PRODUCT SET ESTIMATES FOR NON-COMMUTATIVE GROUPS

TERENCE TAO

ABSTRACT. We develop the Plünnecke-Ruzsa and Balog-Szemerédi-Gowers theory of sum set estimates in the non-commutative setting, with discrete, continuous, and metric entropy formulations of these estimates. We also develop a Freiman-type inverse theorem for a special class of 2-step nilpotent groups, namely the Heisenberg groups with no 2-torsion in their centre.

1. INTRODUCTION

The field of *additive combinatorics* is concerned with the structure and size properties of sum sets such as $A + B := \{a + b : a \in A, b \in B\}$ for various sets A and B (in some additive group G). One also considers partial sum sets such as $A \overset{E}{+} B := \{a + b : (a, b) \in E\}$ for some[1] $E \subset A \times B$. There are many deep and important results in this theory, but we shall mention three particularly important ones. Firstly, there are the *Plünnecke-Ruzsa sum-set estimates*, which roughly speaking asserts that if one sum-set such as $A + B$ is small, then other sum-sets such as $A - B$, $A + B + B$, $A + A$, etc. are also small; see e.g. [29], [30], [28], [36]. Then there is the *Balog-Szemerédi-Gowers theorem* [1], [17], which roughly speaking asserts that if a partial sum-set $A \overset{E}{+} B$ is small for some dense subset E of $A \times B$, then there are large subsets A', B' of A, B respectively whose *complete* sum-set $A' + B'$ is also small. Finally, there are *inverse sum set theorems*, of which *Freiman's theorem* [16] (see also [2], [31], [9], [28], [36]) is the most famous: it asserts that if A is a finite non-empty subset of a torsion-free abelian group (such as $\mathbf{Z}^d$) with $A + A$ small, then A can be efficiently contained in the sum of $O(1)$ arithmetic progressions. These three families of results have had many applications, perhaps most strikingly to the work of Gowers [17], [18] on quantitative bounds for Szemerédi's theorem and to the work of Bourgain and co-authors [4], [5], [6], [7] on exponential sum estimates in finite fields. We refer the reader to [36] for a more detailed treatment of these topics.

The above results are usually phrased in the discrete setting, with A and B being a finite subset of an abelian group such as the lattice $\mathbf{Z}^d$, and with the cardinality $|A|$ of a set A used as a measure of size. However, it is easy to transfer these discrete results to a continuous setting, for instance when A and B are open bounded subsets

[1]We use E here instead of the more traditional G, as we are reserving G for the ambient group.

of a Euclidean space $\mathbf{R}^d$, and using Lebesgue measure $\mu(A)$ rather than cardinality to measure the size of a set. Indeed one can pass from the continuous case to the discrete case (possibly losing constants which are exponential in the dimension d) by discretizing $\mathbf{R}^d$ to a fine lattice such as $\varepsilon \cdot \mathbf{Z}^d$, applying the discrete sum-set theory, and then taking limits as $\varepsilon \to 0$. For similar reasons, there is little difficulty in transferring the sum-set theory to a metric entropy setting, in which the size of a set A in $\mathbf{R}^d$ is measured using a covering number $\mathcal{N}_\varepsilon(A)$. See for instance Propositions 7.1, 7.3 below for examples of these transference techniques.

In this paper we present analogues of the Plünnecke-Ruzsa and Balog-Szemerédi theorems in the non-commutative setting, in which G is now a multiplicative group and one studies the size of product sets $A \cdot B := \{a \cdot b : a \in A, b \in B\}$ and partial product sets $A \overset{E}{\cdot} B := \{a \cdot b : (a,b) \in E\}$; the theory for inverse sumset estimates is significantly more complicated and does not seem to easily extend to the non-commutative setting (as the Fourier-analytic techniques are substantially less effective in this case), though we are able to obtain an inverse theorem for a class of Heisenberg groups, see Theorem 7.12 below. The other two results, however, are more elementary in nature, and much of the theory carries over surprisingly easily to this setting. The one result which fails utterly is the Plünnecke magnification inequality [29], which is valid only for commutative graphs and has no known counterpart in the non-commutative setting. Fortunately, one can use other, more elementary combinatorial arguments as a substitute for the Plünnecke inequalities (as was done in [36]), albeit at the cost of degrading the exponents in the estimates slightly. Another significant issue is that there is no obvious relationship between the size of $A \cdot B$ and of $B \cdot A$ in the non-commutative setting; consider for instance the case when A is a subgroup of G, and B is a right coset of A. Fortunately, there is a residual relationship between the sets $A \cdot A^{-1}$ and $A^{-1} \cdot A$ in that if one product set is small then a large portion of the other product set is small (see Lemma 4.3 below). This key observation allows us to get around the obstruction of non-commutativity and recover almost all the standard sum-set theory, though in some cases one has to throw out a small exceptional set in order to proceed. For instance, if A is the union of a subgroup and a point, then $A \cdot A$ will be small, but higher products such as $A \cdot A \cdot A$ can be large; however, by throwing out this exceptional point one can control all products.

Finally, the passage between the discrete setting, the continuous setting, and the metric entropy setting is not as automatic in the non-commutative setting (such as for the group $SU(2)$) as it is in the case of Euclidean spaces $\mathbf{R}^d$, because there are usually no good analogues of the discrete subgroups $\varepsilon \cdot \mathbf{Z}^d$ in the general setting. Fortunately, the continuous and discrete theories are almost identical, so much so that we shall treat the two in a unified manner. One can then pass from the continuous setting to a metric entropy setting by standard volume packing arguments, provided that the metric structure is sufficiently compatible with the group structure. Ideally one wants the metric to be bi-invariant, but this is usually only possible when the group G is compact. For non-compact groups such as $SL(2, \mathbf{R})$, the metric entropy results that we present here are only satisfactory when all the sets under consideration are contained in a fixed bounded set, in which case the metric structure will be approximately bi-invariant (or more precisely, the group

operations are Lipschitz) and the metric balls will obey a doubling property, in which case the volume packing arguments go through without difficulty.

Our main results are as follows. In both the continuous and discrete setting, we classify sets of small tripling, showing that such sets are nothing more than dense subsets of a type of set that we call an *approximate group*; see Theorem 3.10. As for sets of small doubling (or pairs of sets with small product set), we have a slightly different classification, showing that such sets can be covered efficiently by left or right-translates of an approximate group (Theorem 4.6). For pairs of sets with small *partial* product set, we show that such sets have large *intersection* with translates of an approximate group of comparable size (Theorem 5.4). In Section 6 we extend these results to the metric entropy setting, given some mild hypotheses on the metric. Finally, in Section 7 we discuss the inverse product set problem (the noncommutative generalisation of the inverse sum set problem) and present a new theorem in this direction in the context of Heisenberg groups in the absence of 2-torsion. All of these results are *polynomially reversible* in the sense that we can pass from one class of sets to an equivalent class and then back to the original class, losing only polynomial factors in the parameter K (which should be thought of as a type of doubling constant).

The author thanks Jean Bourgain for encouragement, and for raising the issue of the metric entropy case in the non-commutative setting. He also thanks Imre Ruzsa for very detailed comments and suggestions, and Emmanuel Kowalski for corrections. This work developed from some earlier unpublished notes of the author [35], as well as from portions of the author's book with Van Vu [36]. In particular, the discrete versions of the results here can largely be found in [36, §2.7], although in some cases the proofs are assigned as exercises rather than given in full. The differences between the discrete and the continuous arguments are mostly notational in nature.

2. Setup and notation

We now give the unified framework in which to present the discrete and continuous non-commutative sum-set (or more precisely product-set) theory.

Definition 2.1 (Multiplicative groups)**.** A *multiplicative group* will be a topological group G (thus the group operation $(x, y) \mapsto x \cdot y$ and the inversion operation $x \mapsto x^{-1}$ are continuous), equipped with a *Haar measure* μ, which for us will be a non-negative Radon measure on G which is invariant under left and right translation and inversion, thus $\mu(x \cdot A) = \mu(A \cdot x) = \mu(A^{-1}) = \mu(A)$ for all measurable A in G, where $x \cdot A := \{x \cdot y : y \in G\}$, $A \cdot x := \{y \cdot x : y \in G\}$, and $A^{-1} := \{x^{-1} : x \in G\}$. We denote the multiplicative identity by $1 = 1_G$. We also make the mild non-degeneracy assumption that every non-empty open set has non-zero measure. A *multiplicative set* will be any non-empty open precompact set A in G; note that we necessarily have $0 < \mu(A) < \infty$. Given two multiplicative sets A and B we define their product $A \cdot B := \{a \cdot b : a \in A, b \in B\}$; observe that this is also a multiplicative set, as is the inverse set A^{-1}.

Remarks 2.2. The hypotheses that a multiplicative set is open and precompact (and that μ is Radon) will allow us to avoid many technical issues concerning

measurability and integrability, and we shall in fact not discuss these issues here. Note that we are implicitly assuming that G is locally compact, since otherwise there will be no multiplicative sets to consider. One could weaken the translation and inversion invariance properties of the measure somewhat (so that the group operations only preserve the measure *approximately*) but this would introduce a number of measure-dependent constants into the estimates below and we will not do so here. However, such a generalisation would be useful for studying non-unimodular Lie groups.

We now give the two main examples of multiplicative groups.

Example 2.3 (Discrete case). Let G be an abstract group (not necessarily abelian). Then we can equip this group with the discrete topology and counting measure $\mu(A) = |A|$ to obtain a multiplicative group. In this case, the multiplicative sets are simply the finite non-empty sets.

Example 2.4 (Unimodular Lie group case). Let G be a finite-dimensional unimodular Lie group. This is a finite-dimensional manifold and thus comes with a standard topology, and a standard Haar measure (defined up to a normalizing scalar). The multiplicative sets in this case are the non-empty bounded open sets.

Remark 2.5. In the commutative setting one can pass between the discrete and continuous cases above by standard discretisation arguments, but the connection between the two is less clear in the non-commutative setting. Nevertheless we shall be able to treat both of these cases in a completely unified manner.

Remark 2.6. Observe that the hypotheses on the measure μ are preserved if we multiply the measure μ by a positive constant. Thus all the estimates we present in this paper will be invariant under this symmetry; roughly speaking, this means that the number of times μ appears on the left-hand side of an equality will always equal the number of times μ appears on the right-hand side. (Certain quantities such as the Ruzsa distance $d(A, B)$ and the doubling constant K will be dimensionless, whereas the multiplicative energy $\mathrm{E}(A, B)$, which we define below, has the units of μ^3.)

Henceforth we fix the multiplicative group G (and the measure μ). In the next few sections we study how the measure of various products such as $\mu(A \cdot B)$, $\mu(A \cdot A)$, $\mu(A \cdot A \cdot A)$, etc. of multiplicative sets are related.

We shall use the notation $X = O(Y)$, $Y = \Omega(X)$, $X \lesssim Y$ or $Y \gtrsim X$ to denote the statement that $X \leq CY$ for an absolute constant C (not depending on the choice of group G or on any other parameters). We also use $X \sim Y$ to denote the estimates $X \lesssim Y \lesssim X$. If we wish to indicate dependence of the constant on an additional parameter, we will subscript the notation appropriately, thus for instance $X \sim_n Y$ denotes that $X \leq C_n Y$ and $Y \leq C_n X$ for some C_n depending on n.

3. Ruzsa distance, and tripling sets

To measure the multiplicative structure inherent in a multiplicative set A, or a pair A, B of multiplicative sets, it is convenient to introduce two measurements, the *Ruzsa distance* and the *multiplicative energy*. In this section we focus on the Ruzsa distance and applications to sets of small tripling.

Definition 3.1 (Ruzsa distance). Let A and B be multiplicative sets. We define the *(left-invariant) Ruzsa distance* $d(A, B)$ to be the quantity

$$d(A, B) := \log \frac{\mu(A \cdot B^{-1})}{\mu(A)^{1/2}\mu(B)^{1/2}}.$$

We now justify the terminology "left-invariant[2] Ruzsa distance".

Lemma 3.2 (Ruzsa triangle inequality). *Let A, B, C be multiplicative sets. Then we have $d(A, B) \geq 0$, $d(A, B) = d(B, A)$, and $d(A, C) \leq d(A, B) + d(B, C)$. Also we have $d(x \cdot A, x \cdot B) = d(A, B)$ for all $x \in G$.*

Remark 3.3. This inequality was first established in the discrete case in [30] (initially in the commutative case, but the argument extends easily to non-commutative settings).

Proof From translation invariance we have $\mu(A \cdot B^{-1}) \geq \mu(A \cdot b^{-1}) = \mu(A)$ for any $b \in B$. Similarly $\mu(A \cdot B^{-1}) \geq \mu(B^{-1}) = \mu(B)$. Taking geometric means we obtain $\mu(A \cdot B^{-1}) \geq \mu(A)^{1/2}\mu(B)^{1/2}$ and hence $d(A, B) \geq 0$. The symmetry property $d(A, B) = d(B, A)$ follows from the fact that $B \cdot A^{-1}$ is the inverse of $A \cdot B^{-1}$. Finally, to show the triangle inequality $d(A, C) \leq d(A, B) + d(B, C)$, it will suffice to show the inequality

$$\mu(A \cdot C^{-1})\mu(B) \leq \mu(A \cdot B^{-1})\mu(B \cdot C^{-1}).$$

To prove this, we rewrite the right-hand side as a double integral

$$\int_G \int_G 1_{A \cdot B^{-1}}(x)1_{B \cdot C^{-1}}(y) \, d\mu(x)d\mu(y),$$

where 1_A denotes the indicator function of A. Making the substitution $x = z \cdot y^{-1}$ and using the translation invariance and Fubini's theorem, we can rewrite this as

$$\int_G \left[\int_G 1_{A \cdot B^{-1}}(z \cdot y^{-1})1_{B \cdot C^{-1}}(y) \, d\mu(y) \right] d\mu(z).$$

Now if z lies in $A \cdot C^{-1}$, then we have $z = a \cdot c^{-1}$ for some $a \in A$ and $c \in C$, and then $1_{A \cdot B^{-1}}(z \cdot y^{-1})1_{B \cdot C^{-1}}(y) = 1$ whenever $y \in B \cdot c^{-1}$. Since $B \cdot c^{-1}$ has the same measure as B, the above integral is at least as large as

$$\int_{A \cdot C^{-1}} \mu(B) \, d\mu(z) = \mu(A \cdot C^{-1})\mu(B)$$

and the claim follows.   ∎

---

[2]One could also define a right-invariant Ruzsa distance $\tilde{d}(A, B) := d(A^{-1}, B^{-1}) = \log \frac{\mu(B^{-1} \cdot A)}{\mu(A)^{1/2}\mu(B)^{1/2}}$, but we will not need that notion here.

We caution that $d(A, A) = \log \frac{\mu(A \cdot A^{-1})}{\mu(A)}$ will usually not be zero. Also, $d(A, B) \neq d(A^{-1}, B^{-1})$ and $d(A, B) \neq d(A \cdot x, B \cdot x)$ in general.

From the Ruzsa triangle inequality we see in particular that

$$d(A, A) \leq 2d(A, B) \tag{1}$$

for all multiplicative sets A, B.

For any integer $n \geq 1$ and any multiplicative set A, let A^n denote the n-fold product set

$$A^n = A \cdot \ldots \cdot A = \{a_1 \ldots a_n : a_1, \ldots, a_n \in A\}.$$

In the commutative setting, the Plünnecke-Ruzsa inequalities show that if A^2 is comparable in size to A, then A^n is also comparable in size to A for any fixed n. The same statement is not necessarily true in the non-commutative setting: for instance, in the discrete setting if we take $A = H \cup \{x\}$, where H is a finite subgroup of G and x lies outside of the normalizer of H, then A^2 has size comparable to A, but A^3 can be much larger. However, (as was observed in [22] in the discrete non-commutative case, although the basic argument is essentially in [33]) it turns out that once A^3 is under control, then so are all other combinations of A and A^{-1}:

Lemma 3.4. *Let A be a multiplicative set such that $\mu(A^3) \leq K\mu(A)$. Then for any signs $\epsilon_1, \ldots, \epsilon_n \in \{-1, 1\}$ we have $\mu(A^{\epsilon_1} \ldots A^{\epsilon_n}) \leq K^{O_n(1)}\mu(A)$.*

Proof Let us first observe from hypothesis that $\mu(A) \leq \mu(A^2) \leq \mu(A^3) \leq K\mu(A)$. This implies that $d(A, A^{-1}) \leq \log K$ and $d(A^2, A^{-1}) \leq \log K$. By the triangle inequality we thus have $d(A^2, A) = O(\log K)$, and thus $\mu(A \cdot A \cdot A^{-1}) \leq K^{O(1)}\mu(A)$, which implies that $d(A, A \cdot A^{-1}) = O(\log K)$. By the triangle inequality again this implies $d(A \cdot A^{-1}, A^{-1}) = O(\log K)$, hence $\mu(A \cdot A^{-1} \cdot A) \leq K^{O(1)}\mu(A)$. In particular $d(A, A^{-1} \cdot A) = O(\log K)$, so again by the triangle inequality $d(A^{-1}, A^{-1} \cdot A) = O(\log K)$, and hence $\mu(A^{-1} \cdot A \cdot A) \leq K^{O(1)}\mu(A)$. With all these bounds (and taking inverses) we can already establish the lemma when $n = 3$, which also implies the lemma when $n < 3$.

Now we assume inductively that the lemma is already proven for all $n < n_0$ for some $n_0 \geq 4$, and wish to prove it for $n = n_0$. To establish the bound on $A^{\epsilon_1} \ldots A^{\epsilon_n}$ it suffices to establish the bound

$$d(A^{\epsilon_1} \ldots A^{\epsilon_{n-2}}, A^{-\epsilon_n} \cdot A^{-\epsilon_{n-1}}) = O_n(\log K).$$

But since the lemma is already proven for $n - 1$, we have

$$d(A^{\epsilon_1} \ldots A^{\epsilon_{n-2}}, A) = O_n(\log K)$$

and since the lemma is already proven for 3, we have

$$d(A, A^{-\epsilon_n} \cdot A^{-\epsilon_{n-1}}) = O(\log K)$$

and so the claim follows from the triangle inequality.   ∎

Remark 3.5. One can weaken the condition $\mu(A^3) \leq K|A|$ to $\sup_{a \in A} |A \cdot a \cdot A| \leq K|A|$; see Corollary 4.8.

One can analyze the behavior of tripling sets further, by the following covering lemma.

Lemma 3.6 (Ruzsa covering lemma). *Let A, B be multiplicative sets such that $\mu(A \cdot B) \leq K\mu(A)$ (resp. $\mu(B \cdot A) \leq K\mu(A)$). Then there exists a finite set X contained inside B of cardinality at most K such that $B \subseteq A^{-1} \cdot A \cdot X$ (resp. $B \subseteq X \cdot A \cdot A^{-1}$).*

Remark 3.7. For the commutative version of this lemma (in discrete or continuous settings), see [32], [27], [36].

Proof By the reflection invariance of μ it will suffice to prove the claim when $\mu(A \cdot B) \leq K\mu(A)$. Let X be a subset of B with the property that the sets $A \cdot x$ for $x \in X$ are disjoint. Since $\mu(A \cdot B) \leq K\mu(A)$ we see that such a set X must have cardinality at most K. Now let X be such a set which is maximal with respect to set inclusion (which one can construct for instance using the greedy algorithm). Then for any $b \in B$ we must have $A \cdot b$ intersecting $A \cdot x$ for some $x \in X$, which implies that $b \in A^{-1} \cdot A \cdot X$. The claim follows. ∎

We can now give a classification of sets of small tripling.

Definition 3.8 (Approximate groups). A multiplicative set H is said to be a *K-approximate group* if it is symmetric (so $H^{-1} = H$) and there exists a finite symmetric set $X \subset H^2$ of cardinality at most K such that $H \cdot H \subseteq X \cdot H$.

Remark 3.9. In [36], the additional condition $X \cdot H \subset H \cdot X \cdot X$ was also imposed. Our methods for constructing approximate groups also yield this additional property (see Theorem 3.10 below), though we will not use this additional hypothesis in our arguments and thus omit it from the definition of an approximate group.

Note from the symmetry assumptions that if $H \cdot H \subset X \cdot H$, then $H \cdot H \subset H \cdot X$ also. Iterating this we see that $H^n \subseteq X^{n-1} \cdot H, H \cdot X^{n-1}$ for all $n \geq 1$. Thus approximate groups have small tripling. It turns out that this is essentially the only way that a multiplicative set can have small tripling:

Theorem 3.10. *Let $K \geq 1$, and let A be a multiplicative set. Then the following three statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statements hold for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

   (i) *We have the tripling bound $\mu(A^3) \lesssim K^{O(1)}\mu(A)$.*
   (ii) *We have $\mu(A^{\epsilon_1} \ldots A^{\epsilon_n}) \sim_n K^{O_n(1)}\mu(A)$ for all $n \geq 1$ and all signs $\epsilon_1, \ldots, \epsilon_n \in \{-1, +1\}$.*
   (iii) *There exists a $O(K^{O(1)})$-approximate group H of size $\mu(H) \sim K^{O(1)}\mu(A)$ which contains A.*

Proof The implication (i) $\implies$ (ii) is Lemma 3.4, while the reverse implication (ii) $\implies$ (i) is trivial. The implication (iii) $\implies$ (i) is also trivial:

$$\mu(A^3) \leq \mu(H^3) \lesssim K^{O(1)}\mu(H) \lesssim K^{O(1)}\mu(A).$$

It remains to show that (i) implies (iii). Set $H_0 := A \cup \{1\} \cup A^{-1}$ and $H := H_0^3$, then from Lemma 3.4 we see that $\mu(H) \sim K^{O(1)}\mu(A)$. Clearly H also contains A, so it remains to show that H is a $O(K^{O(1)})$-approximate group. Certainly H is symmetric. From Lemma 3.4 we have $\mu(H_0 \cdot H^2) \lesssim K^{O(1)}\mu(A)$, and hence from Lemma 3.6 we can find a finite set Y in H^2 of cardinality $O(K^{O(1)})$ such that

$$H^2 \subseteq H_0^{-1} \cdot H_0 \cdot Y \subseteq H \cdot Y.$$

If we set $X := Y \cup Y^{-1}$ then X is symmetric and (from symmetry of H) we conclude that $H^2 \subseteq H \cdot X, X \cdot H$. But since X is contained in H^2, we also have

$$H \cdot X \subseteq H \cdot H^2 = H^2 \cdot H \subseteq X \cdot H \cdot H \subseteq X \cdot X \cdot H$$

and similarly $H \cdot X \subseteq H \cdot X \cdot X$. The claim follows. ∎

An inspection of the above proof reveals the following more precise implication of (i) from (iii):

Corollary 3.11. *Let A be a multiplicative set such that $\mu(A^3) \leq K\mu(A)$. Then the set $H := (A \cup \{1\} \cup A^{-1})^3$ is a $O(K^{O(1)})$-approximate group. In particular, if A is symmetric and contains 1, then A^3 is a $O(K^{O(1)})$-approximate group.*

4. Convolution and multiplicative energy

To study sets of small doubling, rather than small tripling, it is convenient to introduce another measure of multiplicative structure between two multiplicative sets, namely the *multiplicative energy*. Given two absolutely integrable functions f, g on G, we define their *convolution* $f * g$ in the usual manner as

$$f * g(x) := \int_G f(y)g(y^{-1}x) \, d\mu(y) = \int_G f(xy^{-1})g(y) \, d\mu(y).$$

As is well-known, convolution is bilinear and associative (though not necessarily commutative), and the convolution of two absolutely integrable functions is continuous. Convolution is not commutative in general, but we do have the identity

$$f * g(1) = g * f(1), \tag{2}$$

which reflects the fact that $xx^{-1} = x^{-1}x = 1$. If f is supported on A and g is supported on B then $f * g$ is supported on $A \cdot B$. If we use $\tilde{f}(x) := f(x^{-1})$ to denote the reflection of f, we observe the reflection property

$$\widetilde{f * g} = \tilde{g} * \tilde{f} \tag{3}$$

(which reflects the fact that $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$ for $x, y \in G$) and the trace formula

$$\int_G f(x)g(x) \, d\mu(x) = f * \tilde{g}(1) = \tilde{g} * f(1) = \tilde{f} * g(1) = f * \tilde{g}(1). \tag{4}$$

If A is a multiplicative set, we use 1_A to denote the indicator function of A; note that this is an absolutely integrable function.

Definition 4.1. Let A, B be multiplicative sets. We define the *multiplicative energy* $\mathrm{E}(A, B)$ between these two sets to be the quantity

$$\mathrm{E}(A, B) := \int_G [1_A * 1_B(x)]^2 \ d\mu(x).$$

Remark 4.2. In the discrete setting (Example 2.3), we have

$$\mathrm{E}(A, B) = |\{(a, b, a', b') \in A \times B \times A \times B : a \cdot b = a' \cdot b'\}|. \tag{5}$$

In the notation of Gowers [17], the quantity $\mathrm{E}(A, B)$ thus counts the number of *multiplicative quadruples* in $A \times B \times A \times B$. In the commutative case, this quantity has a useful representation in terms of the Fourier transform (see e.g. [17], [36]), which has a number of applications, for instance in proving Freiman's inverse sumset theorem. In the non-commutative case it is also possible to use the non-commutative Fourier transform to represent this energy, but the resulting formulae are not as tractable. In particular, no analogue of Freiman's theorem is currently known in the general non-commutative setting. Fortunately, we will be able to use the properties of the convolution algebra (most notably its associativity, the reflection property (3), and the trace property (4)) to compensate for the lack of a convenient Fourier-analytic description of the energy; in particular, we will not need to understand the representation theory of the underlying group G.

A simple application of Fubini's theorem and change of variables shows that

$$\int_G 1_A * 1_B(x) \ d\mu(x) = \mu(A)\mu(B) \tag{6}$$

and

$$1_A * 1_B(x) \leq \min(\mu(A), \mu(B)) \leq \mu(A)^{1/2}\mu(B)^{1/2} \tag{7}$$

and hence by Hölder's inequality we have the upper bound

$$\mathrm{E}(A, B) \leq \mu(A)^{3/2}\mu(B)^{3/2}. \tag{8}$$

Also, since $1_A * 1_B$ is supported on $A \cdot B$, we have a lower bound from (6) and Cauchy-Schwarz:

$$\mathrm{E}(A, B) \geq \frac{\mu(A)^2\mu(B)^2}{\mu(A \cdot B)}. \tag{9}$$

In general, we do not have $\mathrm{E}(A, B) = \mathrm{E}(B, A)$ (although one can use (3) to show that $\mathrm{E}(A, B) = \mathrm{E}(B^{-1}, A^{-1})$). On the other hand, we do have the following important identity, which can be viewed as a weak form of commutativity.

Lemma 4.3. *For any multiplicative set A, we have $\mathrm{E}(A, A^{-1}) = \mathrm{E}(A^{-1}, A)$.*

Remark 4.4. This identity is especially striking since there is no relation between the size of $A \cdot A^{-1}$ and $A^{-1} \cdot A$ in general. For instance, if H is a multiplicative set which is also a subgroup of G, and $A := (x \cdot H) \cup H$ for some x not in the normaliser of H, then $A \cdot A^{-1}$ has about the same size as H, but $A^{-1} \cdot A$ can be much larger. In the discrete case (Example 2.3) one can prove this lemma using the identity (5) and the observation that $a \cdot b = a' \cdot b'$ if and only if $b \cdot (b')^{-1} = a^{-1} \cdot a'$.

Proof From (4), (3) (and associativity) we have

$$\mathrm{E}(A, A^{-1}) = 1_A * 1_{A^{-1}} * \widetilde{1_A * 1_{A^{-1}}}(1) = 1_A * 1_{A^{-1}} * 1_A * 1_{A^{-1}}(1).$$

Similarly we have $\mathrm{E}(A^{-1}, A) = 1_{A^{-1}} * 1_A * 1_{A^{-1}} * 1_A(1)$. The claim then follows from (2). ∎

This lemma has the following useful consequence.

Proposition 4.5. *Let A be a multiplicative set such that $\mu(A \cdot A^{-1}) \leq K\mu(A)$. Then there exists a symmetric multiplicative set S such that $\mu(S) \geq \mu(A)/2K$ and*

$$\mu(A \cdot S^n \cdot A^{-1}) \leq 2^n K^{2n+1} \mu(A) \tag{10}$$

for all integers $n \geq 1$.

Proof From Lemma 4.3 and (9) we have

$$\int_G \mu(A \cap (A \cdot x))^2 \, d\mu(x) = 1_{A^{-1}} * 1_A(x)^2 \, d\mu(x)$$
$$= \mathrm{E}(A^{-1}, A)$$
$$= \mathrm{E}(A, A^{-1})$$
$$\geq \mu(A)^4 / \mu(A \cdot A^{-1})$$
$$\geq \mu(A)^3 / K.$$

Now we define S as

$$S := \{x \in G : \mu(A \cap (A \cdot x)) > \mu(A)/2K\}.$$

It is easy to see that S is a symmetric multiplicative set. From (6) we see that

$$\int_G \mu(A \cap (A \cdot x)) \, d\mu(x) = \mu(A)^2$$

and thus

$$\int_{G \backslash S} \mu(A \cap (A \cdot x))^2 \, d\mu(x) \leq \mu(A)^3 / 2K.$$

Subtracting this from the preceding estimate, we conclude

$$\int_S \mu(A \cap (A \cdot x))^2 \, d\mu(x) \geq \mu(A)^3 / 2K.$$

Bounding $\mu(A \cap (A \cdot x))$ by $\mu(A)$, we conclude in particular that $\mu(S) \geq \mu(A)/2K$.

It remains to prove (10). Let us consider the quantity

$$\int_{(A \cdot A^{-1})^{n+1}} 1_{A \cdot S^n \cdot A^{-1}}(y_0 \ldots y_n) \, d\mu(y_0) \ldots d\mu(y_n). \tag{11}$$

On the one hand, this quantity is clearly bounded above by $\mu(A \cdot A^{-1})^{n+1} \leq K^{n+1}\mu(A)$. Now let us obtain a lower bound. We rewrite this quantity as

$$\int_{A \cdot S^n \cdot A^{-1}} \int_{(A \cdot A^{-1})^n} 1_{A \cdot A^{-1}}(y_{n-1}^{-1} \ldots y_0^{-1} x) \, d\mu(y_0) \ldots d\mu(y_{n-1}) d\mu(x).$$

Suppose that we can show that

$$\int_{(A \cdot A^{-1})^n} 1_{A \cdot A^{-1}}(y_{n-1}^{-1} \ldots y_0^{-1} x) \, d\mu(y_0) \ldots d\mu(y_{n-1}) \geq (\mu(A)/2K)^n \tag{12}$$

for all $x \in A \cdot S^n \cdot A^{-1}$ Then we can bound (11) from below by $\mu(A \cdot S^n \cdot A^{-1})(\mu(A)/2K)^n$, which will establish (10).

It remains to show (12). Let $x \in A \cdot S^n \cdot A^{-1}$ be arbitrary. We can write $x = a_0 s_1 \ldots s_n b_{n+1}^{-1}$ where $a_0, b_{n+1} \in A$ and $s_1, \ldots, s_n \in S$. If we make the successive change of variables

$$y_0 = a_0 b_1^{-1}; \quad y_1 = b_1 s_1 b_2^{-1}; \quad \ldots \quad ; y_{n-1} = b_{n-1} s_{n-1} b_n^{-1}$$

then we observe that

$$y_{n-1}^{-1} \ldots y_0^{-1} x = b_n s_n b_{n+1}^{-1}$$

and we can rewrite the left-hand side of (12) as

$$\int_{G^n} 1_{A \cdot A^{-1}}(a_0 b_1^{-1}) \prod_{i=1}^{n} 1_{A \cdot A^{-1}}(b_i s_i b_{i+1}^{-1}) \, d\mu(b_1) \ldots d\mu(b_n).$$

Note that if $b_1, \ldots, b_n \in A$ and $b_1 s_1, \ldots, b_n s_n \in A$ then the integrand here is equal to 1. Hence we can bound (12) from below by

$$\int_{G^n} \prod_{i=1}^{n} 1_A(b_i) 1_A(b_i s_i) \, d\mu(b_1) \ldots d\mu(b_n) = \prod_{i=1}^{n} \mu(A \cap (A \cdot s_i)).$$

The claim now follows from the definition of S. ∎

Now we can classify sets of small doubling using approximate groups.

Theorem 4.6. *Let $K \geq 1$, and let A, B be multiplicative sets. Then the following two statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statement holds for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

(i) *We have the product bound $\mu(A \cdot B) \lesssim K^{O(1)} \mu(A)^{1/2} \mu(B)^{1/2}$ (or equivalently, $d(A, B^{-1}) \lesssim 1 + \log K$).*

(ii) *There exists a $O(K^{O(1)})$-approximate group H of size $\mu(H) \lesssim K^{O(1)} \mu(A)^{1/2} \mu(B)^{1/2}$ and a finite set X of cardinality $O(K^{O(1)})$ such that $A \subset X \cdot H$ and $B \subset H \cdot X$.*

Proof The implication (ii) $\implies$ (i) is trivial:

$$\mu(A \cdot B) \leq \mu(X \cdot H \cdot H \cdot X) \leq |X|^2 \mu(H^2) \lesssim K^{O(1)} \mu(H) \lesssim K^{O(1)} \mu(A)^{1/2} \mu(B)^{1/2}.$$

Now we show that (i) implies (ii). From (1) we have $d(A, A) \lesssim 1 + \log K$, thus $\mu(A \cdot A^{-1}) \lesssim K^{O(1)} \mu(A)$. Applying Proposition 4.5, we obtain a symmetric multiplicative set S with $\mu(S) \gtrsim K^{O(1)} \mu(A)$ such that

$$\mu(A \cdot S^3 \cdot A^{-1}) \lesssim K^{O(1)} \mu(A). \tag{13}$$

In particular we have $\mu(S), \mu(A \cdot S) \lesssim K^{O(1)} \mu(A)$, which implies that $d(A, S) \lesssim 1 + \log K$. We also see that $\mu(S^3) \lesssim K^{O(1)} \mu(S)$, so by Theorem 3.10 we can

find a $O(K^{O(1)})$-approximate group H of size $O(K^{O(1)})\mu(A)$ which contains S. In particular we have $\mu(S \cdot H) \leq \mu(H^2) \lesssim K^{O(1)}\mu(A)$, and hence $d(S, H) \lesssim 1 + \log K$. From the triangle inequality we conclude $d(A, H) \lesssim 1 + \log K$, thus $\mu(A \cdot H) \lesssim K^{O(1)}\mu(A)$. Applying Lemma 3.6, there exists a finite set Y of cardinality $O(K^{O(1)})$ such that $A \subset Y \cdot H \cdot H$; since H is a $O(K^{O(1)})$-approximate group, we can thus find another finite set Z of cardinality $O(K^{O(1)})$ such that $A \subset Z \cdot H$. Now since $d(A, B^{-1}) \lesssim 1 + \log K$, the triangle inequality also gives $d(B^{-1}, H) \lesssim 1 + \log K$, so by arguing as before we can find a finite set W of cardinality $O(K^{O(1)})$ such that $B^{-1} \subseteq W \cdot H$. The claim now follows by taking $X := Z \cup W^{-1}$.   ∎

One can of course specialize this to theorem to the case $A = B$, to characterize sets of small doubling:

Corollary 4.7. *Let $K \geq 1$, and let A be a multiplicative set. Then the following two statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statement holds for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

  (i) *We have the product bound $\mu(A \cdot A) \lesssim K^{O(1)}\mu(A)$ (or equivalently, $d(A, A^{-1}) \lesssim 1 + \log K$).*
  (ii) *There exists a $O(K^{O(1)})$-approximate group H of size $\mu(H) \lesssim K^{O(1)}\mu(A)$ and a finite set X of cardinality $O(K^{O(1)})$ such that $A \subset (X \cdot H) \cap (H \cdot X)$.*

As one consequence of this corollary, we can obtain the following strengthening of Lemma 3.4 which was conjectured to us by Imre Ruzsa (private communication):

Corollary 4.8. *Let A be a multiplicative set such that $\mu(A \cdot a \cdot A) \leq K\mu(A)$ for all $a \in A$, and such that $\mu(A^2) \leq K\mu(A)$. Then $\mu(A^3) \lesssim K^{O(1)}\mu(A)$, and in particular the conclusions of Lemma 3.4 hold (with slightly worse implied constants).*

Proof By Corollary 4.7, we may find a $O(K^{O(1)})$-approximate group H with $\mu(H) \lesssim K^{O(1)}\mu(A)$ and a finite set X of cardinality $O(K^{O(1)})$ such that $A \subset X \cdot H, H \cdot X$. By removing useless elements of X if necessary we may assume that $X \subset (A \cdot H) \cup (H \cdot A)$. Then

$$\mu(A^3) \leq \mu(X \cdot H \cdot H \cdot X \cdot H) \lesssim K^{O(1)}\mu(H \cdot H \cdot X \cdot H).$$

But by Definition 3.8, $H \cdot H$ is covered by $O(K^{O(1)})$ left-translates of H, thus

$$\mu(A^3) \lesssim K^{O(1)}\mu(H \cdot X \cdot H) \lesssim K^{O(1)} \sup_{x \in (A \cdot H) \cup (H \cdot A)} \mu(H \cdot x \cdot H)$$

where the last inequality follows from the properties of X. Thus it suffices to show that

$$\mu(H \cdot x \cdot H) \lesssim K^{O(1)}\mu(A)$$

for all $x \in A \cdot H$ or $x \cdot H \cdot A$. Splitting x into factors in H and A and noting once again that $H \cdot H$ can be covered by $O(K^{O(1)})$ left-translates (and hence right-translates, by symmetry) of H, we reduce to showing that

$$\mu(H \cdot a \cdot H) \lesssim K^{O(1)}\mu(A) \tag{14}$$

for all $a \in A$.

Fix a. We already know that $\mu(A \cdot a \cdot A) \leq K\mu(A)$, thus

$$d(A, A^{-1} \cdot a^{-1}) \leq \log K.$$

On the other hand, since

$$\mu(A \cdot H) \leq \mu(X \cdot H \cdot H) \lesssim K^{O(1)}\mu(H)$$

and $\mu(H) \lesssim K^{O(1)}\mu(A)$ we see that

$$d(A, H) \lesssim 1 + \log K.$$

By the triangle inequality we thus have

$$d(H, A^{-1} \cdot a^{-1}) \lesssim 1 + \log K$$

or equivalently

$$d(H \cdot a, A^{-1}) \lesssim 1 + \log K.$$

Now

$$\mu(H \cdot A) \leq \mu(H \cdot H \cdot X) \lesssim K^{O(1)}\mu(H)$$

so by arguing as before we have

$$d(H, A^{-1}) \lesssim 1 + \log K$$

and thus by the triangle inequality

$$d(H \cdot a, H) \lesssim 1 + \log K$$

and the claim (14) follows. ∎

5. The Balog-Szemerédi-Gowers theorem

In this section we develop with the non-commutative, continuous analogue of Balog-Szemerédi-Gowers theory. We first give a preliminary version of this lemma, in which we start with $1/K$ of a product set $A \cdot B$ being under control, and end up with $1 - \varepsilon$ of another product set $(A') \cdot (A')^{-1}$ being under control.

Lemma 5.1 (Weak Balog-Szemerédi-Gowers theorem). *Let A, B, C be multiplicative sets such that*

$$\mu(C) \leq K'\mu(A)^{1/2}\mu(B)^{1/2}$$

and

$$\mu \otimes \mu(\{(a,b) \in A \times B : a \cdot b \in C\}) \geq \mu(A)\mu(B)/K$$

for some $K, K' \geq 1$, where $\mu \otimes \mu$ denotes product measure on $G \times G$. Let $0 < \varepsilon < 1$. Then there exists a multiplicative set A' contained in A, and a multiplicative set D, such that

$$\mu(A') \geq \frac{\mu(A)}{\sqrt{2}K} \tag{15}$$

and

$$\mu(D) \leq \frac{2(KK')^2}{\varepsilon}\mu(A) \tag{16}$$

and

$$\mu \otimes \mu(\{(a,a') \in A' \times A' : a \cdot (a')^{-1} \in D\} \geq (1-\varepsilon)\mu(A')^2. \tag{17}$$

Proof By hypothesis on C we have

$$\int_B (\int_A 1_C(a \cdot b) \, d\mu(a)) d\mu(b) \geq \mu(A)\mu(B)/K.$$

By Cauchy-Schwarz we conclude that

$$\int_B (\int_A 1_C(a \cdot b) \, d\mu(a))^2 d\mu(b) \geq \mu(A)^2 \mu(B)/K^2$$

which we rearrange as

$$\int_A \int_A (\int_B 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(b)) \, d\mu(a) d\mu(a') \geq \mu(A)^2 \mu(B)/K^2.$$

Let $\Omega \subset A \times A$ be the set of all (a, a') such that

$$\int_B 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(b) \leq \frac{\varepsilon}{2K^2} \mu(B)$$

then we clearly have

$$\int_A \int_A 1_\Omega(a, a') (\int_B 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(b)) \, d\mu(a) d\mu(a') \leq \varepsilon \mu(A)^2 \mu(B)/2K^2$$

and hence

$$\int_A \int_A (1 - \frac{1}{\varepsilon} 1_\Omega(a, a')) (\int_B 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(b)) \, d\mu(a) d\mu(a') \geq \mu(A)^2 \mu(B)/2K^2.$$

We rewrite this as

$$\int_B (\int_A \int_A (1 - \frac{1}{\varepsilon} 1_\Omega(a, a')) 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(a) d\mu(a')) d\mu(b) \geq \mu(A)^2 \mu(B)/2K^2$$

and hence by the pigeonhole principle there exists $b \in B$ such that

$$\int_A \int_A (1 - \frac{1}{\varepsilon} 1_\Omega(a, a')) 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(a) d\mu(a') \geq \mu(A)^2/2K^2.$$

If we fix this b and set $A' := \{a \in A : (a, b) \in E\}$, we conclude that

$$\mu(A')^2 \geq \int_{A'} \int_{A'} (1 - \frac{1}{\varepsilon} 1_\Omega(a, a')) \, d\mu(a) d\mu(a') \geq \mu(A)^2/2K^2 \geq 0$$

which in particular implies (15). Also we see from the above inequality that

$$\mu \otimes \mu((A' \times A') \cap \Omega) \leq \varepsilon \mu(A')^2.$$

Thus if we define

$$D := \{a \cdot (a')^{-1} : a, a' \in A'; (a, a') \notin \Omega\}$$

then we have (17). Now suppose that $d = a \cdot (a')^{-1}$ lies in D for some $a, a' \in A'$ and $(a, a') \notin \Omega$. From definition of Ω we have

$$\int_B 1_C(a \cdot b) 1_C(a' \cdot b) \, d\mu(b) > \frac{\varepsilon}{2K^2} \mu(B)$$

and hence by the substitution $c := a' \cdot b$

$$\int_G 1_C(d \cdot c) 1_C(c) \, d\mu(c) > \frac{\varepsilon}{2K^2} \mu(B).$$

Integrating this over all $d \in D$, we obtain

$$\int_G \int_G 1_C(d \cdot c) 1_C(c) \, d\mu(c) d\mu(d) > \frac{\varepsilon}{2K^2} \mu(B)\mu(D).$$

Using Fubini's theorem and making the change of variables $c' = d \cdot c$ we see that the left-hand side is just $\mu(C)^2 \leq (K')^2 \mu(A)\mu(B)$, and (16) follows. ∎

Now we extend the Balog-Szemerédi-Gowers theorem to pairs A, B of multiplicative sets (of comparable size).

Theorem 5.2 (Balog-Szemerédi-Gowers theorem). *Let A, B be multiplicative sets such that $\mathrm{E}(A, B) \geq \mu(A)^{3/2}\mu(B)^{3/2}/K$. Then there exist multiplicative sets A''', B''' contained in A, B respectively such that $\mu(A''') \geq \frac{\mu(A)}{8\sqrt{2}K}$, $\mu(B''') \geq \frac{\mu(B)}{8K}$, and $\mu(A''' \cdot B''') \lesssim K^8 \mu(A)^{1/2}\mu(B)^{1/2}$.*

Proof By hypothesis we have

$$\int_G (1_A * 1_B(x))^2 \, d\mu(x) \geq \mu(A)^{3/2}\mu(B)^{3/2}/K.$$

If we let C denote the (open, precompact) set

$$C := \{x \in G : 1_A * 1_B(x) > \mu(A)^{1/2}\mu(B)^{1/2}/2K\}$$

then we see from (6) that

$$\int_{G \setminus C} (1_A * 1_B(x))^2 \, d\mu(x) \leq \mu(A)^{3/2}\mu(B)^{3/2}/2K$$

and hence

$$\int_C (1_A * 1_B(x))^2 \, d\mu(x) \geq \mu(A)^{3/2}\mu(B)^{3/2}/2K. \tag{18}$$

In particular, C is non-empty (and is thus a multiplicative set), while from (6) and Markov's inequality we have

$$\mu(C) \leq 2K\mu(A)^{1/2}\mu(B)^{1/2}. \tag{19}$$

Also, from (18) and (7) we have

$$\int_C 1_A * 1_B(x) \, d\mu(x) \geq \mu(A)\mu(B)/2K.$$

By Fubini's theorem and a change of variables, the left-hand side can be rearranged as

$$\int_A (\int_B 1_C(ab) \, d\mu(b)) d\mu(a).$$

If we thus let A' be the (open precompact) subset of A defined by

$$A' := \{a \in A : \int_B 1_C(ab) \, d\mu(b) > \mu(B)/4K\}$$

then

$$\int_{A \setminus A'} (\int_B 1_C(ab) \, d\mu(b)) d\mu(a) \leq \mu(A)\mu(B)/4K$$

and hence

$$\int_{A'} (\int_B 1_C(ab) \, d\mu(b)) d\mu(a) \geq \mu(A)\mu(B)/4K. \tag{20}$$

In particular, A' is non-empty (and is thus a multiplicative set). Using the trivial bound $\int_B 1_C(ab) \, d\mu(b) \leq \mu(B)$, we also see that

$$\mu(A') \geq \mu(A)/4K.$$

Let us thus write $\mu(A) = L\mu(A')$ for some $1 \leq L \leq 4K$. From (19) we then have

$$\mu(C) \leq 2KL^{1/2}\mu(A')^{1/2}\mu(B)^{1/2}$$

while from (20) we have

$$\mu \otimes \mu(\{(a, b) \in A' \times B : a \cdot b \in C\}) \geq \mu(A')\mu(B)L/4K.$$

We can thus apply Lemma 5.1 (with $\varepsilon := 1/32K$, and with A, K, K' replaced by $A', 4K/L, 2KL^{1/2}$) to find a multiplicative set A'' contained in A' (and hence in A), and a multiplicative set D, such that

$$\mu(A'') \geq \frac{\mu(A')L}{4\sqrt{2}K} = \frac{\mu(A)}{4\sqrt{2}K}$$

and

$$\mu(D) \leq \frac{2(4K/L)^2(2KL^{1/2})^2}{1/32K}\mu(A') \lesssim K^5\mu(A')/L.$$

In particular, we have

$$\mu(D) \lesssim K^6\mu(A'')/L^2 \lesssim K^6\mu(A''). \tag{21}$$

Also we have

$$\mu \otimes \mu(\{(a, a') \in A'' \times A'' : a \cdot (a')^{-1} \in D\}) \geq (1 - \frac{1}{32K})\mu(A'')^2.$$

We can rewrite the latter estimate as

$$\int_{A''} \mu(\{a' \in A'' : a \cdot (a')^{-1} \notin D\}) \, d\mu(a) \leq \frac{1}{32K}\mu(A'')^2$$

so if we set

$$A''' := \{a \in A'' : \mu(\{a' \in A'' : a \cdot (a')^{-1} \notin D\}) \leq \frac{1}{16K}\mu(A'')\}$$

then by Markov's inequality we have

$$\mu(A''') \geq \mu(A'')/2 \geq \frac{\mu(A)}{2^{3.5}K}.$$

Since A'' is a subset of A', we have

$$\int_B 1_C(ab) \, d\mu(b) > \mu(B)/4K \text{ for all } a \in A''$$

and hence upon integrating in a and Fubini's theorem

$$\int_B (\int_{A''} 1_C(ab) \, d\mu(a)) d\mu(b) \geq \mu(A'')\mu(B)/4K.$$

Hence if we define the (open precompact) subset B''' of B by

$$B''' := \{b \in B : \int_{A''} 1_C(ab) \, d\mu(a) > \mu(A'')/8K\}$$

then we have by similar arguments to before that

$$\int_{B'''} (\int_{A''} 1_C(ab) \, d\mu(a)) d\mu(b) \geq \mu(A'')\mu(B)/8K;$$

since $\int_{A''} 1_C(ab) \, d\mu(a) \leq \mu(A'')$, we have in particular that

$$\mu(B''') \geq \mu(B)/8K.$$

In particular B''' is non-empty and is hence a multiplicative set.

Now let $c = ab$ for some $a \in A'''$ and $b \in B'''$. From definition of B''' we have

$$\mu(\{a' \in A'' : a'b \in C\}) > \mu(A'')/8K$$

while from definition of A''' we have

$$\mu(\{a' \in A'' : a \cdot (a')^{-1} \notin D\}) \leq \frac{1}{16K}\mu(A'').$$

Thus

$$\mu(\{a' \in A'' : a'b \in C, a \cdot (a')^{-1} \in D\}) > \mu(A'')/16K$$

so by setting $x := a'b$ (so that $a \cdot (a')^{-1} = cx^{-1}$) we have

$$\int_G 1_C(x)1_D(cx^{-1}) \, d\mu(x) > \mu(A'')/16K.$$

Integrating this over all $c \in A''' \cdot B'''$ we conclude

$$\int_G \int_G 1_C(x)1_D(cx^{-1}) \, d\mu(x)d\mu(c) \geq \mu(A'')\mu(A''' \cdot B''')/16K.$$

But the left-hand side is $\mu(C)\mu(D)$, so we have

$$\mu(A''' \cdot B''') \lesssim \frac{K\mu(C)\mu(D)}{\mu(A'')}.$$

Applying (19), (21), we conclude

$$\mu(A''' \cdot B''') \lesssim K^8 \mu(A)^{1/2}\mu(B)^{1/2}$$

as desired.   ∎

Remark 5.3. There are a number of variants of this theorem, for instance one could replace the hypothesis that $\mathrm{E}(A, B)$ is large by the hypothesis that a partial product set $A \overset{E}{\cdot} B$ is small (with a suitable largeness hypothesis on E). One can then refine the above theorem by requiring the additional conclusion that E has large intersection with $A' \times B'$; see for instance [24], [3] for some examples of this type of refinement. Other variants of the lemma and its proof can be found in [34], [10]. The power of 7 can probably be lowered further but we shall not attempt to do so here.

This gives us a characterisation of pairs of multiplicative sets of large multiplicative energy.

Theorem 5.4. *Let $K \geq 1$, and let A, B be multiplicative sets. Then the following four statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statements hold for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

  (i) *We have the energy bound $\mathrm{E}(A, B) \gtrsim K^{O(1)}\mu(A)^{3/2}\mu(B)^{3/2}$.*
  (ii) *There exists an open subset $E \subset A \times B$ of measure $\mu \otimes \mu(E) \gtrsim K^{O(1)}\mu(A)\mu(B)$ such that $\mu(\{a \cdot b : (a, b) \in E\}) \lesssim K^{O(1)}\mu(A)^{1/2}\mu(B)^{1/2}$.*

(iii) *There exists multiplicative sets A', B' contained in A, B respectively such that $\mu(A') \sim K^{O(1)}\mu(A)$, $\mu(B') \sim K^{O(1)}\mu(B)$, and $\mu(A' \cdot B') \sim K^{O(1)}\mu(A)^{1/2}\mu(B)^{1/2}$.*

(iv) *There exists a $O(K^{O(1)})$-approximate group H of size $\mu(H) \sim K^{O(1)}\mu(A)^{1/2}\mu(B)^{1/2}$, and elements $x, y \in G$ such that $\mu(A \cap (x \cdot H)) \sim K^{O(1)}\mu(A)$ and $\mu(B \cap (H \cdot y)) \sim K^{O(1)}\mu(B)$.*

Proof The implication (i) $\implies$ (iii) follows from Theorem 5.2 and the trivial bounds $\mu(A') \leq \mu(A)$, $\mu(B') \leq \mu(B)$, and $\mu(A' \cdot B') \geq \mu(A')^{1/2}\mu(B')^{1/2}$. Now we show that (iii) implies (iv). Note that the trivial bound $\mu(A' \cdot B') \geq \max(\mu(A'), \mu(B'))$ already implies that $\mu(B) \sim K^{O(1)}\mu(A)$. Using Theorem 4.6, we can find a $O(K^{O(1)})$-approximate group H of measure $\mu(H) \sim K^{O(1)}\mu(A)$ and a finite set X of cardinality $O(K^{O(1)})$ such that $A' \subseteq X \cdot H$ and $B' \subseteq H \cdot X$. By the pigeonhole principle we can thus find $x, y \in X$ such that $\mu(A' \cap (x \cdot H)) \gtrsim K^{O(1)}\mu(A)$ and $\mu(B' \cap (H \cdot y)) \gtrsim K^{O(1)}\mu(B)$. Since we have the trivial bounds $\mu(A' \cap (x \cdot H)) \leq \mu(A)$ and $\mu(B' \cap (H \cdot y)) \leq \mu(B)$, the claim (iv) follows.

Next, we show that (iv) implies (ii). If we set $E := (A \cap (x \cdot H)) \times (B \cap (H \cdot y))$, then we have the desired lower bound on $\mu \otimes \mu(E)$, and we have the upper bound

$$\mu(\{a \cdot b : (a,b) \in E\}) \leq \mu((x \cdot H) \cdot (H \cdot y)) = \mu(H^2) \sim K^{O(1)}\mu(A)^{1/2}\mu(B)^{1/2}$$

as desired.

Finaly we show that (ii) implies (i). If we let $C := \{a \cdot b : (a,b) \in E\}$ then we have

$$\int_C 1_A * 1_B(x) \ d\mu(x) \gtrsim K^{O(1)}\mu(A)\mu(B)$$

and (i) easily follows from the Cauchy-Schwarz inequality.   ■

6. Metric entropy analogues

In some applications involving non-discrete groups (e.g. Lie groups), it is not the measure or cardinality of a set which is of interest, but rather its entropy with respect to a metric.

Definition 6.1 (Metric entropy). Let X be a metric space and $\varepsilon > 0$. The *metric entropy* (or *Kolmogorov entropy*) $\mathcal{N}_\varepsilon(X)$ is defined to be the least number of open balls of radius ε needed to cover X.

Remark 6.2. There are several other formulations of metric entropy which are essentially equivalent to each other. For instance, it is easy to see that the largest ε-separated subset of X has cardinality between $\mathcal{N}_\varepsilon(X)$ and $\mathcal{N}_{\varepsilon/2}(X)$. Similarly, if X is a subspace of a larger metric space Y, one can easily check that the number of open balls of radius ε in Y needed to cover X lies between $\mathcal{N}_{2\varepsilon}(X)$ and $\mathcal{N}_\varepsilon(X)$. We shall shortly impose a volume doubling condition which will imply that $\mathcal{N}_\varepsilon(X)$ and $\mathcal{N}_{2\varepsilon}(X)$ are comparable in magnitude, and so we will not need to distinguish between these slightly different concepts of entropy.

In order for the sum set theory to extend to metric entropy, we need some mild compatibility conditions between the metric structure, the group structure, and the measure structure. We axiomatize these as follows.

Definition 6.3 (Reasonable metrics). We say that a multiplicative group $G = (G, d_G)$ equipped with a metric d_G a is *locally reasonable metric group* if the following properties hold:

(i) The topology on G is compatible with the metric d_G (thus the open balls in d_G generate the topology). Also, we assume that all closed balls are compact (thus G is locally compact).

(ii) The group operations are locally Lipschitz continuous. More precisely, for every compact set $\mathbf{K} \subseteq G$ we have the estimates

$$d_G(g \cdot x, g \cdot y), d_G(x \cdot g, y \cdot g), d(x^{-1}, y^{-1}) \sim_{\mathbf{K}, G} d(x, y)$$

for all $x, y, g \in \mathbf{K}$.

(iii) We have local volume doubling. More precisely, for any $R > 0$ we have

$$\mu(\mathbf{B}(1, 2r)) \sim_{R, G} \mu(\mathbf{B}(1, r))$$

for all $0 < r < R$, where $\mathbf{B}(x, r)$ is the open metric ball of radius r centred at x.

If the implied constants in the $\sim_{\mathbf{K}, G}$ and $\sim_{R, G}$ notation can be chosen to be independent of $\mathbf{K}$ and R (but still dependent on the group G), we say that the group is *globally reasonable*.

Examples 6.4. The Euclidean space $\mathbf{R}^d$ with the usual metric and additive group structure is globally reasonable. Any compact Lie group with a smooth Riemannian metric will also be globally reasonable. If one metric is locally (resp. globally) reasonable, then any other metric bilipschitz equivalent to it will also be locally (resp. globally) reasonable. If a locally reasonable metric group is compact, then it is automatically globally reasonable. If G is a group of linear transformations on a finite-dimensional normed vector space (with the usual topology), then the operator norm metric $d_G(x, y) := \|x - y\|_{\mathrm{op}}$ is locally reasonable (and thus globally reasonable, if G is compact). On the other hand, groups such as $SL_2(\mathbf{R})$ will not support any globally reasonable metric, due to the non-compact nature of the conjugacy classes.

As we shall shortly see, when the metric is locally reasonable, all bounded sets have finite metric entropy for each $\varepsilon > 0$. In this paper we shall be concerned with the bounded-dimensional regime, in which we allow all constants to depend on the implied constants in the $\sim_{\mathbf{K}, G}$ and $\sim_{R, G}$ notation appearing in the above definition. The issue of precise behaviour of constants on the dimension (and on other characteristics of the group) in the high-dimensional regime is an interesting one, but we will not pursue it here.

With the above assumptions on the metric, the metric entropy can be estimated accurately by the measure of various sets.

Lemma 6.5 (Multiplicative structure of balls). *Let G be a locally reasonable metric group. Let $\mathbf{K}$ be a compact subset of G, and let $R > 0$ be a compact set.*

(i) *(Approximate normality of balls) There exists constants $0 < c_{\mathbf{K},R,G} < C_{\mathbf{K},R,G} < \infty$ such that we have the inclusions*

$$X \cdot \mathbf{B}(1, c_{\mathbf{K},R,G}\varepsilon) \subseteq \mathbf{B}(1, \varepsilon) \cdot X \subseteq X \cdot \mathbf{B}(1, C_{\mathbf{K},R,G}\varepsilon)$$

and

$$\mathbf{B}(1, c_{\mathbf{K},R,G}\varepsilon) \cdot X \subseteq X \cdot \mathbf{B}(1, \varepsilon) \subseteq X \cdot \mathbf{B}(1, C_{\mathbf{K},R,G}\varepsilon)$$

for any $X \subseteq \mathbf{K}$ and $0 < \varepsilon < R$.

(ii) *(Doubling property) For any $x \in \mathbf{K}$, $0 < \varepsilon < R$, and $A > 0$, we have*

$$\mu(\mathbf{B}(x, A\varepsilon)) \sim_{A,\mathbf{K},R,G} \mu(\mathbf{B}(1, \varepsilon))$$

(iii) *(Self-covering property) For any $x \in \mathbf{K}$, $0 < \varepsilon < R$, and $A > 0$, we can cover $\mathbf{B}(x, A\varepsilon)$ by $O_{A,\mathbf{K},R,G}(1)$ balls of radius ε.*

If G is globally reasonable, then we can omit the dependence on $\mathbf{K}$ and R in the above estimates.

Proof If $x \in X \subseteq K$ and $0 < \varepsilon < R$, then $\mathbf{B}(x, \varepsilon)$ is contained in a compact set $\tilde{\mathbf{K}} = \tilde{\mathbf{K}}_{\mathbf{K},R}$ which is independent of x and ε. From the locally Lipschitz property we then have $d_G(x, y) \sim_{\mathbf{K},R,G} d_G(1, x^{-1} \cdot y) \sim_{\mathbf{K},R,G} d_G(1, y \cdot x^{-1})$ for all $x \in X$ and $y \in \mathbf{B}(x, \varepsilon)$, and the claim (i) follows.

In view of (i), we see that to prove (ii) it suffices to do so when $x = 1$. But then this follows by iterating the volume doubling property.

Finally, we prove (iii). Let S be a maximal ε-separated subset of $\mathbf{B}(x, A\varepsilon)$. Clearly the balls $\mathbf{B}(s, \varepsilon)$ with $s \in S$ cover $\mathbf{B}(x, A\varepsilon)$. Also, the balls $\mathbf{B}(s, \varepsilon/2)$ with $s \in S$ are disjoint subsets of $\mathbf{B}(x, (A + 1/2)\varepsilon)$, and thus

$$\sum_{s \in S} \mu(\mathbf{B}(s, \varepsilon/2)) \leq \mathbf{B}(x, (A + 1/2)\varepsilon).$$

Applying (ii) we obtain $|S| = O_{A,\mathbf{K},R,G}(1)$, and the claim follows. (One could also have proceeded using Lemma 3.6.)

If d_G is globally reasonable, then an inspection of the above arguments shows that the constants which depended on $\mathbf{K}$ and R are now uniform in those parameters. ∎

Lemma 6.6 (Relationship between entropy and measure). *Let G be a locally reasonable metric group. Let $\mathbf{K}$ be a compact subset of G, and let $R > 0$. Then for every $0 < \varepsilon < R$ and every $X \subseteq \mathbf{K}$ we have*

$$\mathcal{N}_\varepsilon(X) \sim_{\mathbf{K},R,G} \frac{\mu(X \cdot \mathbf{B}(1, \varepsilon))}{\mu(\mathbf{B}(1, \varepsilon))} \tag{22}$$

and

$$\mathcal{N}_\varepsilon(X) \sim_{\mathbf{K},R,G} \mathcal{N}_{2\varepsilon}(X). \tag{23}$$

In particular

$$\mu(X \cdot \mathbf{B}(1,\varepsilon)) \sim_{\mathbf{K},R,G} \mu(X \cdot \mathbf{B}(1,2\varepsilon)). \tag{24}$$

Also, the ball $\mathbf{B}(1,\varepsilon)$ is approximately normal in the sense that

$$\mu(\mathbf{B}(1,\varepsilon) \cdot X \cdot Y) \sim_{\mathbf{K},R,G} \mu(X \cdot \mathbf{B}(1,\varepsilon) \cdot Y) \sim_{\mathbf{K},R,G} \mu(X \cdot Y \cdot \mathbf{B}(1,\varepsilon))$$

$$\tag{25}$$

for any $X, Y \subseteq \mathbf{K}$. In particular we have $\mu(X \cdot \mathbf{B}(1,\varepsilon)) \sim_{\mathbf{K},R,G} \mu(\mathbf{B}(1,\varepsilon) \cdot X)$.

If G is globally reasonable, then we can replace $\sim_{\mathbf{K},R,G}$ by $\sim_G$ in the above estimates.

Remark 6.7. Note that no measurability conditions are required on X and Y, since sets such as $X \cdot \mathbf{B}(1,\varepsilon)$ are automatically open and precompact. These types of inequalities are well known for Euclidean space, and the assumptions we have placed on the metric will allow us to extend the Euclidean space arguments to this more general setting without difficulty.

Proof Fix X, ε. Let $0 < c = c_{\mathbf{K},R,G} \le 1$ be a small constant to be chosen later. Let S be a maximal $c\varepsilon$-separated subset of X, then the balls $\mathbf{B}(s, c\varepsilon/2)$ for $s \in S$ are disjoint, and the balls $\mathbf{B}(s, c\varepsilon)$ cover X. In particular $\mathcal{N}_\varepsilon(X) \le |S|$. By Lemma 6.5(i), the balls $\mathbf{B}(s, c\varepsilon/2)$ are all contained in $X \cdot \mathbf{B}(1,\varepsilon)$ if c is sufficiently small, and thus

$$\sum_{s \in S} \mu(\mathbf{B}(s, c\varepsilon/2)) \le \mu(X \cdot \mathbf{B}(1,\varepsilon)).$$

Applying Lemma 6.5(ii) we have

$$\sum_{s \in S} \mu(\mathbf{B}(s, c\varepsilon/2)) \sim_{\mathbf{K},R,G} |S|\mu(\mathbf{B}(1,\varepsilon)) \ge \mathcal{N}_\varepsilon(X)\mu(\mathbf{B}(1,\varepsilon))$$

thus obtaining the upper bound in (22).

Now we obtain the lower bound in (22). Let $\{\mathbf{B}(s,\varepsilon) : s \in S\}$ be any covering of X by ε-balls with $S \subseteq X$. Our task is to show that $\mu(X \cdot \mathbf{B}(1,\varepsilon)) = O_{\mathbf{K},R,G}(|S|\mu(\mathbf{B}(1,\varepsilon)))$. Since $X \cdot \mathbf{B}(1,\varepsilon)$ is covered by $\mathbf{B}(s,\varepsilon) \cdot \mathbf{B}(1,\varepsilon)$, it thus suffices by the union bound to establish that

$$\mu(\mathbf{B}(s,\varepsilon) \cdot \mathbf{B}(1,\varepsilon)) = O_{\mathbf{K},R,G}(\mu(\mathbf{B}(1,\varepsilon))).$$

But from Lemma 6.5(i) we have $\mathbf{B}(s,\varepsilon) \cdot \mathbf{B}(1,\varepsilon) \subseteq \mathbf{B}(s, C\varepsilon)$ for some $C = O_{\mathbf{K},R,G}(1)$, and the claim then follows from Lemma 6.5(ii).

Now we establish (23). The bound $\mathcal{N}_{2\varepsilon}(X) \le \mathcal{N}_\varepsilon(X)$ is trivial, so it suffices to establish the reverse bound $\mathcal{N}_\varepsilon(X) = O_{\mathbf{K},R,G}\mathcal{N}_{2\varepsilon}(X)$. Let $\{\mathbf{B}(s, 2\varepsilon) : s \in S\}$ be a covering of X by balls of radius 2ε for some $S \subseteq X$; we may take $|S| = \mathcal{N}_{2\varepsilon}(X)$. It will suffice to show that X can be covered by $O_{\mathbf{K},R,G}(|S|)$ balls of radius ε with centres in X. By Remark 6.2, this will follow if we can cover X by $O_{\mathbf{K},R,G}(|S|)$ balls of radius $\varepsilon/2$ whose centres do not necessarily lie in X. But this follows from Lemma 6.5(iii).

Finally, the claim (25) follows easily from Lemma 6.5(i) and (24).

If G is globally reasonable, then an inspection of the above arguments shows that the constants which depended on $\mathbf{K}$ and R are now uniform in those parameters. ■

Lemma 6.6 allows us to pass back and forth between entropies and measures, after paying various normalizing factors of $\mu(\mathbf{B}(1,\varepsilon))$. Using this lemma, one can transfer[3] most of the continuous estimates of preceding sections to entropy ones, though if the metric d_G is merely locally reasonable instead of globally reasonable, then one has to restrict the sets in question to a fixed compact region. We shall focus attention on the three main results of previous sections, namely Theorems 3.10, 4.6, 5.4. We state these results for locally reasonable metric groups, but there is an obvious variant for globally reasonable metric groups in which the dependencies of the constants on $\mathbf{K}$ and R are dropped.

Theorem 6.8. *Let G be a locally reasonable metric group. Let $\mathbf{K}$ be a compact set in G, let $R > 0$, let $0 < \varepsilon < R$, let $K \geq 1$, and let $A \subseteq \mathbf{K}$ be non-empty. Then the following three statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statements hold for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

 (i) *We have the tripling bound $\mathcal{N}_\varepsilon(A^3) \lesssim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)$.*
 (ii) *We have $\mathcal{N}_\varepsilon(A^{\epsilon_1} \ldots A^{\epsilon_n}) \sim_{\mathbf{K},R,G,n} K^{O_n(1)} \mathcal{N}_\varepsilon(A)$ for all $n \geq 1$ and all signs $\epsilon_1, \ldots, \epsilon_n \in \{-1, 1\}$.*
 (iii) *There exists a $O_{\mathbf{K},R,G}(K^{O(1)})$-approximate group H of with $\mathcal{N}_\varepsilon(H) \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)$ which contains A, and is contained in a compact set $\tilde{\mathbf{K}} = \tilde{\mathbf{K}}(\mathbf{K}, R)$ depending only on $\mathbf{K}$ and R.*

Proof Let us first prove that (iii) implies (i). By Lemma 6.5(i) we have

$$A^3 \subseteq H^3 \subseteq X \cdot X \cdot H$$

where X is the set of cardinality at most $O_{\mathbf{K},R,G}(K^{O(1)})$ associated to H. Using Lemma 6.6 we conclude that

$$\begin{aligned}
\mathcal{N}_\varepsilon(A^3) &\lesssim_{\mathbf{K},R,G} \mu(X \cdot X \cdot H \cdot \mathbf{B}(1,\varepsilon))/\mu(\mathbf{B}(1,\varepsilon)) \\
&\lesssim_{\mathbf{K},R,G} |X|^2 \mu(H \cdot \mathbf{B}(1,\varepsilon))/\mu(\mathbf{B}(1,\varepsilon)) \\
&\lesssim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(H) \\
&\sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)
\end{aligned}$$

which is (i).

A similar argument shows that (iii) implies (ii) and is left to the reader. Since (ii) trivially implies (i), it remains to show that (i) implies (iii). From the hypothesis on A and Lemma 6.6 we have

$$\mu(A^3 \cdot \mathbf{B}(1,\varepsilon)) \lesssim_{\mathbf{K},R,G} K^{O(1)} \mu(A \cdot \mathbf{B}(1,\varepsilon)).$$

---

[3]An alternate approach would be to repeat the *proofs* of the previous estimates in the metric entropy setting. That approach also works, and in fact leads to slightly better implied constants in the $O()$ notation, however the repetition of the arguments would be rather boring and we have elected instead to illustrate the transference approach.

Applying many applications of (25), (24) we conclude that

$$\mu((A \cdot \mathbf{B}(1, \varepsilon))^3) \lesssim_{\mathbf{K}, R, G} K^{O(1)} \mu(A \cdot \mathbf{B}(1, c\varepsilon)).$$

By Theorem 3.10 there exists a $O_{\mathbf{K}, R, G}(K^{O(1)})$-approximate group H which contains $A \cdot \mathbf{B}(1, c\varepsilon)$, and which obeys the estimate

$$\mu(H) \lesssim_{\mathbf{K}, R, G} K^{O(1)} \mu(A \cdot \mathbf{B}(1, \varepsilon)) \sim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A) \mu(\mathbf{B}(1, \varepsilon)).$$

From the proof of Theorem 3.10, and the hypothesis that $A \subseteq \mathbf{K}$ and $0 < \varepsilon < R$, we also see that $H \subseteq \tilde{K}$ for some compact $\tilde{K} = \tilde{K}(\mathbf{K}, R)$. Then by Lemma 6.6

$$\begin{aligned}
\mathcal{N}_\varepsilon(H) &\lesssim_{\mathbf{K}, R, G} \mu(\mathbf{B}(1, \varepsilon) \cdot H)/\mu(\mathbf{B}(1, \varepsilon)) \\
&\leq \mu(H \cdot H)/\mu(\mathbf{B}(1, \varepsilon)) \\
&\lesssim_{\mathbf{K}, R, G} K^{O(1)} \mu(H)//\mu(\mathbf{B}(1, \varepsilon)) \\
&\lesssim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A)
\end{aligned}$$

and (iii) follows. ∎

Now we give the metric entropy analogue of Theorem 4.6.

Theorem 6.9. *Let G be a locally reasonable metric group. Let $\mathbf{K}$ be a compact set in G, let $0 < \varepsilon < R$, let $K \geq 1$, and let $A, B \subseteq \mathbf{K}$ be non-empty. Then the following two statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statement holds for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

- (i) *We have the product bound $\mathcal{N}_\varepsilon(A \cdot B) \lesssim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}$.*
- (ii) *There exists a $O_{\mathbf{K}, R, G}(K^{O(1)})$-approximate group H with $\mathcal{N}_\varepsilon(H) \lesssim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}$ and a finite set X of cardinality $O_{\mathbf{K}, R, G}(K^{O(1)})$ such that $A \subset X \cdot H$ and $B \subset H \cdot X$. Furthermore, H and X lie in a compact set $\tilde{K} = \tilde{K}(\mathbf{K}, R)$ depending only on $\mathbf{K}$ and R.*

Proof First we show that (ii) implies (i). We have a set $Y \subset H^2$ of cardinality $O_{\mathbf{K}, R, G}(K^{O(1)})$ such that $H \cdot H \subseteq H \cdot Y$, which implies that $A \cdot B \subseteq X \cdot H \cdot Y \cdot X$. From Lemma 6.6 we then have

$$\begin{aligned}
\mathcal{N}_\varepsilon(A \cdot B) &\leq \mathcal{N}_\varepsilon(X \cdot H \cdot Y \cdot X) \\
&\lesssim_{\mathbf{K}, R, G} \mu(X \cdot H \cdot Y \cdot X \cdot \mathbf{B}(1, \varepsilon))/\mu(\mathbf{B}(1, \varepsilon)) \\
&\lesssim_{\mathbf{K}, R, G} |X|^2 |Y| \mu(H \cdot \mathbf{B}(1, \varepsilon))/\mu(\mathbf{B}(1, \varepsilon)) \\
&\lesssim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(H) \\
&\lesssim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}
\end{aligned}$$

which is (i).

Now we show that (i) implies (ii). From the hypothesis and Lemma 6.6 we have

$$\mu(A \cdot \mathbf{B}(1,\varepsilon) \cdot B \cdot \mathbf{B}(1,\varepsilon)) \lesssim_{\mathbf{K},R,G} \mu(A \cdot B \cdot \mathbf{B}(1,\varepsilon))$$
$$\lesssim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(A \cdot B)\mu(\mathbf{B}(1,\varepsilon))$$
$$\lesssim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}\mu(\mathbf{B}(1,\varepsilon))$$
$$\sim_{\mathbf{K},R,G} K^{O(1)}\mu(A \cdot \mathbf{B}(1,\varepsilon))^{1/2}\mu(B \cdot \mathbf{B}(1,\varepsilon))^{1/2}.$$

Applying Theorem 4.6, we can find a $O_{\mathbf{K},R,G}(K^{O(1)})$-approximate group H and a set X such that $A \cdot \mathbf{B}(1,\varepsilon) \subset X \cdot H$ and $B \cdot \mathbf{B}(1,\varepsilon) \subset H \cdot X$, with the bounds

$$\mu(H) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mu(A \cdot \mathbf{B}(1,\varepsilon))^{1/2}\mu(B \cdot \mathbf{B}(1,\varepsilon))^{1/2}$$
$$\lesssim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}\mu(\mathbf{B}(1,\varepsilon)).$$

and $|X| \lesssim_{\mathbf{K},R,G} K^{O(1)}$. We then have

$$\mathcal{N}_\varepsilon(H) \lesssim_{\mathbf{K},R,G} \mu(\mathbf{B}(1,\varepsilon) \cdot H)/\mu(\mathbf{B}(1,\varepsilon))$$
$$\leq \mu(A \cdot \mathbf{B}(1,\varepsilon) \cdot H)/\mu(\mathbf{B}(1,\varepsilon))$$
$$\leq \mu(X \cdot H \cdot H)/\mu(\mathbf{B}(1,\varepsilon))$$
$$\lesssim_{\mathbf{K},R,G} |X|K^{O(1)}\mu(H)/\mu(\mathbf{B}(1,\varepsilon))$$
$$\lesssim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}$$

and (ii) follows. $\blacksquare$

Now we turn to developing a metric entropy analogue of Theorem 5.4. This will be a bit trickier as we shall need an "ε-approximate" version of the multiplicative energy $\mathrm{E}(A,B)$. There are a number of essentially equivalent ways to do so, each of which are at least somewhat artificial; for sake of concreteness we shall fix one such as follows. Given any $A, B \subset G$ and $\varepsilon > 0$, the set

$$Q_\varepsilon(A,B) := \{(a,b,a',b') \in A \times B \times A \times B : d(a \cdot b, a' \cdot b') \leq \varepsilon\}$$

of approximately multiplicative quadruples is a subset of G^4, which we view as a metric space with the metric

$$d_{G^4}((x_1,x_2,x_3,x_4),(y_1,y_2,y_3,y_4)) := \sum_{i=1}^4 d_G(x_i,y_i).$$

We then define the ε-approximate multiplicative energy $\mathrm{E}_\varepsilon(A,B)$ to be the quantity $\mathcal{N}_\varepsilon(Q_\varepsilon(A,B))$. Note that if A, B are finite sets, then this quantity will equal the usual (discrete) multiplicative energy (5) for ε sufficiently small.

Theorem 6.10. *Let G be a locally reasonable metric group. Let $\mathbf{K}$ be a compact set in G, let $0 < \varepsilon < R$, let $K \geq 1$, and let $A, B \subseteq \mathbf{K}$ be non-empty. Then the following four statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statement holds for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

(i) *We have the energy bound $\mathrm{E}_\varepsilon(A,B) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{3/2}\mathcal{N}_\varepsilon(B)^{3/2}$.*
(ii) *There exists a subset $E \subset A \times B$ of entropy $\mathcal{N}_\varepsilon(E) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)\mathcal{N}_\varepsilon(B)$ such that $\mathcal{N}_\varepsilon(\{a \cdot b : (a,b) \in E\}) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}$. (Of course, we measure the entropy of E using the product metric on G^2.)*

(iii) *There exists subsets A', B' of A, B respectively such that $\mathcal{N}_\varepsilon(A') \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)$, $\mathcal{N}_\varepsilon(B') \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(B)$, and $\mathcal{N}_\varepsilon(A' \cdot B') \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}$.*

(iv) *There exists a $O(K^{O(1)})$-approximate group H with $\mathcal{N}_\varepsilon(H) \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}$, and elements $x, y \in G$ such that $\mathcal{N}_\varepsilon(A \cap (x \cdot H)) \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)$ and $\mathcal{N}_\varepsilon(B \cap (H \cdot y)) \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(B)$. Furthermore, H, x, y lie in a compact set $\tilde{\mathbf{K}} = \mathbf{K}(\mathbf{K}, R)$ that depends only on $\mathbf{K}$ and R.*

Proof Let us first show that (iv) implies (iii). We set $A' := A \cap (x \cdot H)$ and $B' := B \cap (H \cdot y)$. The bounds on $\mathcal{N}_\varepsilon(A')$ and $\mathcal{N}_\varepsilon(B')$ are obvious. From Lemma 6.6 and the trivial estimate $\mu(X \cdot Y) \geq \mu(X)^{1/2} \mu(Y)^{1/2}$ we see that $\mathcal{N}_\varepsilon(A' \cdot B') \gtrsim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(A')^{1/2} \mathcal{N}_\varepsilon(B')^{1/2}$, which gives the lower bound on $\mathcal{N}_\varepsilon(A' \cdot B')$. To obtain the upper bound, we use Lemma 6.6 to compute

$$\mathcal{N}_\varepsilon(A' \cdot B') \leq \mathcal{N}_\varepsilon(x \cdot H \cdot H \cdot y)$$
$$\lesssim_{\mathbf{K},R,G} \mu(x \cdot H \cdot H \cdot y \cdot \mathbf{B}(1,\varepsilon))/\mu(\mathbf{B}(1,\varepsilon))$$
$$\lesssim_{\mathbf{K},R,G} \mu(\mathbf{B}(1,\varepsilon) \cdot H \cdot H)/\mu(\mathbf{B}(1,\varepsilon)).$$

But $H \cdot H$ is covered by $O_{\mathbf{K},R,G}(K^{O(1)})$ right-translates of H, and so

$$\mathcal{N}_\varepsilon(A' \cdot B') \lesssim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(H) \sim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}$$

as desired.

Now we show that (iii) implies (ii). We take $E := A' \times B'$. The bound on $\mathcal{N}_\varepsilon(\{a \cdot b : (a,b) \in E\})$ is obvious, while by considering products of ε-separated sets it is easy to establish a bound of the form

$$\mathcal{N}_{\varepsilon/100}(E) \gtrsim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(A') \mathcal{N}_\varepsilon(B') \gtrsim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A) \mathcal{N}_\varepsilon(B).$$

The claim then follows from (23) (note that if G is locally reasonable then so is G^2).

Now we show that (ii) implies (i). Let E' be a maximal 100ε-separated subset of E, then by (23)

$$|E'| \geq \mathcal{N}_{100\varepsilon}(E) \sim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(E) \gtrsim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A) \mathcal{N}_\varepsilon(B).$$

Let D be a maximal $\varepsilon/2$-separated subset of $\{a \cdot b : (a,b) \in E\}$, thus

$$|D| \leq \mathcal{N}_{\varepsilon/4}(\{a \cdot b : (a,b) \in E\}) \sim_{\mathbf{K},R,G} \mathcal{N}_\varepsilon(\{a \cdot b : (a,b) \in E\}) \lesssim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{1/2} \mathcal{N}_\varepsilon(B)^{1/2}.$$

Observe that for every $(a,b) \in E'$, the product $a \cdot b$ lies within $c\varepsilon$ of an element of D, thus

$$\sum_{x \in D} |\{(a,b) \in E' : d_G(a \cdot b, x) \leq \varepsilon/2\}| \geq |E'| \gtrsim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A) \mathcal{N}_\varepsilon(B).$$

Applying Cauchy-Schwarz we conclude that

$$\sum_{x \in D} |\{(a,b) \in E' : d_G(a \cdot b, x) \leq \varepsilon/2\}|^2 \gtrsim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{3/2} \mathcal{N}_\varepsilon(B)^{3/2}.$$

Observe that if $(a,b), (a',b') \in E'$ and $x \in D$ are such that $d_G(a \cdot b, x), d_G(a' \cdot b', x) \leq \varepsilon/2$ then $(a, b, a', b') \in Q_\varepsilon(A, B)$. Thus

$$|Q_\varepsilon(A, B) \cap (E' \times E')| \gtrsim_{\mathbf{K},R,G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{3/2} \mathcal{N}_\varepsilon(B)^{3/2}.$$

But $E' \times E'$ is clearly ε-separated, thus

$$\mathrm{E}_\varepsilon(A, B) = \mathcal{N}_\varepsilon(Q_\varepsilon(A, B)) \gtrsim_{\mathbf{K}, R, G} K^{O(1)} \mathcal{N}_\varepsilon(A)^{3/2} \mathcal{N}_\varepsilon(B)^{3/2}$$

as desired.

Finally, we show that (i) implies (iv), which is the most difficult implication. Let $C = C_{\mathbf{K}, R, G}$ be a large constant to be chosen later. Let $\overline{A} := A \cdot \mathbf{B}(1, C\varepsilon)$ and $\overline{B} := B \cdot \mathbf{B}(1, C\varepsilon)$, thus from Lemma 6.6 we see that $\overline{A}$, $\overline{B}$ are multiplicative sets with

$$\mu(\overline{A}) \sim_{\mathbf{K}, R, G, C} \mathcal{N}_\varepsilon(A)\mu(\mathbf{B}(1, \varepsilon)); \quad \mu(\overline{B}) \sim_{\mathbf{K}, R, G, C} \mathcal{N}_\varepsilon(B)\mu(\mathbf{B}(1, \varepsilon)). \tag{26}$$

Now consider the quantity $\mathrm{E}(\overline{A}, \overline{B})$. We can rewrite this as

$$\begin{aligned}
\mathrm{E}(\overline{A}, \overline{B}) &= \int_G (1_{\overline{A}} * 1_{\overline{B}})(x)^2 \, d\mu(x) \\
&= \int_{\overline{A}} \int_{\overline{B}} 1_{\overline{A}} * 1_{\overline{B}}(a \cdot b) \, d\mu(a)d\mu(b) \\
&= \int_{\overline{A}} \int_{\overline{B}} \int_{\overline{A}} 1_{\overline{B}}((a')^{-1} \cdot a \cdot b) \, d\mu(a')d\mu(a)d\mu(b).
\end{aligned}$$

Now observe that if C is large enough, we see that for any $x \in G$, the set $B \cdot \mathbf{B}(1, \sqrt{C}\varepsilon)$ intersects $\mathbf{B}(x, \sqrt{C}\varepsilon)$ only when $x \in \overline{B}$. This (and Lemma 6.5(ii)) leads to the pointwise estimate

$$1_{\overline{B}}(x) \gtrsim_{\mathbf{K}, R, G} \frac{1}{\mu(\mathbf{B}(1, \varepsilon))} \int_{B \cdot \mathbf{B}(1, \sqrt{C}\varepsilon)} 1_{\mathbf{B}(x, \sqrt{C}\varepsilon)}(b)d\mu(b)$$

and hence

$$\mathrm{E}(\overline{A}, \overline{B}) \gtrsim_{\mathbf{K}, R, G} \frac{1}{\mu(\mathbf{B}(1, \varepsilon))} \int_{\overline{A}} \int_{\overline{B}} \int_{\overline{A}} \int_{B \cdot \mathbf{B}(1, \sqrt{C}\varepsilon)} 1_{\mathbf{B}((a')^{-1} \cdot a \cdot b, \sqrt{C}\varepsilon)}(b') \, d\mu(b')d\mu(a')d\mu(a)d\mu(b).$$

Now observe (from the local Lipschitz property) that if $(a, b, a', b') \in Q_\varepsilon(A, B) \cdot B_{G^4}(1, \varepsilon)$, where $B_{G^4}(1, \varepsilon)$ denotes the ball in G^4, then $d(a \cdot b, a' \cdot b') \lesssim_{\mathbf{K}, R, G} \varepsilon$, and hence (if C is large enough) $b' \in (a')^{-1} \cdot a \cdot b, \sqrt{C}\varepsilon)$. Thus we have

$$\mathrm{E}(\overline{A}, \overline{B}) \gtrsim_{\mathbf{K}, R, G} \frac{1}{\mu(\mathbf{B}(1, \varepsilon))} \mu^{\otimes 4}(Q_\varepsilon(A, B) \cdot B_{G^4}(1, \varepsilon))$$

and hence (by the analogue[4] of Lemma 6.6 for G^4)

$$\mathrm{E}(\overline{A}, \overline{B}) \gtrsim_{\mathbf{K}, R, G} \frac{\mu^{\otimes 4}(B_{G^4}(1, \varepsilon))}{\mu(\mathbf{B}(1, \varepsilon))} \mathcal{N}_\varepsilon(Q_\varepsilon(A, B)) \gtrsim_{\mathbf{K}, R, G} \mu(\mathbf{B}(1, \varepsilon))^3 \mathrm{E}_\varepsilon(A, B).$$

We henceforth fix C to be a suitably large quantity depending on $\mathbf{K}, R, G$, and thus can omit the dependence of C in the estimates which follow. By hypothesis on $\mathrm{E}_\varepsilon(A, B)$ and (26), we thus have

$$\mathrm{E}(\overline{A}, \overline{B}) \gtrsim_{\mathbf{K}, R, G} K^{O(1)} \mu(\overline{A})^{3/2} \mu(\overline{B})^{3/2}.$$

---

[4]Here we need the easily verified fact that the direct product of finitely many locally reasonable metric groups is still locally reasonable.

Applying Proposition 5.4 we can thus locate a $O_{\mathbf{K},R,G}(K^{O(1)})$-approximate group H of size

$$\mu(H) \sim_{\mathbf{K},R,G} K^{O(1)}\mu(\overline{A})^{1/2}\mu(\overline{B})^{1/2}$$

and elements $x, y \in G$ such that $\mu(\overline{A} \cap (x \cdot H)) \sim_{\mathbf{K},R,G} K^{O(1)}\mu(\overline{A})$ and $\mu(\overline{B} \cap (H \cdot y)) \sim_{\mathbf{K},R,G} K^{O(1)}\mu(\overline{B})$. An inspection of the proof of Proposition 5.4 reveals that H, x, y are also contained in a compact set that depends only on $\mathbf{K}$ and R. Note that from the trivial bounds $\mu(\overline{A} \cap (x \cdot H)) \leq \mu(\overline{A})$ and $\mu(\overline{B} \cap (H \cdot y)) \leq \mu(\overline{B})$ we can conclude that $\overline{A}$ and $\overline{B}$ are comparable in size:

$$\mu(\overline{A}) \sim_{\mathbf{K},R,G} K^{O(1)}\mu(\overline{B}).$$

From (26) we thus have entropy comparability also:

$$\mathcal{N}_\varepsilon(A) \sim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(B). \tag{27}$$

From (26) again, we have a good bound on the measure of H:

$$\mu(H) \sim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}\mu(\mathbf{B}(1,\varepsilon)). \tag{28}$$

However to get a good bound on the *entropy* of H we need to estimate $\mu(H \cdot \mathbf{B}(1,\varepsilon))$. This we shall do by means of the Ruzsa triangle inequality. Observe that

$$\mu([\overline{A} \cap (x \cdot H)] \cdot H) \leq \mu(x \cdot H \cdot H) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mu(H)$$

since H is an approximate group. From our bounds on $\mu(H)$ and $\mu(\overline{A} \cap (x \cdot H))$ we conclude that

$$d(\overline{A} \cap (x \cdot H), H^{-1}) \leq O(\log K) + O_{\mathbf{K},R,G}(1).$$

Next, we observe that

$$\mu([\overline{A} \cap (x \cdot H)] \cdot \mathbf{B}(1,\varepsilon)) \leq \mu(\overline{A} \cdot \mathbf{B}(1,\varepsilon)) = \mu(A \cdot \mathbf{B}(1,C\varepsilon) \cdot \mathbf{B}(1,\varepsilon))$$

and so from Lemma 6.6 we have

$$\mu([\overline{A} \cap (x \cdot H)] \cdot \mathbf{B}(1,\varepsilon)) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)\mu(\mathbf{B}(1,\varepsilon)).$$

This gives a bound on the Ruzsa distance:

$$d(\overline{A} \cap (x \cdot H), \mathbf{B}(1,\varepsilon)^{-1}) \leq \log \mathcal{N}_\varepsilon(A) + O(\log K) + O_{\mathbf{K},R,G}(1).$$

By the triangle inequality we conclude that

$$d(H^{-1}, \mathbf{B}(1,\varepsilon)^{-1}) \leq \log \mathcal{N}_\varepsilon(A) + O(\log K) + O_{\mathbf{K},R,G}(1)$$

which implies that

$$\mu(H \cdot \mathbf{B}(1,\varepsilon)) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)\mu(\mathbf{B}(1,\varepsilon)). \tag{29}$$

Combining this with Lemma 6.6, (28), (27), and the trivial lower bound $\mu(H \cdot \mathbf{B}(1,\varepsilon)) \geq \mu(H)$ we conclude that

$$\mathcal{N}_\varepsilon(H) \sim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)^{1/2}\mathcal{N}_\varepsilon(B)^{1/2}.$$

We are nearly done with establishing (iv), but there is one slight problem: we have shown that $\overline{A}$ and $\overline{B}$ have large intersection (in the measure sense) with translates of H, but we need A and B to have large intersection (in the entropy sense) with translates of H. There are a number of ways to resolve this; one is as follows. Observe that

$$\mathcal{N}_\varepsilon(\overline{A} \cap (x \cdot H)) \gtrsim_{\mathbf{K},R,G} \mu(\overline{A} \cap (x \cdot H))/\mu(\mathbf{B}(1,\varepsilon)) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A).$$

Thus there exists a ε-separated subset $\overline{A'}$ of $\overline{A} \cap (x \cdot H)$ of cardinality $\gtrsim_{\mathbf{K},R,G}$ $K^{O(1)}\mathcal{N}_\varepsilon(A)$. Using (23) one can refine this subset to be $C\varepsilon$-separated for any fixed $C = O_{\mathbf{K},R,G}(1)$ without degrading the cardinality of $\overline{A'}$ significantly. By construction of $\overline{A'}$, each element of $\overline{A'}$ is at a distance $O_{\mathbf{K},R,G}(\varepsilon)$ to an element of A. This shows that there exists an ε-separated subset A' of A of cardinality $\gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)$, with each element of A' at a distance $O_{\mathbf{K},R,G}(\varepsilon)$ to an element of $x \cdot H$. If we then set $\tilde{H} := H \cdot \mathbf{B}(1, C\varepsilon)$ for some sufficiently large $C = O_{\mathbf{K},R,G}(1)$, we see that A' is contained in $x \cdot \tilde{H}$ and so $\mathcal{N}_\varepsilon(A \cap (x \cdot \tilde{H})) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)$. A similar argument yields $\mathcal{N}_\varepsilon((y \cdot \tilde{H}) \cap B) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(B)$. Now we need to pass from $\tilde{H}$ back to H. Recall that as H is an approximate group, $H \cdot H$ can be covered by $O_{\mathbf{K},R,G}(K^{O(1)})$ left-translates of H, hence $H \cdot \tilde{H}$ can be covered by $O_{\mathbf{K},R,G}(K^{O(1)})$ left-translates of $\tilde{H}$. Combining this with (28), (29), (24) we have

$$\mu(H \cdot \tilde{H}) \lesssim_{\mathbf{K},R,G} K^{O(1)}\mu(H).$$

Applying Lemma 3.6 we can thus cover $\tilde{H}$ by $O_{\mathbf{K},R,G}(K^{O(1)})$ left-translates of H. In particular we can cover $x \cdot \tilde{H}$ by $O_{\mathbf{K},R,G}(K^{O(1)})$ sets of the form $x' \cdot H$, and so by the pigeonhole principle we have $\mathcal{N}_\varepsilon(A \cap (x' \cdot H)) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(A)$ for some x', which one can easily verify is contained in a compact set $\tilde{\mathbf{K}}(\mathbf{K}, R)$ depending only on $\mathbf{K}$ and R. A similar argument (using (25) to move the $\mathbf{B}(1, \varepsilon)$ factors around as necessary) gives $\mathcal{N}_\varepsilon(B \cap (H \cdot y')) \gtrsim_{\mathbf{K},R,G} K^{O(1)}\mathcal{N}_\varepsilon(B)$ for some $y' \in \tilde{\mathbf{K}}(\mathbf{K}, R)$, and (iv) follows.   ∎

One can of course develop metric entropy analogues of many of the other estimates from previous sections (such as the Ruzsa triangle inequality). We leave the details to the reader.

7. Inverse theorems

The above theory reduces the study of sets of small doubling or tripling (or pairs of sets with small product set, small partial product set, or large multiplicative energy) to that of studying approximate groups, at least if one is prepared to lose polynomial factors in the constants and (in the locally reasonable metric entropy setting) one restricts all sets to a compact region. There remains of course the question of how to effectively classify these approximate groups; we refer to this as the *inverse product set problem* (or the *inverse sum set problem*, in the abelian additive setting). At present, there is not even a reasonable conjecture as to what such objects should look like; there are obvious examples of approximate groups, such as genuine groups, geometric progressions[5], and (given sufficient commutativity) the direct sum of other approximate groups, but it is not clear in general what the statement should be[6].

---

[5]In the abelian case, the group G is usually written additively, and it is then the *arithmetic* progressions which are relevant here. However as we are considering the non-commutative setting we are forced to depart from the usual additive notation and work instead with geometric progressions.

[6]Indeed, the problem can be viewed as a robust version of the problem of classifying all the subgroups of a given group G, which is already quite a difficult problem, especially for highly non-abelian groups such as the permutation group S_n. In some cases it seems that the class of

There are however a number of special cases which are well understood. If G is a discrete abelian r-torsion group for some small $r > 1$ (thus $x^r = 1$ for all r) and A is a finite non-empty subset of G then it is known that $|A \cdot A| = O(|A|)$ if and only if A can be contained in a finite subgroup H of G of size $O(|A|)$; see [32]. If G is instead a discrete abelian torsion-free group, and A is a finite non-empty subset of G, then a famous theorem of Freiman [16] (see also [2], [31], [9]) shows that $|A \cdot A| = O(|A|)$ if and only if A is contained in the product P of $O(1)$ geometric progressions, whose total cardinality is $O(|A|)$. These results were unified in [20], in which A was now a finite non-empty subset of an arbitrary abelian group G, and the result now being that $|A \cdot A| = O(|A|)$ if and only if A is contained in the product P of $O(1)$ geometric progressions and a finite subgroup of G, whose total cardinality is $O(|A|)$. See [36] for a presentation of all of these abelian results. Apart from the (important) issue of quantifying the dependence of constants here, this is a satisfactory resolution of the inverse sum set problem in the discrete setting.

In the abelian setting it is also easy to pass to the continuous setting and the metric entropy setting. For instance, we have

Proposition 7.1 (Continuous version of Freiman's theorem). *Let $d \geq 1$. Let A be an open bounded non-empty subset of $\mathbf{R}^d$ such that $\mu(A + A) \leq K\mu(A)$ for some $K \geq 2^d$, where μ denotes Lebesgue measure. Then there exists an $\varepsilon > 0$ and a set P which is the sum of $O_K(1)$ arithmetic progressions in $\mathbf{R}^d$ such that $A \subseteq P + \mathbf{B}(0, \varepsilon)$ and $\mu(P + \mathbf{B}(0, \varepsilon)) \sim_K \mu(A)$.*

Remark 7.2. Note that the trivial inclusion $A + A \supset 2 \cdot A$ (or the Brunn-Minkowski inequality) shows that K cannot be less than 2^d. In the converse direction, it is easy to see that if $A \subseteq P + \mathbf{B}(0, \varepsilon)$ and $\mu(P + \mathbf{B}(0, \varepsilon)) \sim_K \mu(A)$, where P is the sum of $O(1)$ arithmetic progressions, then $\mu(A + A) \sim_K \mu(A)$. One can certainly use the arguments in [2], [31], [21] to quantify the exact dependence on K in the above proposition but we will not attempt to do so here. It is also not difficult to modify the above proposition to replace the Euclidean space $\mathbf{R}^d$ with a torus such as $\mathbf{R}^d/\mathbf{Z}^d$ by a lifting argument; we omit the details.

Proof Since $A + A$ is open, we have $A + A = \bigcap_{\varepsilon > 0} A + A + \mathbf{B}(0, \varepsilon)$. By the monotone convergence theorem, we can thus find an $\varepsilon > 0$ such that $\mu(A + A + B(0, \varepsilon)) \sim \mu(A + A) \sim_K \mu(A)$.

Now let $\tilde{A} := (A + B(0, \varepsilon/2)) \cap (\frac{\varepsilon}{10d} \cdot \mathbf{Z}^d)$, thus A is a finite non-empty set. From the inclusions

$$A \subseteq \tilde{A} + B(0, \varepsilon) \text{ and } \tilde{A} + \tilde{A} + B(0, \frac{\varepsilon}{100d}) \subseteq A + A + B(0, \varepsilon)$$

one easily verifies the estimate

$$\mu(A) \lesssim_K |\tilde{A}|\varepsilon^d \leq |\tilde{A} + \tilde{A}|\varepsilon^d \lesssim_K \mu(A).$$

Note that any dependencies on d of the implied constant can be converted to a dependency on K since $K \geq 2^d$. In particular we have $|\tilde{A} + \tilde{A}| \sim_K |\tilde{A}|$. Applying

---

approximate subgroups of G is not much "richer" the class of genuine subgroups of G, in the sense that one can express approximate subgroups as dense subsets of combinations of genuine subgroups and related objects such as geometric progressions, but this might not be true for sufficiently complicated groups G.

Freiman's theorem (see e.g. [16], [2], [31], [9], [28], [36]) we can thus place $\tilde{A}$ inside a set $P \subset \mathbf{R}^d$ of cardinality $|P| \sim_K |\tilde{A}|$ which is the sum of $O_K(1)$ arithmetic progressions. Since $A \subseteq \tilde{A} + B(0, \varepsilon) \subseteq P + B(0, \varepsilon)$, we have

$$\mu(A) \le \mu(P + B(0, \varepsilon)) \lesssim_d |P|\varepsilon^d \sim_K |\tilde{A}|\varepsilon^d \sim_K \mu(A)$$

and the claim follows. ∎

Proposition 7.3 (Entropy version of Freiman's theorem). *Let $d \ge 1$ and $\varepsilon > 0$. Let A be a bounded non-empty subset of $\mathbf{R}^d$ such that $\mathcal{N}_\varepsilon(A + A) \le K\mathcal{N}_\varepsilon(A)$ for some $K \ge 1$. Then there exists a set P which is the sum of $O_{K,d}(1)$ arithmetic progressions in $\mathbf{R}^d$ such that $A \subseteq P + B(0, \varepsilon)$ and $|P| \sim_{K,d} \mathcal{N}_\varepsilon(A)$.*

Remark 7.4. The commentary in Remark 7.2 also applies in this setting. For instance, if $A \subseteq P + B(0, \varepsilon)$ and $|P| \sim_{K,d} \mathcal{N}_\varepsilon(A)$ and P is the sum of $O_{K,d}(1)$ arithmetic progressions then it is easy to see that $\mathcal{N}_\varepsilon(A + A) \sim_{K,d} \mathcal{N}_\varepsilon(A)$.

Proof Again, we take $\tilde{A} := (A + B(0, \varepsilon/2)) \cap (\frac{\varepsilon}{10d} \cdot \mathbf{Z}^d)$. From Lemma 6.6 (and the global reasonableness of $\mathbf{R}^d$) we have $|\tilde{A}| \sim_d \mathcal{N}_\varepsilon(A)$ and $|\tilde{A} + \tilde{A}| \lesssim_d \mathcal{N}_\varepsilon(A + A)$, and thus $|\tilde{A} + \tilde{A}| \sim_{K,d} |\tilde{A}|$. We now argue as in the proof of Proposition 7.1. ∎

We now turn to the noncommutative setting. Here our understanding is only satisfactory for a few special noncommutative groups; in the general case it is not even clear what the correct statement of an inverse product setting theorem should be, let alone how to prove it. We shall restrict our attention to the discrete setting, in other words in understanding those finite non-empty sets A for which $|A \cdot A| = O(|A|)$ (or $|A \cdot A \cdot A| = O(|A|)$), or for classifying finite non-empty $O(1)$-approximate groups; in view of the preceding results it seems likely that the transferral of the discrete results to a continuous or metric entropy setting will not be too difficult.

Inverse product set theorems for groups of affine or projective mappings on the real or complex line or projective line, and hence to groups such as $SL_2(\mathbf{R})$ or $SL_2(\mathbf{C})$, were studied in [11, 12, 13, 14, 15]. A typical result here is that if $A \subset SL_2(\mathbf{C})$ is a finite non-empty set such that $|A \cdot A^{-1}| = O(|A|)$, then A is contained inside $O(1)$ left-cosets of an abelian subgroup of $SL_2(\mathbf{C})$.

The case $G = SL_2(\mathbf{Z}/p\mathbf{Z})$, with p a large prime, was studied in [22]. In particular it was shown that if $A \subset G$ had size $p^\varepsilon \le |A| \le p^{-\varepsilon}|G|$ for some $\varepsilon > 0$, and A was not contained in any proper subgroup of G, then one had the tripling estimate $|A \cdot A \cdot A| \ge p^\delta |A|$ for some $\delta = \delta(\varepsilon) > 0$. Thus the only sets of small tripling are those sets which are very small, very large, or are contained in a proper subgroup (e.g. a geometric progression containing the identity). Using the machinery in this section one can also obtain a classification of sets of small doubling, which we leave as an exercise to the reader.

Another interesting example arises in the work of Lindenstrauss [25], in which G is now the *lamplighter group* $\mathbf{Z} \times (\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$ with group law $(i, a) \cdot (j, b) = (i+j, \sigma^j a + b)$, where σ is the standard shift on $(\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}}$. There it was shown that the group G

contains no Følner sequence of sets of small doubling constant, despite G being amenable (and solvable).

The case of very small doubling, e.g. $|A \cdot A| \leq 2|A|$, was treated in [26], [8], [23] in the torsion-free non-commutative case. In this case one has $|A \cdot A| \geq 2|A| - 1$, with equality only holding when A is a geometric progression.

We were unable to say anything new about the inverse product set problem for general groups. However for discrete groups G which have a normal subgroup H, it turns out that one can exploit the short exact sequence

$$\{1\} \to H \to G \to G/H \to \{1\}$$

to split the inverse product set problem for G into the inverse product set problem for H and G/H separately, together with the problem[7] of classifying a certain type of "approximate group homomorphism" from an approximate subgroup of G/H into G. To motivate matters, let us first see how *genuine* subgroups of G (as opposed to approximate groups) split under this short exact sequence.

Lemma 7.5 (Splitting lemma, group case). *Let H be a normal subgroup of a group G, and let $A \subset G$. Let $\pi : G \to G/H$ be the canonical projection. Then the following are equivalent.*

(i) *A is a subgroup of G.*
(ii) *There exists a subgroup B of H, a subgroup C of G/H, and a partial inverse $\phi : C \to G$ to π (i.e. $\pi(\phi(x)) = x$ for all $x \in C$) with the property*

$$\phi(x)B = B\phi(x) \text{ for all } x \in C \tag{30}$$

(thus ϕ takes values in the normaliser of B) and the quotiented homomorphism property

$$\phi(xy) \in \phi(x)\phi(y)B \text{ for all } x, y \in C \tag{31}$$

and such that A has the representation

$$A = \bigcup_{x \in C} \phi(x)B. \tag{32}$$

In particular, $|A| = |C||B|$.

Remark 7.6. One way to view this lemma is to think of G as a principal H-bundle over G/H. Then A is a principal B-bundle over C that takes values in the normaliser of the structure group B, and which collapses to a group homomorphism from C to G when quotiented out by B.

Proof Let us first verify that (ii) implies (i). It is easy to verify from (30), (31) that $\phi(0) \in B$ and $\phi(x^{-1}) \in \phi(x)B$ for all $x \in C$; from these facts and (30), (31), (32) we quickly see that A contains the identity, that $A^{-1} = A$, and that $A \cdot A = A$; in other words, A is a subgroup of G.

---

[7]This problem seems to be somewhat difficult, however it does appear to be fractionally simpler than the original inverse product set problem on G, so the reduction is not entirely trivial. It is somewhat analogous to the reduction of the (open) "polynomial Freiman-Ruzsa conjecture" to a conjecture concerning approximate homomorphisms in [19].

Now let us verify that (i) implies (ii). Let $C := \pi(A)$ and $B := A \cap K$; it is easy to see that B and C are subgroups of K and C/K respectively. Let $\phi : C \to A$ be an arbitrary partial inverse to the map $\pi : A \to C$ (which exists thanks to the axiom of choice), then we have (32). Since $A \cdot A = A$ we conclude (31). To verify (30), observe that $\phi(x)B\phi(x)^{-1}$ lies in $A \cdot A \cdot A^{-1} = A$, but also lies in the normal group K, and must therefore lie in $A \cap K = B$. This shows the inclusion $\phi(x)B \subseteq B\phi(x)$, and the other inclusion is proven similarly. ∎

We now present an analogue of the above lemma for approximate groups.

Lemma 7.7 (Splitting lemma, approximate group case)**.** *Let H be a normal subgroup of a discrete multiplicative group G, and let $\pi : G \to G/H$ be the canonical homomorphism. Note that H and G/H are also discrete multiplicative groups. Let $A \subset G$, and let $K \geq 1$. Then the following three statements are equivalent, in the sense that if one of them holds for one choice of implied constant in the $O()$ and $\lesssim$ notation, then the other statement holds for a different choice of implied constant in the $O()$ and $\lesssim$ notation:*

(i) *We have $|A \cdot A \cdot A| \lesssim K^{O(1)}|A|$.*
(ii) *There exists a $O(K^{O(1)})$-approximate group $\tilde{A}$ of size $|\tilde{A}| \sim K^{O(1)}|A|$ which contains A.*
(iii) *There exist $O(K^{O(1)})$-approximate groups $B_1 \subseteq B_2 \subseteq B_3 \subset H$ and $C \subset G/H$ with*

$$|B_3| \lesssim K^{O(1)}|B_1|, \tag{33}$$

together with a partial inverse $\phi : C^3 \to G$ to π, with $\phi(1) = 1$ and $\phi(x^{-1}) = \phi(x)^{-1}$ for all $x \in C^3$, such that

$$\phi(x)B_i \subseteq B_{i+1}\phi(x); \quad B_i\phi(x) \subseteq \phi(x)B_{i+1} \quad \text{for all } x \in C, i = 1, 2 \tag{34}$$

and

$$\phi(x)\phi(y)\phi(z) \in \phi(xyz)B_3 \text{ for all } x, y, z \in C \tag{35}$$

with the containment

$$A \subseteq \bigcup_{x \in C} \phi(x)B_1 \tag{36}$$

and the cardinality bound

$$|A| \gtrsim K^{-O(1)}|B_1||C|. \tag{37}$$

Remark 7.8. One can extend this lemma to the continuous setting provided that the topology of H is well-behaved (e.g. the projection map π should be continuous and open, and in particular H should be closed) and one can "disintegrate" the measure μ on G into the measures on H-cosets of G, integrated against the measures on G/H; this can for instance be done if G is a finite-dimensional Lie group and H is a closed Lie subgroup. We omit the details. There is likely to also be an entropy analogue of this lemma under reasonable assumptions on the metric but we will not describe these here. The approximate homomorphism ϕ is only defined on C^3, but an inspection of the proof below shows that one could in fact extend it to C^n for any fixed n, and have a sequence $B_1 \subseteq \ldots \subseteq B_n$ of nested approximate groups

of comparable size, with suitable modifications to (33), (34), (35); again, we omit the details.

Proof The equivalence of (i) and (ii) follows from Theorem 3.10. Now let us see that (iii) implies (i). From (36) we have

$$A \cdot A \cdot A \subseteq \bigcup_{x,y,z \in C} \phi(x) B_1 \phi(x_2) B_1 \phi(x_3) B_1.$$

From repeated application of (34) we have

$$\phi(x) B_1 \phi(y) B_1 \phi(z) B_1 \subseteq \phi(x) \phi(y) \phi(z) B_3 B_2 B_1$$

and hence by (35)

$$A \cdot A \cdot A \subseteq \bigcup_{x,y,z \in C} \phi(xyz) B_3 B_3 B_2 B_1 \subseteq \bigcup_{w \in C^3} \phi(w) B_3^4$$

and thus $|A^3| \leq |C^3||B_3^4|$. The claim now follows from (37), (33), and the hypothesis that C and B_3 are $O(K^{O(1)})$-approximate groups.

It remains to show that (ii) implies (iii). By replacing A by $\tilde{A}$ if necessary we may assume that A is itself a $O(K^{O(1)})$-approximate group; in particular, A is symmetric, contains 1, and (by Theorem 3.10) we have $|A^n| \lesssim_n K^{O_n(1)}|A|$ for all $n \geq 1$. Since π is a homomorphism, we easily see that $\pi(A)$ is also a $O(K^{O(1)})$-approximate group. Thus we shall set $C := \pi(A)$. Now we construct the B_i by the formulae

$$B_1 := (A^2 \cap H)^3; \quad B_2 := (A^8 \cap H)^3; \quad B_3 := (A^{26} \cap H)^3.$$

Observe that if $a, a' \in A$ lie in the same fiber of π (i.e. in the same coset of H), then $a' \in (A^2 \cap H)a$. Since A intersects exactly $|C|$ fibers of π, we conclude that $|A| \leq |C||A^2 \cap H|$. On the other hand, observe that if $a \in A$, then the set $(A^{2n} \cap H)a$ lies in A^{2n+1}, and is also contained in the same fiber of π as a. This implies that $|A^{2n+1}| \geq |C||A^{2n} \cap H|$ for all $n \geq 1$. Since $|A^{2n+1}| \lesssim_n K^{O_n(1)}|A|$, we conclude that $|A^{2n} \cap H| \sim_n K^{O_n(1)}|A^2 \cap H|$ for all $n \geq 1$. Since $(A^{2n} \cap H)^3 \subseteq A^{6n} \cap H$, we conclude that $A^{2n} \cap H$ has a tripling constant of $O_n(K^{O_n(1)})$. Applying Corollary 3.11 (observing that $A^{2n} \cap H$ is symmetric and contains 1) we thus see that B_1, B_2, B_3 are all $O(K^{O(1)})$-approximate groups. Note that the estimates established here also give (37) and (33).

Since $C = \pi(A)$, we have $C^3 = \pi(A^3)$. Using the axiom of choice (which is actually unnecessary here, since A^3 and C^3 are finite sets), we may select a partial inverse $\phi : C^3 \to A^3$ to π, which takes values in C on A; since C and A are both symmetric and contain the origin, there is no difficulty requiring $\phi(1) = 1$ and $\phi(x^{-1}) = \phi(x)^{-1}$ for all $x \in C^3$. If a and $\phi(x)$ lie in the same fiber of π, then as mentioned before we have $a \in \phi(x)(A^2 \cap H)$, which implies (36).

Now observe that if $x \in C$, then $\phi(x)(A^{2n} \cap H)^3 \phi(x)^{-1}$ lies in $A^{6n+2} \cap H$, and hence $\phi(x) B_i \phi(x)^{-1}$ lies in B_{i+1} for $i = 1, 2$. This proves the inclusions $\phi(x) B_i \subseteq \phi(x) B_{i+1}$; the reverse inclusions in (34) are proven similarly.

Finally, for $x, y, z \in C$, we observe that $\phi(x)\phi(y)\phi(z)\phi(xyz)^{-1} \in A^4 \cap H$, and so (35) follows. This concludes the implication of (iii) from (ii). $\blacksquare$

In principle, this splitting lemma should allow one to deduce inverse product set estimates for nilpotent (and perhaps even solvable) groups from the abelian theory. To do this in full generality appears to be rather difficult however, and we shall only demonstrate the situation with a particularly simple nilpotent group, namely a Heisenberg group.

Definition 7.9 (Heisenberg group). Let Z, W be additive abelian groups, and let $\{,\} : Z \times Z \to W$ be an antisymmetric mapping (thus $\{x, y\} = -\{y, x\}$) which is a homomorphism in each of the two variables separately (thus $\{x + y, z\} = \{x, z\} + \{y, z\}$ and $\{x, y + z\} = \{x, y\} + \{x, z\}$). We define the *Heisenberg group* associated to this antisymmetric mapping to be the set $G := Z \times W$ endowed with the group law

$$(z, w) \cdot (z', w') := (z + z', w + w' + \{z, z'\}).$$

One easily verifies that G is a discrete multiplicative group with the "vertical" group $\{0\} \times W$ (which we identify with W) as a normal subgroup (indeed, it lies in the centre of G) and with identity $1_G = (0, 0)$ and inverse $(z, w)^{-1} = (-z, -w)$; the quotient G/W is canonically identified with the "horizontal" group Z. Thus we have the short exact sequence

$$0 \to W \to G \to Z \to 0.$$

Since W and Z are abelian, we thus see that G is a 2-step nilpotent group.

We write $Z \times W$ for the additive group which is the product of Z and W, and $\iota : G \to Z \times W$ for the identity map from G to $Z \times W$. We caution that while ι is a bijection, it is *not* a group homomorphism from the multiplicative group G to the additive group $Z \times W$. Nevertheless, ι does relate the subgroups of G with the subgroups of $Z \times W$ (except for a "2-torsion" issue) as follows. Let $\pi : Z \times W \to Z$ be the canonical projection, and for any $A, B \subset Z$ let $\{A, B\} \subset W \equiv \{0\} \times W$ denote the set $\{A, B\} := \{\{a, b\} : a \in A, b \in B\}$. We also let $\langle \{A, B\} \rangle$ denote the subgroup of $\{0\} \times W$ generated by $\{A, B\}$. Finally, given any $A \subset Z \times W$ we write $2 \cdot A := \{2x : x \in A\} = \{(z + z, w + w) : (z, w) \in A\}$.

Proposition 7.10 (Subgroups of the Heisenberg group). *Let G be a Heisenberg group arising from an antisymmetric mapping $\{,\} : Z \times Z \to W$, and let $A \subset G$ be a multiplicative subgroup of G. Then there exists an additive subgroup $\tilde{A}$ of $Z \times W$ such that*

$$2 \cdot (\tilde{A} + \langle \{\pi(\tilde{A}), \pi(\tilde{A})\} \rangle) \subseteq \iota(A) \subseteq \tilde{A} + \langle \{\pi(\tilde{A}), \pi(\tilde{A})\} \rangle \tag{38}$$

Remark 7.11. In the converse direction, it is easy to verify that given any additive subgroup $\tilde{A}$ of $Z \times W$, that the set $\iota^{-1}(\tilde{A} + \langle \{\pi(\tilde{A}), \pi(\tilde{A})\} \rangle)$ is a multiplicative subgroup of G. Thus the above proposition classifies multiplicative subgroups of G in terms of additive subgroups of $Z \times W$, except for the "2-torsion" issue of having to distinguish a set the additive group $A' := \tilde{A} + \langle \{\pi(\tilde{A}), \pi(\tilde{A})\} \rangle$ from its dilate $2 \cdot A'$. If $Z \times W$ is finitely generated, then the quotient of the additive group A' by $2 \cdot A'$ will be bounded, and so in some sense the above classification of subgroups of G only "loses" a bounded amount of information.

Proof First observe that if (z, w) and (z', w') are in A, then

$$(z, w) \cdot (z', w') \cdot (z, w)^{-1} \cdot (z', w')^{-1} = (0, 2\{z, z'\}).$$

Thus if we set $C := \pi(\iota(A)) \subset Z$, we see that $2 \cdot \{C, C\} \subset A$, and hence (since the Heisenberg group law is additive on $\{0\} \times W$) we also have $2 \cdot \langle\{C, C\}\rangle \subset A$. If we now set $\tilde{A} := \iota(A) + \langle\{C, C\}\rangle$, we easily verify that $\tilde{A}$ is indeed an additive group and obeys the inclusions (38). ∎

We now extend this proposition to approximate groups, though to deal with the 2-torsion issue we shall need to make an additional assumption on the "vertical" group W.

Theorem 7.12 (Approximate subgroups of the Heisenberg group). *Let G be a Heisenberg group arising from an antisymmetric mapping $\{, \} : Z \times Z \to W$ such that W has no 2-torsion (thus if $w \in W$ and $2w = 0$, then $w = 0$), and let $A \subset G$ be a finite nonempty subset of G such that $|A \cdot A \cdot A| \leq K|A|$ for some $K \geq 1$. Then there exists a $O(K^{O(1)})$-approximate additive subgroup $\tilde{A}$ of $Z \times W$, such that*

$$\{\pi(\tilde{A}), \pi(\tilde{A})\} \subseteq \tilde{A} \tag{39}$$

and

$$\iota(A) \subseteq \tilde{A}.$$

Furthermore we have

$$|A| \gtrsim K^{-O(1)}|\tilde{A}|.$$

Remark 7.13. In the converse direction, if $\tilde{A}$ obeys all the above properties, then a comparison of the multiplicative group law for G and the additive group law for $Z \times W$ reveals that

$$\iota(A \cdot A \cdot A) \subseteq 3\tilde{A} + 3\{\pi(\tilde{A}), \pi(\tilde{A})\} \subseteq 6\tilde{A}$$

and hence by the approximate group properties of $\tilde{A}$

$$|A \cdot A \cdot A| \leq 6|\tilde{A}| \lesssim K^{O(1)}|\tilde{A}| \lesssim K^{O(1)}|A|.$$

Thus we have a sharp characterisation of the sets of small tripling in the Heisenberg group G, in the case when no 2-torsion is present in the centre. In principle, one can make this characterisation more explicit by using a version of Freiman's theorem (such as the one in [20]) in the abelian group $Z \times W$ to classify $\tilde{A}$, and then to work with the concrete description of $\tilde{A}$ given by that theorem to determine which approximate groups $\tilde{A}$ obey the constraint (39). Of course once one characterises sets of small tripling, one can use the results of earlier sections to characterise sets of small doubling, or of with small partial product set, etc. These fully explicit descriptions are however rather lengthy to state and we will leave them to the reader.

Proof We apply Lemma 7.7 with $H := \{0\} \times W \equiv W$ to obtain $O(K^{O(1)})$-approximate groups[8] $B_1 \subseteq B_2 \subseteq B_3 \subset W$ and $C \subset Z$, and a partial inverse $\phi : C^3 \to G$ to the projection map $\pi : G \to Z$ with $\phi(0) = (0, 0)$ and $\phi(-x) = -\phi(x)$

---

[8]It is a somewhat unfortunate circumstance that we will be regarding Z and W both as additive groups and multiplicative group (with the same group operation); thus for instance if $B \subset W$ then $B + B = B \cdot B$. We hope the reader will not be unduly confused by this.

that obeys (33), (35), (36), (37). (Because W is abelian, the containments (34) become trivial and will not be needed here.) Since ϕ is a partial inverse to π, we may write $\phi(z) = (z, f(z))$ for some odd function $f : C^3 \to W$. Thus for instance (36) becomes the assertion that $w \in f(z) + B_1$ for all $(z, w) \in A$. From (35) (setting the third element of C to be the identity) we see that for all $z_1, z_2 \in C$ we have

$$(z_1, f(z_1)) \cdot (z_2, f(z_2)) \in (z_1 + z_2, f(z_1 + z_2)) \cdot B_3$$

and hence (expanding out the group multiplication law in coordinates)

$$f(z_1) + f(z_2) + \{z_1, z_2\} \in f(z_1 + z_2) + B_3.$$

Swapping z_1 and z_2 and then subtracting, we see that

$$2 \cdot \{z_1, z_2\} \in B_3 - B_3 \text{ for all } z_1, z_2 \in C.$$

Let $\tilde{B} := (B_3 - B_3) \cap (2 \cdot W)$. The set $2 \cdot W$ is a subgroup of the abelian group W, and so $\tilde{B} + \tilde{B} + \tilde{B} = (3B_3 - 3B_3) \cap (2 \cdot W)$. Since B_3 is a $O(K^{O(1)})$-approximate group, we can cover $3B_3 - 3B_3$ by $O(K^{O(1)})$ translates of B_3, and hence $\tilde{B} + \tilde{B} + \tilde{B}$ can be covered by $O(K^{O(1)})$ translates of B_3, intersected with $2 \cdot W$. Any one of these translates is itself contained in a translate of $(B_3 - B_3) \cap (2 \cdot W)$, and so we see that $|3\tilde{B}| \lesssim K^{O(1)} |\tilde{B}|$. Applying Lemma 3.11 we see that $3\tilde{B}$ is a $O(K^{O(1)})$-approximate group. If we then let

$$B' := \{b \in W : 2 \cdot b \in 3\tilde{B}\}$$

we conclude (since the map $b \mapsto 2 \cdot b$ is a group isomorphism between W and $2 \cdot W$) that B' is also a $O(K^{O(1)})$-approximate group. By construction we have

$$\{C, C\} \subseteq B'. \tag{40}$$

Next, observe that

$$|B' + 3\tilde{B}| = |B' + 2 \cdot B'| \le |3B'| \lesssim K^{O(1)} |B'| = K^{O(1)} |3\tilde{B}|$$

and hence by Lemma 3.6, we can cover B' by $O(K^{O(1)})$ translates of $3\tilde{B} - 3\tilde{B}$, which is contained in $12B_3$. Since B_3 is itself a $O(K^{O(1)})$-approximate group, we can conclude that

$$|nB' + mB_3| \lesssim_{n,m} K^{O_{n,m}(1)} |B_3| \tag{41}$$

for any $n, m \ge 1$. To use this, define the set $A' \subset Z \times W$ by

$$A' := \{(z, w) : z \in C, w \in f(z) + 9B' + B_1\}.$$

From (36) we see that A' contains A. Also we have

$$3\iota(A') = \bigcup_{z_1, z_2, z_3 \in C} (z_1, f(z_1) + 9B' + B_1) + (z_2, f(z_2) + 9B' + B_1) + (z_3, f(z_3) + 9B' + B_1)$$

$$\subseteq \bigcup_{z_1, z_2, z_3 \in C} (z_1 + z_2 + z_3, f(z_1) + f(z_2) + f(z_3) + 27B' + 3B_3).$$

On the other hand, from (35) we have

$$f(z_1) + f(z_2) + f(z_3) + \{z_1, z_2\} + \{z_1, z_3\} + \{z_2, z_3\} \in f(z_1 + z_2 + z_3) + B_3$$

and hence by (40)

$$f(z_1) + f(z_2) + f(z_3) \in f(z_1 + z_2 + z_3) + B_3 + 3B'.$$

Thus we have

$$3\iota(A') \subseteq \bigcup_{z_1,z_2,z_3 \in C} (z_1 + z_2 + z_3, f(z_1 + z_2 + z_3) + 30B' + 4B_3)$$

and hence

$$|3\iota(A')| \leq |C^3||30B' + 4B_3|.$$

Since C is a $O(K^{O(1)})$-approximate group, we thus see from (41) and (37) that

$$|3\iota(A')| \lesssim K^{O(1)}|A|.$$

Since $|\iota(A')| \geq |A|$, we thus see from Lemma 3.11 that the set $\tilde{A} := 3(\iota(A') \cup 0 \cup -\iota(A'))$ is a $K^{O(1)}$-approximate group with

$$|\tilde{A}| \sim K^{O(1)}|A|.$$

Finally, we see from construction that $\pi(\tilde{A}) \subseteq 3C$, and hence $\{\pi(\tilde{A}), \pi(\tilde{A})\} \subseteq 9\{C, C\} \subseteq 9B'$ by (40); since $f(0) = 0$, we obtain (39) as desired. ∎

References

[1] A. Balog, E. Szemerédi, *A statistical theorem of set addition*, Combinatorica, **14** (1994), 263–268.

[2] Y. Bilu, *Structure of sets with small sumset*, Structure theory of set addition. Astrisque No. 258, (1999), xi, 77–108.

[3] J. Bourgain, *On the dimension of Kakeya sets and related maximal inequalities*, Geom. Func. Anal. **9** (1999), 256–282.

[4] J. Bourgain, *Estimates on exponential sums related to the Diffie-Hellman distributions*, Geom. Funct. Anal. **15** (2005), no. 1, 1–34.

[5] J. Bourgain, *Mordell's exponential sum estimate revisited*, J. Amer. Math. Soc. **18** (2005), no. 2, 477–4993.

[6] J. Bourgain, N. Katz, T. Tao, *A sum-product estimate in finite fields, and applications*, Geom. Func. Anal. **14** (2004), 27–57.

[7] J. Bourgain, S. Konyagin, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, C. R. Acad. Sci. Paris, Ser. I **337** (2003), 75–80.

[8] L.V. Brailovsky, G. A. Freiman, *On a product of finite subsets in a torsion-free group*, J. Algebra **130** (1990), 462–476.

[9] M. Chang,*A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.

[10] M. C. Chang, *On problems of Erdős and Rudin*, J. Funct. Anal. **207** (2004), 444–460.

[11] G. Elekes, *On linear combinatorics I*, Combinatorica **17** (1997), 447–458.

[12] G. Elekes, *On linear combinatorics II*, Combinatorica **18** (1998), 13–25.

[13] G. Elekes, *On linear combinatorics III*, Combinatorica **19** (1999), 43–53.

[14] G. Elekes, Z. Király, *On combinatorics of projective mappings*, J. Alg. Combin. **14** (2001), 183–197.

[15] G. Elekes, I. Ruzsa, *The structure of sets with few sums along a graph*, preprint.

[16] G. Freiman, Foundations of a structural theory of set addition. Translated from the Russian. *Translations of Mathematical Monographs,* Vol 37. American Mathematical Society, Providence, R. I., 1973. vii+108 pp.

[17] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), 529–551.

[18] T. Gowers, *A new proof of Szemeredi's theorem*, Geom. Func. Anal. **11** (2001), 465-588.

[19] B. Green, *Finite field models in arithmetic combinatorics*, preprint.

[20] B. Green, I. Ruzsa, *Freiman's theorem in an arbitrary abelian group*, preprint.

[21] B. Green, T. Tao, *Compressions, Convex Geometry and the Freiman-Bilu Theorem*, preprint.

[22] H. Helfgott, *Growth and generation in $SL_2(Z/pZ)$*, preprint.

[23] Y. Hamidoune, A.S. Lladó, O. Serra, *On subsets with small product in torsion-free groups*, Combinatorica **18** (1998), 529–540.

[24] M. Laczkovich, I. Ruzsa, *The number of homothetic subsets*, in The Mathematics of Paul Erdős, Graham and Nešetřil eds. Springer, 1996.

[25] E. Lindenstrauss, *Pointwise theorems for amenable groups*, Invent. Math. **146** (2001), no. 2, 259–295.

[26] J.H.B. Kemperman, *On complexes in a semigroup*, Indag. Math. **18** (1956), 247–254.

[27] V. Milman, *Entropy and asymptotic geometry of non-symmetric convex bodies*, Adv. in Math. **152** (2000), 314–335.

[28] M. Nathanson, *Additive number theory. Inverse problems and the geometry of sumsets*, Graduate Texts in Mathematics 165, Springer-Verlag, New York, 1996.

[29] H. Plünnecke, *Eigenschaften un Abschätzungen von Wirkingsfunktionen*, BMwF-GMD-22 Gesellschaft für Mathematik und Datenverarbeitung, Bonn 1969.

[30] I. Ruzsa, *Sums of finite sets*, Number Theory: New York Seminar; Springer-Verlag (1996), D.V. Chudnovsky, G.V. Chudnovsky and M.B. Nathanson editors.

[31] I. Ruzsa, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. 65 (1994), no. 4, 379–388.

[32] I. Ruzsa, *An analog of Freiman's theorem in groups*, Structure theory of set addition, Astérisque No. 258 (1999), 323–326.

[33] I. Ruzsa, S. Turjányi, *A note on additive bases of integers*, Publ. Math. Debrecen **32** (1985), 101–104.

[34] B. Sudakov, E. Szemerédi, V. Vu, *On a question of Erdős and Moser*, Duke Math. J. **129** (2005), no. 1, 129–155.

[35] T. Tao, *Non-commutative sum set estimates*, unpublished.

[36] T. Tao and V. Vu, *Additive Combinatorics,* book in preparation.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: tao@math.ucla.edu