A REMARK ON SUMSETS IN $\mathbb{F}_q^n$

JORDAN S. ELLENBERG

In this note, we explain how the method of [CLP16] and [EG16] can be extended to show that the sumset $S+T$ of two large subsets S and T of $\mathbb{F}_q^n$ can be expressed "more efficiently" as a union of sumsets of smaller subsets.

Write $M(\mathbb{F}_q^n)$ be the upper bound proved in [EG16] for the size of a subset of $\mathbb{F}_q^n$ with no three-term arithmetic progressions; that is, $M(\mathbb{F}_q^n)$ is the number of monomials in $x_1, \dots, x_n$ with degree at most $(q-1)$ in each variable and total degree at most $(q-1)n/3$. For each q, the bound $M(\mathbb{F}_q^n)$ is bounded above by c^n for some $c < q$.

Theorem 1. *Let $\mathbb{F}_q$ be a finite field and let S, T be subsets of $\mathbb{F}_q^n$. Then there is a subset S' of S and a subset T' of T such that*

- $|S'| + |T'| \leq M(\mathbb{F}_q^n)$;
- $S + T = (S' + T) \cup (S + T')$.

Corollary 2. *Let S be a subset of $\mathbb{F}_q^n$. Then S has a subset S' of size at most $M(\mathbb{F}_q^n)$ such that $S' + S = S + S$.*

Proof. By Theorem 1 there are subsets S_1 and S_2 of S such that $S + S = (S_1 + S) \cup (S + S_2)$ and $|S_1| + |S_2| \leq M(\mathbb{F}_q^n)$. Taking S' to be $S_1 \cup S_2$ we are done. □

The bound proved in [EG16] on subsets of $\mathbb{F}_q^n$ with no three terms in arithmetic progression is also an immediate consequence.

Corollary 3. *A subset S of $\mathbb{F}_q^n$ containing no three-term arithmetic progression has size at most $M(\mathbb{F}_q^n)$.*

Proof. If S has no 3-term arithmetic progression, then $S' + S$ is strictly smaller than $S + S$ for every *proper* subset $S' \subset S$ (because $S' + S$ fails to contain $2s$ if s lies in the complement of S'.) Thus, the subset S' guaranteed by Corollary 2 must be equal to S, whence $|S| = |S'| \leq M(\mathbb{F}_q^n)$. □

We now prove Theorem 1.

Proof. Let V be the space of polynomials in $\mathbb{F}_q[x_1, \dots, x_n]$ with degree at most $(q-1)$ in each variable and total degree at most d, which vanish on the complement of $S + T$. Then $\dim V$ is at least $m_d - q^n + |S + T|$. Write $\mathcal{M}$ for the space of linear functions from the $\mathbb{F}_q$-vector space with basis S to the $\mathbb{F}_q$-vector space with basis T; this can be identified, if we wish, with the space of $|S| \times |T|$ matrices.

For each $P \in V$ we may consider $M(P) \in \mathcal{M}$ whose entries are $P(s+t)_{s \in S, t \in T}$. By the argument of the Croot-Lev-Pach lemma [CLP16] this matrix has rank at most $2m_{d/2}$.

Date: 28 May 2016.

Note that M is an homomorphism from V to $\mathcal{M}$, which is injective: if P lies in the kernel, it vanishes at $S+T$, but P vanishes on the complement of $S+T$ by hypothesis, so P vanishes on every point of $\mathbb{F}_q^n$ and is 0.

We thus can, and do, think of V as a vector subspace of $\mathcal{M}$ of dimension at least $m_d - q^n + |S+T|$, each of whose members has rank at most $2m_{d/2}$.

In order to derive the desired conclusion, we use a theorem of Meshulam [Mes85], which gives lower bounds for the maximum rank attained in a linear space of matrices. Choose an ordering on S and an ordering on T. These choices endow the entries of a matrix in $\mathcal{M}$ with a lexicographic order. If $A \in \mathcal{M}$ is a matrix, we denote by $p(A) \in S \times T$ the location of the lexicographically first nonzero entry of A.

We note that $p(M(P))$ cannot be an arbitrary element of $S \times T$, since $M(P)$ has equal entries at (s,t) and (s',t') whenever $s+t = s'+t'$. In particular, for each $u \in \mathbb{F}_q^n$, write $\phi(u)$ for the lexicographically first (s,t) such that $s+t=u$; then, for each $P \in V$, we have $p(M(P)) = \phi(u)$ for some $u \in \mathbb{F}_q^n$.

By Gaussian elimination, we can find a basis $A_1, \ldots, A_{\dim V}$ for V such that $p(A_1), \ldots, p(A_{\dim V})$ are distinct.

We now apply Meshulam's theorem [Mes85, Theorem 1], which shows that there is a set of $2m_{d/2}$ lines (a line being a row or a column) whose union contains $p(A_i)$ for all i.

This set of lines consists of a subset of S, which we call S_0, and a subset of T, which we call T_0, satisfying $|S_0| + |T_0| = 2m_{d/2}$.

We now have, for $i = 1, \ldots, \dim V$,

$$p(A_i) = (s_i, t_i)$$

with either $s_i \in S_0$ or $t_i \in T_0$. What's more, $s_i + t_i$ and $s_j + t_j$ are distinct whenever i and j are. So the union of $S_0 + T$ with $S + T_0$ contains at least $\dim V$ elements of $S+T$.

Since $\dim V \geq m_d - q_n + |S+T|$, the set U of elements of $S+T$ *not* contained in $(S_0+T) \cup (S+T_0)$ has cardinality at most $q_n - m_d$. Let S_1 be a subset of S of size $q_n - m_d$ such that each $u \in U$ is represented as $s+t$ for some $s \in S_1$. Then taking $S' = S_0 \cup S_1$ and $T' = T_0$, we have that $S' + T \cup S + T' = S + T$; moreover,

$$|S'| + |T'| \leq 2m_{d/2} + q_n - m_d$$

and minimizing over d we get the desired result. □

Question 4. To what extent is the bound on $|S'| + |T'|$ in Theorem 1 sharp?

Question 5. Does Theorem 1 have any nontrivial analogue in cyclic groups? For instance, using the symmetric formulation of Corollary 2 to simplify the statement, is there $f(N) < N$ such that, given $S \in \mathbb{Z}/N\mathbb{Z}$, there is always a subset $S' \subset S$ of size $f(N)$ with $S' + S = S + S$? Behrend's example of a large subset of $\mathbb{Z}/N\mathbb{Z}$ with no three-term arithmetic progressions shows that $f(N)$ would have to be at least $N^{1-\epsilon}$.

References

[CLP16] E. Croot, V. Lev, and P. P. Pach, *Progression-free sets in $\mathbf{Z}_4^n$ are exponentially small* (2016). arXiv preprint 1605.01506.

[EG16] J. S. Ellenberg and D. Gijswijt, *On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression* (2016). preprint.

[Mes85] R. Meshulam, *On the maximal rank in a subspace of matrices*, The Quarterly Journal of Mathematics **36** (1985), no. 2, 225–229.