. *If*

$$\frac{\pi}{g^{-1}(\eta\ g(\pi\beta))} \geq 2$$

then $\beta \leq 1/4$.

Indeed the first inequation leads to

$$\cos \pi\beta \leq \frac{1+\eta-4\beta}{2(1-2\beta)} \leq \frac{1+1/\sqrt{2}-4\beta}{2(1-2\beta)},$$

which is easily shown to imply $\beta \geq 1/4$.

For the second inequation, applying g, it implies

$$\eta g(\pi\beta) \geq g\left(\frac{\pi}{2}\right) = \frac{2}{\pi},$$

and therefore (again due to $\eta \leq 1/\sqrt{2}$) that

$$g(\pi\beta) \geq \frac{2\sqrt{2}}{\pi} = g\left(\frac{\pi}{4}\right)$$

and we obtain $\beta \leq 1/4$.

5. Proof of Theorem 1 concluded: the rectification

We apply Lev's lemma with $\beta = 1/(|t|+1)$ and the value we just obtained for η, namely $\eta = (1-xc)/\sqrt{x}$. We obtain an interval I of length $[p/(|t|+1)] \leq (p-1)/(|t|+1)$ such that $\mathcal{A}_0 = \mathcal{A} \cap I$ contains at least

$$|\mathcal{A}_0| \geq B_t(x,c)|\mathcal{A}| \tag{5}$$

elements, where we denote by $B_t(x,c)$ the lower bound

$$B_t(x,c) = M(\beta,\eta) = M\left(\frac{1}{|t|+1}, \frac{1-xc}{\sqrt{x}}\right).$$

The point is now to notice that the sumset $\mathcal{A}_0 + t\cdot\mathcal{A}_0$ is Freiman isomorphic to the same set seen in the integers. Indeed all the elements in the sumset belong to an interval of integers of length $(1+|t|)[p/(|t|+1)] \leq p-1 < p$, and therefore two sums are equal if and only if they are equal modulo p. Thus, we can apply to the sumset $\mathcal{A}_0$ the lower bound derived in the case of integers (4) and get, in view of (5),

$$\begin{aligned} x|\mathcal{A}| - w(t) = |\mathcal{S}| = |\mathcal{A} + t\cdot\mathcal{A}| &\geq |\mathcal{A}_0 + t\cdot\mathcal{A}_0| \\ &\geq (|t|+1)|\mathcal{A}_0| - w(t) \\ &\geq (|t|+1)B_t(x,c)|\mathcal{A}| - w(t). \end{aligned}$$

We finally obtain

$$x \geq (|t|+1)B_t(x,c). \tag{6}$$

In order to know which bound is worth here (for the maximum), we use the remark stated just after Lev's lemma.

In the case $|t| = 2$, we use the first argument in the maximum defining M. It then follows from (6)

$$x \geq 3B_2(x,c) = 3\left(\frac{1-xc}{\sqrt{x}}\right)$$

or equivalently $x \geq f_2(c)$ by definition of the function f_2 and the result is proved.

As for proving Corollary 1, we only have to solve $f_2(c) = 2.08$ to get

$$c = \frac{1 - 2.08^{3/2}/3}{2.08} = 0.0000209607\cdots = \frac{1}{34410.7\ldots} > \frac{1}{35000}.$$

In the case $|t| \geq 3$, we use the second argument in the maximum defining M, which gives similarly

$$x \geq \frac{\pi}{g^{-1}(\eta\ g(\pi/(|t|+1)))}.$$

Applying g, this can be rewritten as

$$\eta\ g\left(\frac{\pi}{|t|+1}\right) \leq g\left(\frac{\pi}{x}\right),$$

that is

$$(|t|+1)\sin\left(\frac{\pi}{|t|+1}\right)(1-xc) \leq x^{3/2}\sin\left(\frac{\pi}{x}\right).$$

And by definition, this implies

$$x \geq f_t(c)$$

and the result follows.

Acknowledgements: This note was started in Paris in January 2011 after a long discussion with Yahya ould Hamidoune on the case of integers. The author had just the time to announce him the (quantitative version of the) result before he stopped any mathematical activity. The article was then mainly written down while the author was enjoying an invitation to the Isaac Newton Institute for Mathematical Sciences in Cambridge at the occasion of the *Discrete Analysis* Programme, February 2011. He thanks the organizers for this kind invitation and the good working conditions they offered.

References

[1] B. Bukh, *Sums of dilates*, Combin. Probab. Comput. 17 (2008), no. 5, 627–639.
[2] A.-L. Cauchy, *Recherches sur les nombres*, J. École polytechnique 9 (1813), 99–123.
[3] J. Cilleruelo, Y. ould Hamidoune, O. Serra, *On sums of dilates*, Combin. Probab. Comput. 18 (2009), 871–880.
[4] J. Cilleruelo, M. Silva, C. Vinuesa, *A sumset problem*, J. Combin. Number Th. 2 (2010).
[5] S.-S. Du, H.-Q. Cao, Z.-W. Sun, *On a sumset problem for the integers*, arXiv:1011.5438.
[6] G. A. Freiman, *Foundations of a structural theory of set addition*, Trans. AMS Monographs **37**, AMS, 1973.
[7] Y. ould Hamidoune, A. Plagne, *A generalization of Freiman's* $3k-3$ *theorem*, Acta Arith. 103 (2002), no. 2, 147–156.
[8] Y. ould Hamidoune, J. Rué, *On dilates sums*, Combin. Probab. Comput. 20 (2011), 249–256.
[9] V.F. Lev, *Distribution of points on arcs*, Integers 5 (2005), 8pp.
[10] V.F. Lev, *More on points and arcs*, Combinatorial number theory, 347–350, de Gruyter, Berlin, 2007.
[11] Z. Ljujic, *A lower bound for the size of a sum of dilates*, arXiv 11.01.5425.
[12] M. Nathanson, *Inverse problems for linear forms over finite sets of integers*, J. Ramanujan Math. Soc. 23 (2008), no. 2, 151–165.

E-mail address: plagne@math.polytechnique.fr

Centre de Mathématiques Laurent Schwartz École polytechnique 91128 Palaiseau Cedex France