Ruzsa’s good modelling lemma

Ernie Croot

July 23, 2012

1 Introduction

I thought I would give here an intuitive discussion of a certain lemma of Ruzsa which played a central role in the proof of Freiman’s Theorem. This lemma stated that:

Lemma. Suppose that $A$ is a finite set of integers. Then, for any prime $N > 2|kA - kA|$, there exists a subset $A' \subseteq A$ of size at least $|A|/k$ which is Freiman $k$-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$.

Note that if $A$ has “small doubling” then this subset of $\mathbb{Z}/N\mathbb{Z}$ will be “large” – it can be bounded from below in terms of $C$ and in terms of $k$, provided $N$ is chosen to be close to that lower bound $2|kA - kA|$. To see this, let us suppose that $N$ were at most $4|kA - kA|$ (we are on safe ground here because by Bertrand’s postulate there is always a prime between $x$ and $2x$), and suppose that $C = |A + A|/|A|$ is the doubling constant. Then, that subset of $\mathbb{Z}/N\mathbb{Z}$ has size $|A'| > |A|/k$, and therefore its density in $\mathbb{Z}/N\mathbb{Z}$ is at least

$$|A'|/N \geq |A|/4k|kA - kA| \geq 1/4kC^{2k},$$

where the last inequality follows from Ruzsa-Plunnecke-Petridis.
2 Proof of the lemma

The way I think of Ruzsa’s proof is that one can produce lots and lots of Freiman $k$-homomorphisms $\nu$ from “large” subsets $A' \subseteq A$ to subsets of $\mathbb{Z}/N\mathbb{Z}$, each parameterized by some integer $q$ that appears in intermediate steps of the proof, such that there are more choices for $q$ than there are potential obstructions that keep any of the $\nu$ from being a Freiman $k$-isomorphism. So, by a counting argument one discovers that there exists a $q$, and therefore a map $\nu$, which results in a Freiman $k$-isomorphism.

To prove Ruzsa’s lemma, we start by letting $p$ be any prime satisfying
\begin{equation}
  p > k(\text{MAX} A - \text{MIN} A). \tag{1}
\end{equation}

Then, for an integer $1 \leq q \leq p - 1$ (which is necessarily coprime to $p$) we consider the mapping
\[ \varphi_q : A \to \mathbb{Z}/p\mathbb{Z} \\
    a \to qa \pmod{p}. \]

It is obvious that this is a Freiman $k$-homomorphism for all $k$, since it is a group homomorphism (which are necessarily Freiman $k$-homomorphisms for all $k$); however, what takes a little bit of work to see (though not much) is that, in fact, inequality (1) implies that
\[ \varphi_q \] is a Freiman $k$ – isomorphism.

The trouble with working with the group $\mathbb{Z}/p\mathbb{Z}$ to prove Ruzsa’s lemma is that it is potentially too large (much larger than $2|kA - kA|$). So what we want to do is to compress the images of $\varphi_q$ in $\mathbb{Z}/p\mathbb{Z}$ somehow; and, Ruzsa’s idea was to map subsets of $\mathbb{Z}/p\mathbb{Z}$ down to subsets of $\mathbb{Z}/N\mathbb{Z}$, where $N$ is any prime satisfying
\[ N > 2|kA - kA|. \]

Note that this $N$ is potentially quite a bit smaller than $p$, which is good.

Given such an $N$ we are now faced with a problem, which is that if we let $\psi$ be any mapping from $\mathbb{Z}/p\mathbb{Z}$ down to $\mathbb{Z}/N\mathbb{Z}$, it cannot be an injective Freiman $k$-homomorphism, let alone an injective group homomorphism.
However, if we restrict ourselves to an integer interval $I$ of residues mod $p$ of width at most $p/k$, then on that interval we can pick $\psi$ to be a Freiman $k$-homomorphism. We have to be a little careful here in describing this, due to the fact that residues mod $p$ are not integers, so the mapping is tricky to define because of “type” issues: given $I$, choose a representation for the residues in $I$ so that we get consecutive integers, say $I = \{x, x + 1, x + 2, ..., x_n\} \subset \mathbb{Z}$ be the obvious inclusion mapping.

Using this mapping $\iota$ we can now define our mapping

$$
\psi_I : I \to J \subseteq \mathbb{Z}/N\mathbb{Z}
$$

$$
n \to \iota(n) \pmod{N}.
$$

It is straightforward to check that this is a Freiman $k$-isomorphism.

To each $1 \leq q \leq p - 1$ suppose we choose $I_q \subseteq \mathbb{Z}/p\mathbb{Z}$ to be any interval of width $\lfloor p/k \rfloor$ that contains the maximal number of elements of $\varphi_q(A)$. And then let $A'_q \subseteq A$ be those elements of $A$ that map to this interval $I_q$. Clearly we will have

$$
|A'_q| \geq |A|/k.
$$

To prove Ruzsa’s lemma, then, we just need to focus on the following claim.

**Claim.** There exists $1 \leq q \leq p - 1$, such that then the composition $\psi_{I_q} \circ \varphi_q$ is a Freiman $k$-isomorphism when this mapping is restricted to $A'_q$ (and the image is restricted to the appropriate subset of $\mathbb{Z}/N\mathbb{Z}$).

Let $\nu_q := \psi_{I_q} \circ \varphi_q|_{A'_q}$ be one of these restricted mappings. Note that regardless of what $q$ we pick, $\nu_q$ is *always* a Freiman $k$-homomorphism from $A'_q$ into $\mathbb{Z}/N\mathbb{Z}$; however, only special $q$ are “good”, meaning that they result in a $k$-isomorphism.

Now, if $q$ is “bad” then it means that there exist elements

$$
a_1, ..., a_k, a'_1, ..., a'_k \in A'_q,
$$

such that

$$
a_1 + \cdots + a_k \neq a'_1 + \cdots + a'_k,
$$
while

\[ \nu_q(a_1) + \cdots + \nu_q(a_k) \equiv \nu_q(a'_1) + \cdots + \nu_q(a'_k) \pmod{N}. \]

This last statement implies that

\[ \psi_{I_q}(b_1) + \cdots + \psi_{I_q}(b_k) \equiv \psi_{I_q}(b'_1) + \cdots + \psi_{I_q}(b'_k) \pmod{N}, \]

where \( b_i \equiv qa_i \pmod{p} \) and all \( b'_i \equiv qa'_i \pmod{p} \), where \( b_i, b'_i \in I_q \). Since we are already working mod \( N \) we can just remove these \( \psi_{I_q} \)'s and conclude that

\[ b_1 + \cdots + b_k \equiv b'_1 + \cdots + b'_k \pmod{N}. \]

So,

\[ b_1 + \cdots + b_k - b'_1 - \cdots - b'_k = Nm, \quad \text{where } 1 \leq m \leq p/N \]

(Without loss we can assume that this sum of \( b_i \)'s exceeds the sum of \( b'_i \)'s.)

Upon considering this last equation mod \( p \), and upon writing the \( b_i \) and \( b'_i \) back in terms of \( a_i \) and \( a'_i \), we find that it implies that

\[ (Nm)^{-1}(a_1 + \cdots + a_k - a'_1 - \cdots - a'_k) \equiv q^{-1} \pmod{p}. \]

Since this difference of sums of \( a_i \)'s and \( a'_i \)'s is contained in \( kA - kA \), and since there are at most \( p/N \) choices for \( m \) it follows that there can be at most \( (p/N)|kA - kA| \) “bad \( q \”). This number is smaller than \( p-1 \) if \( N > 2|kA - kA| \); and so, assuming \( N \) is this large there are more choices for \( q \) than there are “bad \( q \”). It follows that one of the \( \nu_q \)'s is a Freiman \( k \)-isomorphism out of \( A'_{q} \), thereby proving the lemma.