Additive Combinatorics Lecture 12

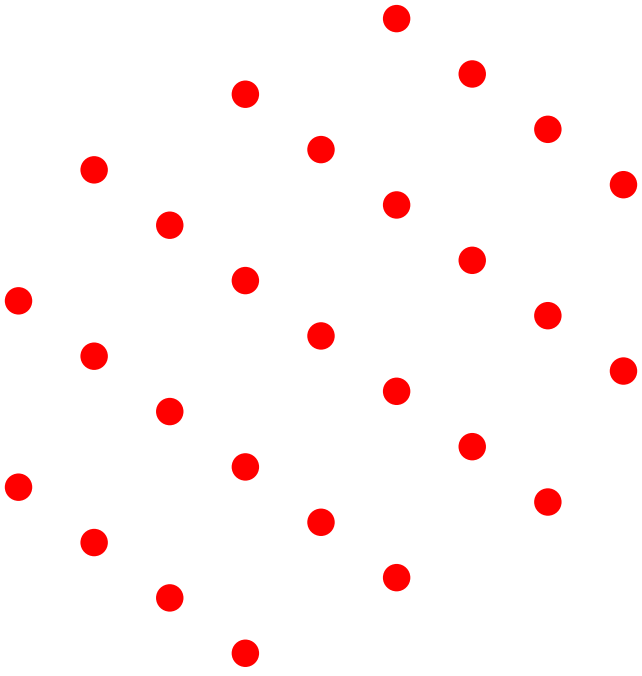
Leo Goldmakher
Scribe: Gal Gross

April 4th, 2014

Last lecture we proved the Bohr-to-gAP proposition, but the final step was a bit mysterious – we invoked Minkowski's second theorem without explanation. Today we will describe and motivate this theorem; this necessitates a brief discussion of the Geometry of Numbers. We will conclude the lecture by starting to develop the final tool necessary for the proof of Freiman-Ruzsa, the notion of Freiman homomorphisms and isomorphisms.

1 Lattices

We begin by discussing lattices in Euclidean space. A *lattice* is an infinite grid of regularly spaced points. For example, here is a two dimensional lattice:



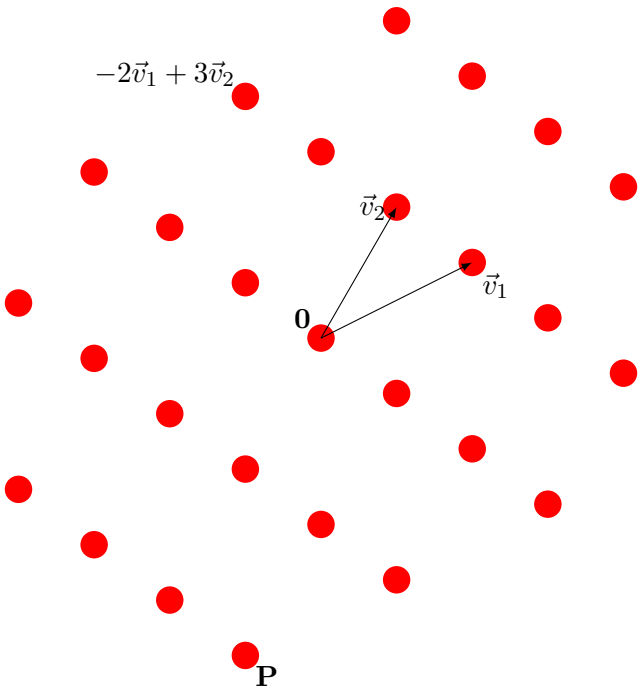
A two-dimensional lattice (the grid extends infinitely in all directions)

What about higher-dimensional lattices? Thinking about this quickly shows that we need a more precise definition of a lattice. Here's a convenient definition in terms of vectors: a d-dimensional lattice Λ is any subgroup of $\mathbb{R}^d$ of the form

$$\Lambda = \mathbb{Z}\vec{v}_1 + \mathbb{Z}\vec{v}_2 + \cdots + \mathbb{Z}\vec{v}_d,$$

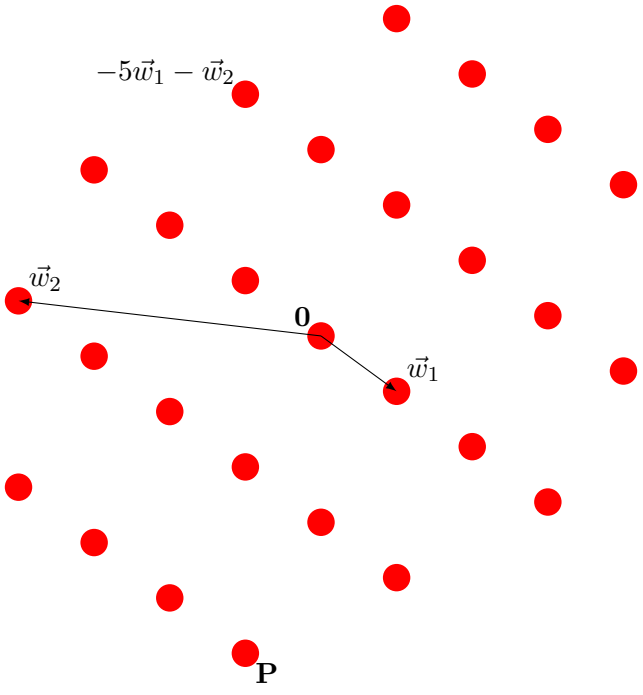
where $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_d$ are vectors in $\mathbb{R}^d$. (We will assume throughout that our lattices are *non-degenerate*, i.e. generated by a set of linearly-independent vectors.) We will sometimes use the notation $\Lambda = \langle \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d \rangle$,

and refer to the collection $\{\vec{v}_i\}$ as a *basis* of Λ. Here is a picture of the same lattice as before, this time with an explicit basis $\{\vec{v}_1, \vec{v}_2\}$. (Every point of the lattice can be written as a linear combination of the basis vectors; I've given one example of this below.)



The lattice $\Lambda = \langle \vec{v}_1, \vec{v}_2 \rangle$

Note that the choice of basis of a lattice isn't unique. This is unfortunate, because it changes the way we label the points in the lattice. For example, here is a different basis for the lattice pictured above:



The same lattice Λ *as above can be generated by a different basis:* $\Lambda = \langle \vec{w}_1, \vec{w}_2 \rangle$

Exercise 1. In each of the above two pictures, express the point **P** indicated in terms of the given basis.

In practice, one often proves things about lattices using algebra. Here are some simple examples of this.

Exercise 2. Recall that $\mathbb{Z}^2 = \{(a, b) : a, b \in \mathbb{Z}\}$.

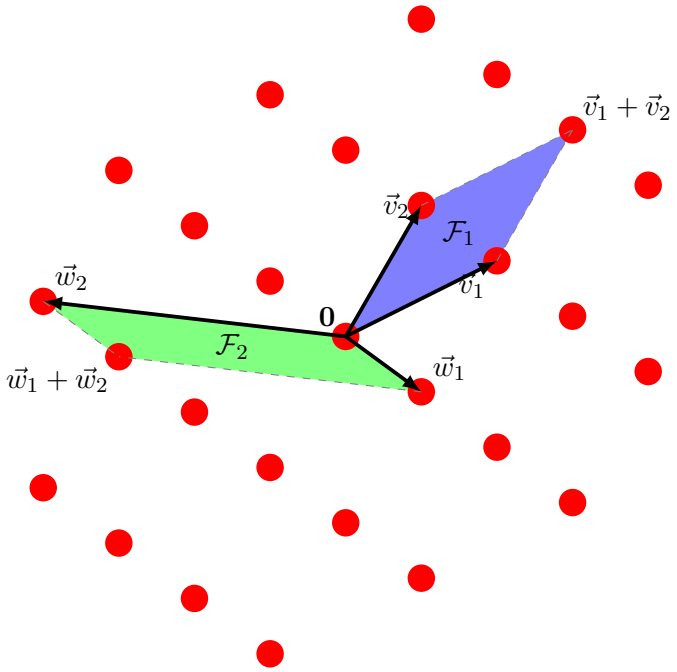
(a) Prove that $\mathbb{Z}^2 = \langle (1, 0), (0, 1) \rangle$.

(b) Prove that $\mathbb{Z}^2 = \langle (1, 0), (1, 1) \rangle$.

(c) Prove that $\mathbb{Z}^2 = \langle (2014, 1), (2015, 1) \rangle$.

(d) Give the simplest description you can of the lattice $\langle (1, 1), (2, 0) \rangle$. (Include all relevant proofs, of course!)

As we pointed out above, it's a bit frustrating that there are multiple bases of any given lattice (and hence, multiple labellings of the points of the lattice). It would be great if there were some invariant, i.e. some measurable quantity which is independent of the choice of basis. Let's look at a picture:



The same lattice can have multiple fundamental parallelepipeds, e.g. $\mathcal{F}_1$ and $\mathcal{F}_2$

Given a basis of a d-dimensional lattice $\Lambda = \langle \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d \rangle$, the *fundamental parallelepiped* of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d\}$ is the smallest parallelepiped whose edges contain the basis vectors. Another way to think about it is as the parallelepiped whose vertices are

$$\epsilon_1 \vec{v}_1 + \epsilon_2 \vec{v}_2 + \cdots + \epsilon_d \vec{v}_d$$

with all the $\epsilon_i = 0$ or 1. For technical reasons, it's convenient to exclude some of the boundary from the parallelepiped. Here's the formal definition: the fundamental parallelepiped $\mathcal{F}$ of the lattice basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_d\}$ is defined to be

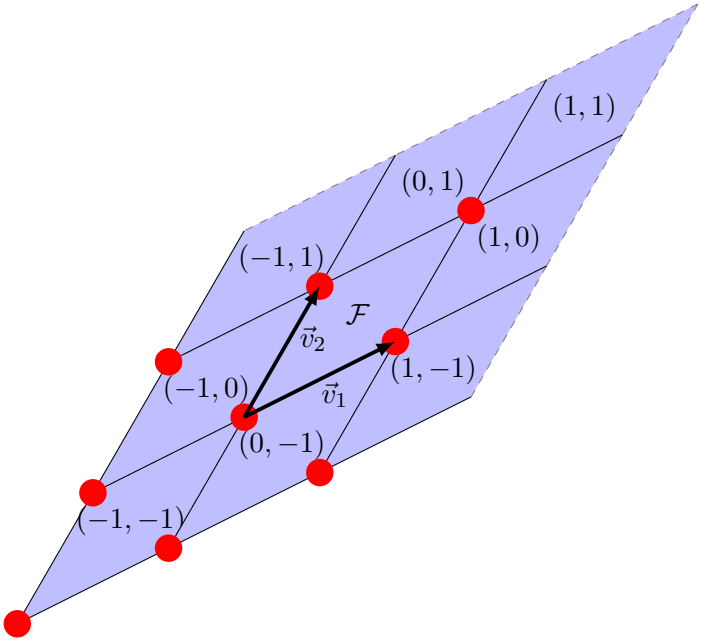
$$\mathcal{F} := \{\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_d \vec{v}_d : \alpha_i \in [0, 1) \ \forall i\}$$

Exercise 3.

(a) Draw the lattice $\langle (2, 1), (1, 0) \rangle$ and the corresponding fundamental parallelogram. Indicate which boundaries are excluded by drawing them as dashed lines. Similarly, indicate which points are included or excluded.

(b) Describe the fundamental parallelepiped of the lattice basis $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

What's so special about a fundamental parallelepiped? One nice observation is that we can use it to tile the whole plane. Here's an illustration:



Tiling the plane by translations of a fundamental parallelogram $\mathcal{F}$.
A region labelled (α, β) is the translation of $\mathcal{F}$ by $\alpha\vec{v}_1 + \beta\vec{v}_2$.

In the picture above, any point in the plane lies in a unique translation of the fundamental parallelogram $\mathcal{F}$. This might remind you of the construction of the floor function; in order to define $\lfloor x \rfloor$, we need the fact that every $x \in \mathbb{R}$ can be written *uniquely* in the form $n + \epsilon$ with $n \in \mathbb{Z}$ and $\epsilon \in [0, 1)$. A similar statement holds for points in Euclidean space.

Exercise 4. Suppose Λ is a d-dimensional lattice with a fundamental parallelepiped $\mathcal{F}$. Prove that any $\vec{x} \in \mathbb{R}^d$ can be written uniquely in the form

$$\vec{x} = \vec{n} + \vec{\epsilon},$$

where $\vec{n} \in \Lambda$ and $\vec{\epsilon} \in \mathcal{F}$.

Another nice feature of a fundamental parallelepiped is that its volume is an invariant of the lattice. In other words, a given lattice Λ may have many different fundamental parallelepipeds, but they all have the same volume. To prove this fact, it is helpful to introduce matrices into our description of lattices.

Any invertible $d \times d$ matrix M can be used to generate a lattice $\Lambda = M \cdot \mathbb{Z}^d$. It turns out that every lattice in $\mathbb{R}^d$ can be represented in this way! This is terrific news, because it allows us to employ the full arsenal of linear algebra to study lattices.

Exercise 5. Suppose we have a lattice $\Lambda = \langle \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_d \rangle$. Determine a matrix M (in terms of the $\vec{b}_i$) such that $\Lambda := M \cdot \mathbb{Z}^d$. [*Hint: think about how matrix multiplication works.*]

Exercise 6. Suppose the matrices M and M' generate the same lattice, with corresponding fundamental parallelepipeds $\mathcal{F}$ and $\mathcal{F}'$.

(a) Prove that there exists a unimodular[1] matrix U such that $M' = MU$.

(b) Prove that the volume of $\mathcal{F}$ is $|\det(M)|$.

The previous exercise implies that every fundamental parallelepiped of a given lattice has the same volume. We can therefore define the *volume of the lattice* Λ to be

$$\mathrm{vol}(\Lambda) := \mathrm{vol}(\mathcal{F}) \qquad \text{for any fundamental parallelepiped } \mathcal{F}.$$

[1] A matrix U is *unimodular* iff $|\det(U)| = 1$.

2 The Geometry of Numbers

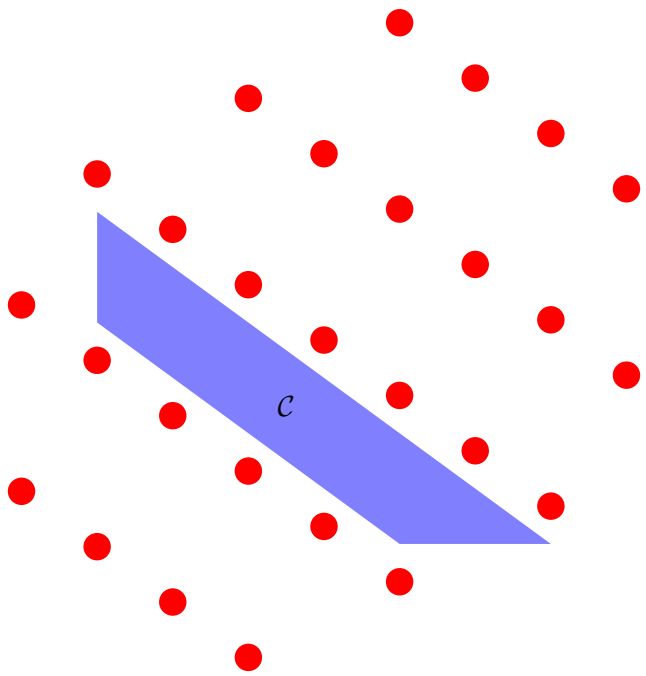
The Geometry of Numbers is primarily concerned with counting the number of lattice points in a given region. We warm up with an exercise.

Exercise 7. Let $B = B(R)$ denote the d-dimensional cube $[-R, R]^d$ centered at the origin, and let Λ be a non-denegerate d-dimensional lattice. Evaluate the limit

$$\lim_{R\to\infty} \frac{|B \cap \Lambda|}{\mathrm{vol}(B)}.$$

If you can prove your answer, great! If not, at least give an informal justification. [*Hint: Pick any fundamental parallelepiped $\mathcal{F}$. How many lattice points does $\mathcal{F}$ contain?*]

Thus, sufficiently large centrally symmetric cubes contain roughly the expected number of lattice points. What about other shapes in other positions? What does sufficiently large mean? One thing is immediately clear: having large volume doesn't guarantee that you contain a lot of lattice points. For example, the region $\mathcal{C}$ in the picture below can be extended arbitrarily far without containing any lattice points.



Arbitrarily large regions might not contain any lattice points

However, we do have the following version of the pigeonhole principle.

Theorem (Blichfeldt's Lemma[2])**.** Suppose $\Lambda \leq \mathbb{R}^d$ is a lattice and $C \subseteq \mathbb{R}^d$ is an open set such that $\mathrm{vol}(C) > \mathrm{vol}(\Lambda)$. Then $\exists x \neq y \in C$ with $x - y \in \Lambda$.

The following observation will be helpful.

Exercise 8. There exist distinct points $x, y \in C$ with $x - y \in \Lambda$ if and only if there exist distinct $\lambda_1, \lambda_2 \in \Lambda$ such that $(\lambda_1 + C) \cap (\lambda_2 + C) \neq \emptyset$.

[2]Hans Blichfeldt (1873–1945) was born into a poor family in Denmark. After immigrating to the United States at 15 and working on the railroads for six years, he saved enough money to enroll at Stanford. He spent almost the entire rest of his career at Stanford, and was the chair of the math department when he retired in 1938.

Proof of Blichfeldt's Lemma. Rather than proving Blichfeldt's Lemma directly, we prove the contrapositive. Precisely, we will show that if all the translates $\lambda+C$ (with $\lambda \in \Lambda$) are pairwise disjoint, then $\mathrm{vol}(C) \leq \mathrm{vol}(\Lambda)$.

Let $B := (-R, R)^d$ be a large box centered at the origin, and set

$$U = U_R := \bigcup_{\lambda \in B \cap \Lambda} (\lambda + C).$$

The strategy of the proof is to estimate the volume of U in two different ways. First, assuming that all the $\lambda + C$'s are pairwise disjoint, Exercise 7 implies

$$\mathrm{vol}(U) = \sum_{\lambda \in B \cap \Lambda} \mathrm{vol}(C) = \mathrm{vol}(C) \cdot |B \cap \Lambda| = \Big(1 + o(1)\Big)(2R)^d \frac{\mathrm{vol}(C)}{\mathrm{vol}(\Lambda)}$$

where $o(1) \to 0$ as $R \to \infty$. On the other hand, U is a subset of the box $\Big(- R - \mathrm{diam}(C), R + \mathrm{diam}(C)\Big)^d$, so

$$\mathrm{vol}(U) \leq 2^d \Big(R + \mathrm{diam}(C)\Big)^d.$$

It follows that

$$\Big(1 + o(1)\Big) \frac{\mathrm{vol}(C)}{\mathrm{vol}(\Lambda)} \leq \left(1 + \frac{\mathrm{diam}(C)}{R}\right)^d.$$

Letting $R \to \infty$ yields

$$\mathrm{vol}(C) \leq \mathrm{vol}(\Lambda)$$

as claimed. □

As a corollary we obtain the following beautiful result, which asserts that any nice sufficiently large set must contain a nontrivial lattice point.

Theorem (Minkowski's 1st Theorem)**.** Given a lattice $\Lambda \leq \mathbb{R}^d$, and suppose $C \subseteq \mathbb{R}^d$ is open, convex, and centrally symmetric (i.e. symmetric about the origin). If $\mathrm{vol}(C) > 2^d \mathrm{vol}(\Lambda)$, then C contains a nonzero lattice point.

Exercise 9. Proof of Minkowski's 1st Theorem.

(a) Let C be as in the statement of the theorem. Prove that $C = \frac{1}{2}C - \frac{1}{2}C$.

(b) Conclude the proof. [*Hint: Apply Blichfeldt's Lemma to* $\frac{1}{2}C$.]

Given a lattice $\Lambda \subseteq \mathbb{R}^d$ and *any* open convex centrally-symmetric set $C \subseteq \mathbb{R}^d$, it is evident that sufficiently large dilations $\lambda \cdot C := \{\lambda x : x \in C\}$ must contain a lattice point; Minkowski's 1st theorem makes precise what *sufficiently large* means. To see this, let λ_1 denote the least dilation of C which contains a nonzero lattice point. Since $\mathrm{vol}(\lambda \cdot C) = \lambda^d \mathrm{vol}(C)$, Minkowski's 1st theorem implies

$$\lambda_1^d \leq 2^d \left(\frac{\mathrm{vol}(\Lambda)}{\mathrm{vol}(C)}\right). \tag{1}$$

Actually, there's a subtle flaw in the above discussion: λ_1 isn't well-defined, because C is an open set. Fortunately, this is easy to fix. Define

$$\lambda_1 := \inf\{\lambda > 0 : \lambda \cdot C \text{ contains a nonzero lattice point}\}.$$

Now λ_1 is well-defined, and Minkowski's 1st still implies the inequality (1).

Having found one lattice point inside some dilation of C, we can now become more ambitious and ask how much we must dilate by to find more. Let λ_2 denote the smallest dilation of C which contains two *linearly independent* lattice points; as before, for this to be well-defined we set

$$\lambda_2 := \inf\{\lambda > 0 : \lambda \cdot C \text{ contains two linearly independent lattice points}\}.$$

More generally, whenever $2 \leq n \leq d$ let

$$\lambda_n := \inf\{\lambda > 0 : \lambda \cdot C \text{ contains } n \text{ linearly independent lattice points}\};$$

in particular, we have $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. The following theorem is a strengthening of (1):

Theorem (Minkowski's 2nd Theorem)**.** Given a lattice $\Lambda \leq \mathbb{R}^d$, and suppose $C \subseteq \mathbb{R}^d$ is open, convex, and centrally symmetric (i.e. symmetric about the origin). Then

$$\lambda_1 \lambda_2 \cdots \lambda_d \leq 2^d \left(\frac{\text{vol}(\Lambda)}{\text{vol}(C)} \right).$$

Since (1) is equivalent to Minkowski's 1st theorem, we see that Minkowski's 2nd theorem is stronger than the 1st. It's also significantly more difficult to prove, and we shall not do so here.

Finally, we show how to apply this result to the proof of the Bohr-to-gAP proposition (see the very end of the Lecture 11 notes). Recall that Λ was the d-dimensional lattice generated by $N\mathbb{Z}^d$ and some vector $\vec{r} = (r_1, r_2, \ldots, r_d)$, where N is prime and the r_i's are distinct residues (mod N). We chose $\vec{g}_1$ to be a shortest nonzero lattice vector, and $\vec{g}_j$ to be a shortest lattice vector linearly independent of $\vec{g}_i$ for all $i \leq j$. In the last step of the proof, we used the bound

$$\prod_{i=1}^{d} |\vec{g}_i| \ll_d N^{d-1}.$$

Where does this come from? Taking C to be the d-dimensional unit ball[3], we see that $\lambda_i = |\vec{g}_i|$, so Minkowski's 2nd theorem gives the bound

$$\prod_{i=1}^{d} |\vec{g}_i| \leq 2^d \left(\frac{\text{vol}(\Lambda)}{\text{vol}(C)} \right) \ll_d \text{vol}(\Lambda).$$

It therefore suffices to prove that $\text{vol}(\Lambda) \leq N^{d-1}$.

Exercise 10. Let Λ, d, N, and $\vec{r}$ be as above.

(a) Prove that there exists $\vec{v} \in \Lambda$ one of whose coordinates is equal to 1.

(b) Prove that there exists a d-dimensional polytope, all of whose vertices are in Λ, which has volume N^{d-1}.

(c) Prove that $\text{vol}(\Lambda) \leq N^{d-1}$.

Aside from proving Minkowski's 2nd theorem, this completes the proof of the Bohr-to-gAP proposition.

3 Freiman-Homomorphisms

To finish the proof of the Freiman-Ruzsa Theorem, we need a tool which translates proper gAPs in $\mathbb{Z}_N$ into proper gAPs in $\mathbb{Z}$. For example, any homomorphism $\phi : G \to H$ between abelian groups $(G, +)$ and $(H, +)$ preserves arithmetic progressions. To see this, suppose a, b, c is a 3-term arithmetic progression in G. Then $a + c = b + b$, whence $\phi(a) + \phi(c) = \phi(a + c) = \phi(b + b) = \phi(b) + \phi(b)$; it follows that $\phi(a), \phi(b), \phi(c)$ is a 3-term arithmetic progression in H. Moreover, if ϕ is injective, then it preserves properness of an arithmetic progression. It is clear that ϕ preserves gAPs as well as APs.

Note that we did not require the full power of a homomorphism above – all we really needed was that if $a + c = b + b$, then $\phi(a) + \phi(c) = \phi(b) + \phi(b)$. This inspires the following definition.

Definition (Freiman 2-homomorphism)**.** Suppose $(G, +)$ and $(H, +)$ are abelian groups, $A \subseteq G$, and $B \subseteq H$. A map $\phi : A \to B$ is a *Freiman 2-homomorphism* if and only if $\forall a, b, x, y \in A$,

$$a + b = x + y \implies \phi(a) + \phi(b) = \phi(x) + \phi(y).$$

In view of Bogolyubov's Lemma, for our application to the Freiman-Ruzsa Theorem we are interested in gAPs inside $2A - 2A$. Thus rather than equations of the form $x + y = x' + y'$, we will need to consider thing like $x + y - w - z = x' + y' - w' - z'$. We therefore define a suitable generalization of the above.

[3]Recall that the volume of the d-dimensional unit ball is $\frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$.

Definition (Freiman k-homomorphism)**.** Suppose $(G,+)$ and $(H,+)$ are abelian groups, $A \subseteq G$, and $B \subseteq H$. A map $\phi : A \to B$ is a *Freiman k-homomorphism* if and only if $\forall a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_k \in A$,

$$a_1 + a_2 + \cdots + a_k = b_1 + b_2 + \cdots + b_k \implies \phi(a_1) + \cdots + \phi(a_k) = \phi(b_1) + \cdots + \phi(b_k).$$

Exercise 11. Let $(G,+)$, $(H,+)$, and $(K,+)$ be abelian groups, and suppose $A \subseteq G$, $B \subseteq H$, and $C \subseteq K$. Prove the following properties of Freiman k-homomorphisms:

1. What is a Freiman 1-homomorphism from A to B?
2. If $\phi : G \to H$ is a homomorphism, then it is a Freiman k-homomorphism from A to B for any $k \geq 1$.
3. If $\phi : A \to B$ and $\psi : B \to C$ are Freiman k-homomorphisms, then their composition $\psi \circ \phi$ is a Freiman k-homomorphism (provided it exists and is well defined).
4. If $\phi : A \to B$ is a Freiman k-homomorphism, then it is also a Freiman ℓ-homomorphism for all $\ell \leq k$.
5. Give an example where $\phi : A \to B$ is a Freiman k-homomorphism for some $k \geq 2$, but is not a Freiman ℓ-homomorphism for some $\ell > k$.

Exercise 12. Let $(G,+)$ be an abelian group. Fix $a \in G$ and let $T_a : G \to G$ be the translation by a: $g \mapsto g + a$. Show that T_a is a Freiman k-homomorphism from G to itself, for any k.

Exercise 13. Since the composition of any two Freiman k-homomorphisms is a k-homomorphism, the above exercise implies that the translation of any homomorphism from G to itself is also a k-homomorphism for every k. Is the converse to this statement true? More precisely, let $(G,+)$ be an abelian group, and suppose that $\phi : G \to G$ is a Freiman k-homomorphism for every k. Is ϕ necessarily the translation of some homomorphism $\psi : G \to G$?