The Structure of Locally Compact Approximate Groups

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Pietro Kreitlon Carolino

2015
ABSTRACT OF THE DISSERTATION

The Structure of Locally Compact Approximate Groups

by

Pietro Kreitlon Carolino

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2015

Professor Terence Chi-Shien Tao, Chair

A subset $A$ of a locally compact group $G$ is said to be a $K$-approximate group if it contains the identity, is closed under inverses, and if the product set $A \cdot A = \{ab \mid a, b \in A\}$ can be covered by $K$ left translates of $A$. In this work we prove a structure theorem on open precompact $K$-approximate groups that describes them as a combination of compact subgroups, convex sets in Lie algebras of dimension bounded in terms of $K$, and certain generalizations of arithmetic progressions. Our results can be naturally regarded as generalizations of a theorem of Breuillard, Green, and Tao that describes finite approximate groups, and also a theorem of Gleason and Yamabe, which describes locally compact groups as inverse limits of Lie groups.

Along the way we prove continuous analogues of additive combinatorial results of Sanders, Croot, and Sisask, and model-theoretic results of Hrushovski, all of which were first obtained in the finite context.
The dissertation of Pietro Kreitlon Carolino is approved.

Amit Sahai
Igor Pak
Matthias J. Aschenbrenner
Terence Chi-Shien Tao, Committee Chair

University of California, Los Angeles
2015
# Table of Contents

1 Results and Context .................................................. 1
   1.1 Introduction ..................................................... 1
   1.2 Before the Lie Model Theorem .................................. 6
   1.3 The Lie Model Theorem .......................................... 10
   1.4 Proof Strategy and Statement of Results ...................... 13

2 Examples and Non-Examples ......................................... 16

3 Some Special Cases .................................................. 19

4 Definitions and Basic Facts ........................................ 22

5 Sanders-Croot-Sisask Theory ....................................... 26
   5.1 Finding Almost-Periods ......................................... 26
   5.2 Finding Small Neighborhoods ................................... 32

6 Finding a Lie Model .................................................. 40
   6.1 Building a Metric ................................................. 40
   6.2 Local Compactness ............................................... 44
   6.3 Local Compactness Gives Lie Model ............................ 46

7 Strong Approximate Groups and Good Escape Norms ................. 53
   7.1 Extracting Strong Approximate Groups ......................... 53
   7.2 Strong Approximate Groups Have Good Escape Norm .......... 57

8 Reduction to Lie Case in Bounded Dimension ....................... 70
8.1 Good Escape Norm Reduces to Lie Case .......................... 70
8.2 A Continuous Escape Norm ........................................... 72
8.3 A Dimension Bound ...................................................... 77

9 A “Finite Plus Convex” Decomposition ............................... 83

References ........................................................................ 88
CHAPTER 1

Results and Context

1.1 Introduction

Group theory is an old and well-studied branch of mathematics. Much more recently, however, there has been substantial interest in objects that are in some sense close to, but are not quite, groups. In particular, the focus has been on controlled ways to weaken the requirement of closure under the group operation. The following definition is at the root of all the various versions of approximate closure that appear in the literature:

**Definition 1.1.** If $G$ is a group and $A, B$ are subsets of $G$, the product set of $A$ and $B$ is

$$A \cdot B = \{ab \mid a \in A, \ b \in B\}.$$ 

The multiplication symbol is often omitted. Iterated product sets of a set with itself are denoted with exponents, e.g. $A \cdot A \cdot A$ is written $A^3$. If the group is written additively we call the product set a sum set and denote it by $A + B$, and iterated product sets as in $A + A + A = 3A$.

If $A$ is a subset of a group $G$ containing the identity and closed under inverses, then $A$ is a subgroup if and only if $A \cdot A = A$. Otherwise, $A \cdot A$ strictly contains $A$, and one may investigate “how much larger” $A \cdot A$ is than $A$; we regard $A$ as approximately closed under multiplication when $A \cdot A$ is “not much larger” than $A$ in some appropriate sense. We present three notions from the literature that attempt to measure this discrepancy via a number $K$; but in this work we shall mostly be concerned with the third.
Definition 1.2. Let $G$ be a locally compact group with Haar measure $\mu$ and $A$ a subset of $G$ of finite measure. We say that $A$ has \textit{doubling} at most $K$ if $\mu(A \cdot A) \leq K \mu(A)$.

In the case where $G$ is a metric group, one can look at how $A \cdot A$ and $A$ differ in metric entropy. Recall that, if $X$ is a metric space, then for $\varepsilon > 0$ the metric entropy $h(X, \varepsilon)$ is the least number of open balls of radius $\varepsilon$ needed to cover $X$.

Definition 1.3. Let $G$ be a metric group and $A$ a subset of $G$ of finite metric entropy. We say that $A$ has \textit{entropy doubling} at most $K$ if $h(A \cdot A, \varepsilon) \leq K h(A, \varepsilon)$ for all $\varepsilon > 0$.

Finally, one may ask for a purely group-theoretic notion; such is provided by requiring there be group elements that “witness” the measure- and entropy-theoretic bounds.

Definition 1.4. Let $G$ be a group and $A$ a subset of $G$. We say that $A$ has \textit{cover doubling} at most $K$ if $A \cdot A$ can be covered by $K$ left-translates of $A$, and also by $K$ right-translates of $A$. In other words, if there are finite sets $X, Y \subseteq G$, with $K$ elements each, such that $A \cdot A \subseteq X \cdot A$ and $A \cdot A \subseteq A \cdot Y$.

If $A$ is a subgroup of $G$, then it has doubling, entropy doubling, and cover doubling all equal to 1; the intuition, then, is that the higher $K$ is, the further $A$ is from a subgroup.

Cover doubling is arguably the most powerful notion. It not only implies doubling, and entropy doubling when the metric is left- or right-invariant, but it is the only easily iterable one, in the following sense. If $A^2$ can be covered by $K$ left-translates of $A$, it is easy to see that $A^n$ can be covered by $K^{n-1}$ left-translates of $A$, and likewise for right-translates. On the other hand, directly analogous results for doubling and entropy doubling are not true in general, and even in the abelian case, where they do hold, they are rather difficult to obtain \cite{15, 17}.

For these, and other reasons presented in Chapter 4, the following definition has been settled on for the structural investigation of these objects that are close to, but are not quite, groups:
**Definition 1.5** (Approximate groups, preliminary). Let $G$ be a group, $A \subseteq G$ a subset, and $K \geq 1$ a parameter. We say $A$ is a $K$-approximate group if: (i) $1 \in A$; (ii) $A$ is closed under inverses; and (iii) $A$ has cover doubling at most $K$.

**Remark.** This definition is “preliminary” because for the vast majority of this work we will impose some topological regularity conditions on $A$. But they need not concern us for now.

In most contexts of interest (e.g. for $A$ finite, or $A$ open precompact), it is the case that every set $A$ has cover doubling $K$ for some $K$ depending on $A$. Therefore, to avoid triviality, research has focused on the regime where $K$ is independent of $A$, the question being: if $A$ is a $K$-approximate group, can we describe $A$ in terms of $K$ and more explicit group-theoretic objects? In other words, we fix $K$ and imagine that $A$ “grows” in some appropriate sense.

As a consequence, examples of approximate groups are usually families of subsets of groups, where members of the family are all $K$-approximate groups with the same $K$. In this regard approximate groups are like expander graphs: every graph is an expander for some expansion constant, and the interest is in finding infinite families with a uniform (lower bound on the) expansion constant.

With this in mind, examples of approximate groups include: genuine subgroups ($K = 1$); arithmetic progressions centered at 0 in the integers ($K = 2$); symmetric convex sets in $\mathbb{R}^d$ containing the origin ($K \leq 5^d$); homomorphic images and inverse images, and products of approximate groups; and “large” subsets of approximate groups (in the sense that a few translates of the subset suffice to cover the approximate group). We will look at examples in more detail in Chapter 2, but for now a bold conjecture might be: every approximate group can be obtained from the basic examples of genuine subgroups, arithmetic progressions, and convex sets, via the operations of product, homomorphic image and inverse image, and passing to large subsets. This conjecture isn’t true, but it comes remarkably close.

The structure of finite sets of very small doubling — less than 2, say — can be elucidated by simple enough arguments that the problem has even appeared in mathematical olympiads. In the torsion-free abelian case, one can handle doubling up to almost 3 by elementary means.
However, these approaches completely break down even for doubling as low as 10. Thus it may be surprising, even in the finite abelian case, that it is possible to say anything at all about $K$-approximate groups when $K$ is a large number, like 1000. This surprise is to some extent justified; it is not known how to handle 10-approximate groups by methods less involved than those required for general $K$.

Nevertheless, it turns out to be possible to describe approximate groups in a fair amount of detail. After work of several authors [2, 3, 4, 6, 20, 21] characterizing finite approximate groups in various special cases — notably within abelian groups, and groups of Lie type — a general result was proved by Breuillard, Green, and Tao [7]. To state it succinctly we set down some notation.

**Definition 1.6 (Asymptotic notation).** Following analytic convention, we write $X \ll_{a,b,c,...} Y$ or $X = O_{a,b,c,...}(Y)$ to mean that there is a function $C(a,b,c,\ldots)$, depending only on the subscript variables, such that $X \leq C(a,b,c,\ldots) \cdot Y$. We mean the same when we write $Y \gg_{a,b,c,...} X$ and $Y = \Omega_{a,b,c,...}(X)$.

**Theorem 1.7 (BGT, weak form).** Let $A$ be a finite $K$-approximate group. Then there is a subset $A'$ of $A^4$ and a compact subgroup $H \subseteq A'$, normal in $\langle A' \rangle$, such that:

(i) (Largeness) $|A'| \gg_K |A|$;

(ii) (Controlled nilpotence) $\langle A' \rangle / H$ is a nilpotent group of rank and step $O_K(1)$.

Although this result comes with worse bounds than those achieved in work on special cases, it holds in full generality for finite approximate groups. Until the present work, however, no general description of infinite approximate groups has appeared in the literature. Our results generalize the main theorem in [7] to the case where $A$ is an open precompact subset of a topological group.

Before stating our main theorem, let us remark that the proof of Theorem 1.7 heavily uses the body of results and techniques around Hilbert’s fifth problem, in particular the so-called Gleason-Yamabe theorem [10, 23] which describes locally compact groups in terms
of Lie groups. We quote it here, in slightly non-standard form, because our results may also be seen as a quantitative version of this purely qualitative result:

**Theorem 1.8** (Gleason-Yamabe). Let $U$ be a precompact neighborhood of the identity in a locally compact group $G$. Then there is $U' \subseteq U$ and a compact subgroup $H \subseteq U'$, normal in $\langle U' \rangle$, such that:

(i) (Largeness) $U'$ is open, thus $U$ can be covered by finitely many translates of $U'$;

(ii) (Lie structure) $\langle U' \rangle / H$ is a Lie group.

The connection between finite approximate groups and locally compact groups was forged with model-theoretic tools in work of Hrushovski [13] and was widely considered a breakthrough; no fully general description of finite approximate groups is known that substantially departs from this approach. Hrushovski then uses the Gleason-Yamabe theory to connect (asymptotic families of) finite approximate groups to Lie groups, and he deduces structural information on the former from basic facts of Lie theory. These connections animate much of the present work.

Below is a simplified statement of our main theorem. Recall that if $G$ is a Lie group and $\mathfrak{g}$ its Lie algebra, the adjoint action of $G$ associates to $g \in G$ a Lie algebra automorphism $\text{Ad}_g$ of $\mathfrak{g}$. It is defined by the requirement that $g \cdot \exp(tX) \cdot g^{-1} = \exp(t\text{Ad}_gX)$ for all $X \in \mathfrak{g}$ and all $t \in \mathbb{R}$.

**Theorem 1.9** (Main theorem, simple form). Let $A$ be an open precompact $K$-approximate group inside a locally compact group $G$. Then, for every $\varepsilon > 0$, there is a subset $A' \subseteq A^4$ and a compact subgroup $H \subseteq A'$, normal in $\langle A' \rangle$, such that:

(i) (Largeness) $A$ can be covered by $O_{K,\varepsilon}(1)$ left-translates of $A'$;

(ii) (Controlled Lie structure) $\langle A' \rangle / H$ is a Lie group of dimension $O_K(1)$;

Moreover, if we denote by $\mathfrak{a}'$ the Lie algebra of $\langle A' \rangle / H$ and by $A''$ the image of $A'$ in the quotient, then we also have
(iii) (Approximate nilpotence) For $g \in A''$, the spectral radius of $\text{Ad}_g - I$ is $O_K(\varepsilon)$;

(iv) (Finite + convex decomposition) There is a convex set $B \subseteq a'$ such that $A''/\exp(B)$ is a finite $O_{K,\varepsilon}(1)$-approximate local group.

In later chapters we will make (iii)-(iv) a little more precise, but doing so will require additional definitions, which we have elected to postpone for fear of making this Introduction too cumbersome. In the meantime, note the analogy between (i)-(ii) in our main theorem and (i)-(ii) in Gleason-Yamabe. Although we have labeled (iii) “approximate nilpotence” — due to the fact that, in a nilpotent Lie group, $\text{Ad}_g - I$ is a nilpotent map, with spectral radius 0 — we may also regard (iii) as “controlled Lie structure”, further strengthening the case that our result is a quantitative version of Gleason-Yamabe. This will become even more apparent once we state the full version of our theorem, which puts bounds on the operator norm (induced by a certain natural norm) of the Lie bracket on $A''$.

1.2 Before the Lie Model Theorem

The investigation of sum sets is arguably centuries old; it is possible to phrase Goldbach’s conjecture and Lagrange’s four square theorem naturally in terms of them. To wit, if $P$ is the set of positive odd primes, $E$ the set of positive even numbers, and $Q$ the set of perfect squares, then they ask and state, respectively:

**Conjecture 1.10** (Goldbach, 1742). $P + P = E \setminus \{2, 4\}$.

**Theorem 1.11** (Lagrange, 1770). $Q + Q + Q + Q = \mathbb{N}$

Questions about sum sets involving specific sets of a number-theoretic flavor continue to attract interest; more recent results include the proof of Waring’s conjecture (the analogue of Lagrange’s theorem for higher powers), and the so-called weak Goldbach conjecture. Let $O$ be the set of positive odd numbers, and for $k \in \mathbb{N}$ let $Q_k$ be the set of perfect $k$-th powers. Then:
Theorem 1.12 (Hilbert, 1909). For every \( k \in \mathbb{N} \) there is \( \ell \in \mathbb{N} \) such that \( \ell Q_k = \mathbb{N} \).

Theorem 1.13 (Helfgott, 2013). \( P + P + P = O \setminus \{3, 5\} \).

It was perhaps Schnirelmann [19] who first investigated general sum sets (as opposed to particular ones from number theory), although he did so with an eye towards the Goldbach and Waring problems. Roughly speaking, Schnirelmann showed in 1933 that, if an infinite set of positive integers \( S \) isn’t “too sparse”, then for some \( \ell \) we have \( \ell S = \mathbb{N} \).

But it was Freiman, in 1966 [9], who first studied sum sets for their own sake. In particular, he proved a deep theorem characterizing finite sets of integers \( A \) such that \( |A + A| \leq K |A| \) in terms of fairly explicit objects depending only on \( K \). We need a definition in order to state it.

Definition 1.14. A standard progression of rank \( r \) is a subset of \( \mathbb{Z}^r \) of the form \( \{0, 1, \ldots, \ell_1\} \times \cdots \times \{0, 1, \ldots, \ell_r\} \) for some \( \ell_1, \ldots, \ell_r \in \mathbb{N} \). If \( G \) is a group, a progression of rank \( r \) in \( G \) is a set of the form \( x_0 \phi(P) \), where \( x_0 \in G \), \( P \) is a standard progression of rank \( r \), and \( \phi : \mathbb{Z}^r \to G \) is a homomorphism.

Remark. It is easily seen that a standard progression of rank \( r \) has cover doubling \( 2^r \); therefore, a progression of rank \( r \) in an abelian group also has cover doubling \( 2^r \). (See Lemma 4.5.)

Theorem 1.15 (Integer case). Let \( A \) be a finite \( K \)-approximate group inside the integers. Then there is a subset \( A' \) of \( 4A \) such that:

(i) (Largeness) \( A \) is contained in \( A' \);

(ii) (Controlled progression structure) \( A' \) is a progression of rank \( O_K(1) \) in \( \mathbb{Z} \).

Freiman’s theorem, especially after its rediscovery by Ruzsa [16], sparked much interest in generalizations. We describe some of the milestones reached in this line of research, as motivation for the more precise statement of BGT (Theorem 1.24) and our own results.
In 2003 Green and Ruzsa [12] found a generalization of Freiman’s theorem to an arbitrary abelian group, which we state in a slightly weaker form that better fits the story being told here:

**Theorem 1.16** (Abelian case). Let \( A \) be a finite \( K \)-approximate group inside an abelian group \( G \). Then there is a subset \( A' \) of \( 4A \) and a subgroup \( H \subseteq A' \) such that:

(i) (Largeness) \( A \) is contained in \( A' \);

(ii) (Controlled progression structure) \( A'/H \) is a progression of rank \( O_K(1) \) in \( G/H \).

In their paper, Green and Ruzsa obtain effective bounds for the dependence on \( K \) and a more precise description of \( A' \), but that need not concern us here.

The Green-Ruzsa theorem is as satisfying as could have been expected, and it shared with (modern proofs of) Freiman’s theorem the use of Fourier analysis to obtain what we have been calling “progression structure”. However, once one leaves the abelian setting, one loses Fourier analysis. The shortage of periodic functions makes it difficult to locate progression-like objects.

Foreshadowing later developments, in 2009 Breuillard and Green [2] made headway in the torsion-free nilpotent case by embedding the ambient group in a simply-connected nilpotent Lie group (via a result of Mal’cev) and transferring the problem to the Lie algebra, using the fact that, in the simply-connected nilpotent case, the logarithm is a diffeomorphism from the Lie group to its Lie algebra.

Recall that a nilpotent group is said to be of **step** \( s \) if its lower central series ends after \( s \) steps; in other words, if any nested commutator of depth \( s + 1 \) is trivial. Also, the group is said to be of **rank** \( r \) if it can be generated by \( r \) elements. Then we may state a simplified form of the Breuillard-Green result as follows:

**Theorem 1.17** (Torsion-free nilpotent case, I). Let \( A \) be a finite \( K \)-approximate group inside a torsion-free nilpotent group \( G \) of step \( s \). Then there is a subset \( A' \) of \( G \) such that:

(i) (Largeness) \( A \) can be covered by \( O_{K,s}(1) \) translates of \( A' \);
(ii) (Controlled nilpotent structure) \( \langle A' \rangle \) is nilpotent of rank \( O_{K,s}(1) \) (and step \( s \)).

(The full statement of the result, which is quite satisfying, involves a nilpotent analogue of progressions; their version of (ii) could fairly be labeled “progression structure”.)

Around the same time Tao [20] obtained partial results on finite approximate subgroups of solvable groups, by some fairly laborious induction on derived length. His results are cumbersome to state precisely, but the upshot is that \( K \)-approximate groups in solvable groups of derived length \( \ell \), after quotienting by an appropriate subgroup, can be controlled by certain nilpotent objects of size, rank, and step depending only on \( K \) and \( \ell \).

Shortly after the latter two results were proved, Hrushovski [13] published the connection with locally compact groups and Gleason-Yamabe theory, which is a major focus of this work and will be detailed in the following chapters. For now, it is worth mentioning two other results that avoid that connection and make progress on the nilpotent case. In particular they obtain explicit bounds, although we suppress those here.

The first, due to Breuillard, Green and Tao in 2011 [5], borrows a trick of Gleason (which itself became part of Gleason-Yamabe theory) to establish:

**Theorem 1.18** (Torsion-free nilpotent case, II). Let \( A \) be a finite \( K \)-approximate group inside a torsion-free nilpotent group \( G \). Then there is a subset \( A' \) of \( A^4 \) such that:

(i) (Largeness) \( A \) can be covered by \( O_K(1) \) translates of \( A' \);

(ii) (Controlled nilpotent structure) \( \langle A' \rangle \) is nilpotent of rank and step \( O_K(1) \).

The second, due to Tointon in 2014, eliminates the torsion-free requirement:

**Theorem 1.19** (Nilpotent case). Let \( A \) be a finite \( K \)-approximate group inside a nilpotent group \( G \) of step \( s \). Then there is a subset \( A' \) of \( A^{O_{K,s}(1)} \) and a subgroup \( H \) of \( G \), normal in \( \langle A \rangle \), such that:

(i) (Largeness) \( A \) can be covered by \( O_{K,s}(1) \) translates of \( A' \);
(ii) (Controlled nilpotent structure) $\langle A' \rangle / H$ is nilpotent of rank $O_{K,s}(1)$ (and step $s$).

(The full statement of Tointon’s theorem also involves a nilpotent analogue of progressions and can fairly be described as “progression structure”.)

The reader is invited to contrast theorems 1.15, 1.17, and 1.18 on the one hand, with theorems 1.16, 1.19, and 1.7 on the other. It should be apparent that, at least in the finite case, torsion creates the need to quotient our approximate groups by certain genuine subgroups before we are able to describe them in terms of progressions. That stands to reason, since the cartesian product of a $K$-approximate group with an arbitrary finite group is itself a $K$-approximate group, but in general does not have anything like a progression structure.

1.3 The Lie Model Theorem

As we have hinted at previously, in 2009 Hrushovski [13] proved the following theorem by methods from model theory, and stable group theory in particular:

**Theorem 1.20** (Local compactness, rough version). Let $(A_n)$ be an asymptotic family of finite $K$-approximate groups (where $K$ is uniform) and form the ultraproduct $A = \prod_{n \to \alpha} A_n$. It is possible to give $\langle A \rangle$ a topology, generated by certain sets of the form $A'_n = \prod_{n \to \alpha} A'_n$ where each $A'_n$ is a subset of $A_n$, in such a way that $\langle A \rangle$ becomes a locally compact group, with $A^4$ a precompact neighborhood of the identity.

From this and Gleason-Yamabe theory Hrushovski was able to deduce the following crucial result:

**Theorem 1.21** (Lie model, rough version). Let $(A_n)$ be an asymptotic family of finite $K$-approximate groups (where $K$ is uniform) and form the ultraproduct $A = \prod_{n \to \alpha} A_n$. Then there is a Lie group $L$ and a surjective morphism $\pi : \langle A \rangle \to L$ which is preserves a good amount of structure.
(This work includes a whole chapter devoted to a generalization of these theorems, so we will not try to state them in full precision in this introduction. There will be time later.)

With the Lie Model Theorem in hand, Hrushovski employs tools from Lie theory (in particular, commutator bounds) to extract non-trivial information about \( A \), and then, by standard compactness arguments, about asymptotic families of \( K \)-approximate groups. In particular, he locates non-commutative analogues of so-called Bourgain systems; these in turn are a natural generalization of Bohr sets, which play a major role in the proof of Freiman’s theorem.

Still, Hrushovski’s results weren’t quite the direct analogues of the theorems in the preceding section that the community hoped for and conjectured. Specifically, there was missing the explicit controlled nilpotence and/or progression structure.

An old result of Jordan, generalized by Boothby-Wang [1], suggested a way forward. It deals with finding nilpotent-type structure (abelian, in fact) by combining finiteness with Lie structure:

**Theorem 1.22** (Jordan-Boothby-Wang). Let \( F \) be a finite subgroup inside a connected Lie group \( L \). Then there is a subset \( F' \) of \( F \) such that:

(i) (Largeness) \( F \) can be covered by \( O_L(1) \) translates of \( F' \);

(ii) (Nilpotent-type structure) \( F' \) is an abelian group.

By combining an argument in (some proofs of) Jordan’s theorem with Hrushovski’s transference to the Lie setting, Breuillard, Green and Tao [7] were able to establish a general characterization of finite \( K \)-approximate groups more closely analogous to that known in the abelian and nilpotent cases. We need a definition in order to state their result.

**Definition 1.23.** Let \( \Gamma_{r,s} \) be the free nilpotent group of rank \( r \) and step \( s \), and let \( u_1, \ldots, u_r \) be its generators. A standard nilprogression of rank \( r \), step \( s \), and lengths \( \ell_1, \ldots, \ell_r \) is the set of all products involving the \( u_i \) and their inverses such that the number of occurrences of \( u_i \) and \( u_i^{-1} \), taken together, does not exceed \( \ell_i \). A nilprogression of rank \( r \), step \( s \), and
lengths $\ell_1, \ldots, \ell_r$ in a group $G$ is a set of the form $x_0 \phi(P)$, where $x_0 \in G$, $P$ is a standard nilprogression of rank $r$, step $s$, and lengths $\ell_1, \ldots, \ell_r$, and $\phi : \Gamma_{r,s} \to G$ is a homomorphism.

**Theorem 1.24 (BGT).** Let $A$ be a finite $K$-approximate group. Then there is a subset $A'$ of $A^4$ and a subgroup $H \subseteq A'$, normal in $\langle A' \rangle$, such that:

(i) (Largeness) $|A'| \gg_K |A|$;

(ii) (Controlled nilprogression structure) $A'/H$ is a nilprogression of rank and step $O_K(1)$.

Thus there are two separate phenomena at work in the BGT theorem. One, which is not apparent in the statement but plays a vital role in the proof, is that an “efficient covering property” — covering $A^2$ by $K$ translates of $A$ — gives rise to Lie structure via Gleason-Yamabe. The other is that finiteness, in the presence of Lie structure, gives rise to nilpotent structure via a Jordan-type argument.

It is worth remarking that in [7] BGT was proved at a greater level of generality, that of local groups. They need the extra generality at the Jordan-like step of their argument, which uses finiteness of the approximate groups in a crucial way. Since our results are not restricted to finite approximate groups we have no analogue of this step, and are thus able to remain in the setting of global groups for most of this work. The only exception is at the very end, when we use BGT as a black box to decompose an open precompact approximate group into a “convex” part and a finite part; the right concept for the job turns out to be a quotient of local groups.

Finally, it might be illuminating to spell out the several threads of analogy that run through this work on the one hand, and BGT and Gleason-Yamabe on the other. In the table below, occurrences of $O_K(1)$ mean that all measures of complexity and “wildness” — rank and step for nilprogressions; dimension, adjoint action, and bracket for Lie groups — are bounded in the relevant sense by constants depending only on $K$. 

---

12
### Gleason-Yamabe

- $U$, neighborhood of 1
- $U'$, open subset of $U$
- compact normal subgroup $H$
- $U'/H$ is Lie

### BGT

- $A$, finite approximate group
- $A'$, large subset of $A^4$
- normal subgroup $H$
- $A'/H$ is $O_K(1)$-nilprogression

### This work

- $A$, approximate group
- $A'$, large subset of $A^4$
- compact normal subgroup $H$
- $A'' = A'/H$ is $O_K(1)$-Lie
- convex set $B \subseteq a'$
- $A''/\exp(B)$ is finite $O_K(1)$-approximate group

## 1.4 Proof Strategy and Statement of Results

The argument is naturally divided into three parts.

In part one, the goal is a continuous analogue of the Lie Model Theorem, which says that an ultraproduct $A$ of a family of open precompact $K$-approximate groups $(A_n)$ admits a “useful” morphism into a Lie group. To that end we adapt the results of BGT, hewing fairly closely to their strategy: we prove some (continuous analogues of) additive combinatorial theorems, and use those to show that $A$ is locally compact in a certain sense. We then
use Gleason-Yamabe theory to obtain the “useful” morphism. One difference between our approach and that of BGT is that we have elected to use (continuous analogues of) results of Croot and Sisask [8] as our additive combinatorial foundation, as opposed to the results of Sanders [18] employed by BGT. Our continuous analogues of the Croot-Sisask results might be of independent interest.

In part two, the goal is to locate within an approximate group a large subset with some extra regularity, in particular a kind of dilation structure. (We call it a strong approximate group.) We first find strong approximate groups at the ultraproduct level: we leverage the useful morphism constructed in part one to obtain a strong approximate group \( \tilde{A} \) inside \( A^4 \). (Via the Lie model, the dilation structure ultimately comes from that in \( \mathbb{R}^n \).) A standard compactness argument then shows that it is possible to find such large regular sets inside any approximate group, although the bounds on largeness are ineffective. Next, we define a certain norm-like quantity (the escape norm) which makes sense in any approximate group, but we show that it has several useful properties in a strong approximate group — the so-called Gleason lemmas.

In part three, the goal is to describe (a large subset of) the strong approximate group in terms of subgroups, convex sets, and nilprogressions. Starting with a \( K \)-approximate group \( A \), we pass to a large strong \( O_K(1) \)-approximate subgroup \( \tilde{A} \). We then use the escape norm to find a compact normal \( O_K(1) \)-approximate subgroup \( \tilde{A} \) and argue that \( \langle \tilde{A} \rangle/H \) is a Lie group. Next, we define the path norm — a norm-like quantity related to the escape norm, but now on the Lie algebra \( \tilde{a} \) of \( \langle \tilde{A} \rangle/H \). We use it to prove that for every \( \varepsilon > 0 \) there is large subset \( A' \) of \( \tilde{A} \) such that for all \( g \in A' \) the operator norm (and hence spectral radius) of \( \text{Ad}_g - I \) is \( O_K(\varepsilon) \). It then follows that the dimension of \( \langle \tilde{A} \rangle/H \) is \( O_K(1) \). Finally, we show that, if \( B \) is a small ball in the path norm, and \( A' \) is a small ball in the escape norm, then the local group quotient \( A'/\exp(B) \) is discrete, and hence, by precompactness, finite.

Perhaps one additional point of interest is that the path norm is defined uniformly, with \( A \) as just a parameter; this may have useful implications in the case when \( A \) is definable, or in arguments involving ultraproducts.
We may now formulate our main theorem more precisely as follows:

**Theorem 1.25** (Main theorem). Let $A$ be an open precompact $K$-approximate group inside a locally compact group $G$. Then, for every $\varepsilon > 0$, there is a subset $A' \subseteq A^4$ and a compact subgroup $H \subseteq A'$, normal in $\langle A' \rangle$, such that:

(i) (Largeness) $A$ can be covered by $O_{K,\varepsilon}(1)$ left-translates of $A'$;

(ii.a) (Controlled Lie structure, 1) $\langle A' \rangle / H$ is a Lie group of dimension $O_K(1)$;

Moreover, let $\mathfrak{a}'$ be the Lie algebra of $\langle A' \rangle / H$ and $A''$ the image of $A'$ in the quotient. Then there is a norm $\| \cdot \|$ on $\mathfrak{a}'$ such that:

(ii.b) (Controlled Lie structure, 2) For $X, Y \in \mathfrak{a}'$ we have $\| [X, Y] \| \ll_K \|X\| \|Y\|$.

(iii) (Approximate nilpotence) For $g \in A''$ the operator norm (induced by $\| \cdot \|$) of $\text{Ad}_g - I$, and hence its spectral radius, is $O_K(\varepsilon)$;

(iv) (Finite + convex decomposition) There is a convex set $B \subseteq \mathfrak{a}'$ such that $A'' / \exp(B)$ is a finite $O_{K,\varepsilon}(1)$-approximate local group.
CHAPTER 2

Examples and Non-Examples

Before proceeding to general results we pause to consider some examples of infinite approximate groups.

**Example 2.1.** If $A$ is a finite $K$-approximate group inside a discrete group $G$, and $H$ is a compact group, then $A \times H$ is an open precompact $K$-approximate group in $G \times H$.

One may also construct a continuous version of the objects in Freiman’s theorem:

**Example 2.2.** If $P$ is a symmetric progression of rank $r$ in $\mathbb{R}^d$ containing the origin, and $C$ is an open precompact symmetric convex subset of $\mathbb{R}^d$, then $P + C$ is a $2^r 5^d$-approximate group.

**Proof.** Clearly $P + C$ is open, precompact, symmetric, and contains the origin. We know that $P + P \subseteq X + P$ for some finite $X$ of size $2^r$. If we can show that $C + C \subseteq Y + C$ for some finite $Y$ of size $5^d$ we will be done, since

$$
(P + C) + (P + C) = (P + P) + (C + C) \\
\subseteq (X + P) + (Y + C) \\
= (X + Y) + (P + C)
$$

and $|X + Y| \leq |X| |Y| = 2^r 5^d$.

We find the required $Y$ by a standard volume-packing argument. Let $C_\lambda$ be the $\lambda$-dilate of $C$; by convexity and symmetry, $C_\lambda + C_\mu = C_{|\lambda|+|\mu|}$. Let $Y$ be a maximal subset of $C + C$ such that the sets $\{y + C_\frac{1}{2} \mid y \in Y\}$ are pairwise disjoint; then $Y + C_\frac{1}{2} \subseteq C + C + C_\frac{1}{2} = C_\frac{5}{2}$.
It follows that $\text{vol}(Y + C_{1/2}) = |Y| \frac{\text{vol}(C_{1/2})}{\text{vol}(C_{1/2})} \leq \text{vol}(C_{1/2})$. But $\text{vol}(C_{\lambda}) = \lambda^d \text{vol}(C)$, and so $|Y| \leq \frac{\text{vol}(C_{1/2})}{\text{vol}(C_{1/2})} = 5^d$.

We claim that $C + C \subseteq Y + C$ with this choice of $Y$. Indeed, let $x \in C + C$ be arbitrary. By maximality of $Y$, there is $y \in Y$ such that $x + C_{1/2}$ and $y + C_{1/2}$ intersect, say at $x + c' = y + c''$. But then $x = y + (c'' - c')$, which lies in $Y + C_{1/2} + C_{1/2} = Y + C$. \qed

In particular, any open, precompact convex set in dimension $d$ is a $5^d$-approximate group. One may transfer this result from vector spaces to Lie groups via the exponential map:

**Example 2.3.** Let $G$ be a $d$-dimensional Lie group, $\mathfrak{g}$ its Lie algebra, and $C$ an open precompact symmetric convex subset of $\mathfrak{g}$. Then for sufficiently small $\lambda > 0$ the set $\exp(\lambda B)$ is a $10^d$-approximate group.

*Proof.* This follows from the fact that, at small scales near the identity, multiplication in a $d$-dimensional Lie group is close to addition in $\mathbb{R}^d$. \qed

By similar arguments we have

**Example 2.4.** Let $A$ be a finite $K$-approximate group in a $d$-dimensional Lie group $L$ endowed with a left-invariant metric. Let $B(1, \varepsilon)$ be a sufficiently small ball normalized by $A$. Then $A \cdot B(1, \frac{\varepsilon}{2})$ is a $10^dK$-approximate group.

At first blush it may appear that requiring approximate groups to be open doesn’t make much of a difference; one might conjecture that, if $A$ is a $K$-approximate group in every way except for being open, then sufficiently small open neighborhoods of $A$ are $O_K(1)$-approximate groups. The next example shows that this is not the case.

**Example 2.5.** Let $G$ be the group $\text{Aff}^+(2, \mathbb{R})$ of increasing affine transformations of $\mathbb{R}$, i.e. elements of $G$ are maps of the form $x \mapsto ax + b$ with $a > 0$ and the group operation is composition. Let $P = \{-N, \ldots, -1, 0, 1, \ldots, N\}$ and $C = (-0.1, 0.1)$. Then the set

$$Q_N = \{ x \mapsto x + b \in \text{Aff}^+(2, \mathbb{R}) \mid b \in P + C \}$$

17
is a 10-approximate group by Example 2.2. (Note that that the subgroup \( \{1\} \times \mathbb{R} \) of \( \text{Aff}^+(2, \mathbb{R}) \) is isomorphic to \( \mathbb{R} \).) This is true independent of \( N \). On the other hand, we will see in future chapters that if \( x \mapsto ax + b \) belongs to an open precompact \( K \)-approximate group, then there is an \( O_K(1) \) bound on \( b \). Thus for large enough \( N \), although \( Q_N \) is a 10-approximate group, there is no open neighborhood of it that is an \( O(1) \)-approximate group.

Roughly speaking, if \( g \) is the transformation \( x \mapsto ax + b \), then the operator norm of \( \text{Ad}_g - I \) (induced by any norm on the Lie algebra \( g = \mathbb{R}^2 \)) goes to infinity as \( b \to \infty \), but by Theorem 1.25 it should remain bounded by \( O_K(1) \).
CHAPTER 3

Some Special Cases

If one places restrictions on $K$, or in the ambient group, it is possible to use existing techniques to give fairly satisfactory descriptions of infinite $K$-approximate groups; we provide three examples.

The first is folklore in its finite form, and adapts without difficulty to the locally compact setting.

**Theorem 3.1** (Small $K$). Let $S$ be a symmetric measurable subset of a locally compact group $G$ with left Haar measure $\mu$, such that $S^2$ is also measurable. Suppose $S$ has doubling strictly less than $\frac{3}{2}$. Then $S^2$ is a genuine subgroup of $G$.

**Proof.** Clearly $S^2$ contains the identity and is closed under inverses, so we only need to address closure under products.

We start by showing that each element of $S^2$ has “many” representations as a product of elements of $S$. To that end, let $x \in S^2$, say $x = ab$ with $a, b \in S$, and define

$$A_x = \{a' \in S \mid x = a'b' \text{ for some } b' \in S\}.$$ 

Since $ab = a'b'$ iff $a' = ab(b')^{-1}$ and $S$ is symmetric, we have $A_x = S \cap abS$, which is measurable. This means that

$$\mu(A_x) = \mu(S \cap abS) = \mu(a^{-1}S \cap bS)$$

by left invariance. But the sets $a^{-1}S$ and $bS$ each have measure $\mu(S)$ and are contained in $S^2$, which has measure less than $\frac{3}{2}\mu(S)$; it follows that $\mu(a^{-1}S \cap bS) > \frac{1}{2}\mu(S)$.
Now let \( x, y \in S^2 \) be arbitrary; the sets \( A_{x^{-1}} \) and \( A_y \) are both contained in \( S \) and measure more than \( \frac{1}{2} \mu(S) \), so they must overlap, say at \( a_0 \). Then there are \( b, b' \in S \) such that \( x^{-1} = a_0 b \) and \( y = a_0 b' \). But then \( xy = (a_0 b)^{-1} a_0 b' = b^{-1} b' \), which lies in \( S^2 \). Thus \( S^2 \) is closed under products and is a genuine subgroup. \( \square \)

Thus, when a set has small enough doubling, it is a large subset of an actual subgroup; this is in line with our philosophy that approximate groups should be closely related to group-theoretic objects.

If we allow \( K \) to be arbitrary but restrict our attention to certain kinds of groups, then known results suffice to obtain some description of approximate groups. The next two result are the simplest extensions of BGT (which deals with the finite case) and Freiman (which deals with grid-like objects) to infinite contexts.

**Theorem 3.2** (Profinite case). Let \( G \) be a profinite group and \( A \subseteq G \) an open \( K \)-approximate group. Then there is a compact subgroup \( H \), contained in \( A \) and normal in \( G \), such that \( A/H \) is a finite \( K \)-approximate group.

**Proof.** Since \( A \) is open and \( G \) is profinite, there exists a compact subgroup \( H \), contained in \( A \) and normal in \( G \), such that \( G/H \) is finite; in particular, \( A/H \) is finite. By Lemma 4.5 \( A/H \) is a \( K \)-approximate group. \( \square \)

Finally, when the ambient group is \( \mathbb{R}^d \), one can easily adapt Freiman’s theorem to describe open precompact approximate groups. This is a good exercise and appears in [21]:

**Theorem 3.3** (Freiman in \( \mathbb{R}^d \)). Let \( A \) be an open precompact \( K \)-approximate group in \( \mathbb{R}^d \). Then there is a subset \( A' \) of \( 4A \) such that:

(i) (Largeness) \( A \subseteq A' \);

(ii) (Progression + convex decomposition) There is a progression \( P \) in \( \mathbb{R}^d \) of rank \( E_K(1) \) and a convex set \( B \subseteq \mathbb{R}^d \) such that \( A' = P + B \).
The key to the preceding theorem is that there are arbitrarily dense subgroups of $\mathbb{R}^d$ isomorphic to $\mathbb{Z}^d$; one then approximates any open set by a set of the form $X + B$ where $X$ is contained in some such subgroup and $B$ is an arbitrarily small ball. As it turns out, the small doubling condition on $A$ transfers to $X$, which is then described via Freiman’s theorem.

Unfortunately this approach cannot succeed in groups that are from abelian, due to lack of any surrogate of the subgroups $\varepsilon\mathbb{Z}^d$, $\varepsilon > 0$. So, starting with the next chapter, we take up the general case.
The purpose of this chapter is to set down some definitions and recall useful lemmas from the literature. As we have already mentioned, our focus will be on open precompact approximate groups inside locally compact groups. One reason for this focus is that the most natural analogue of cardinality in an infinite additive combinatorial context is Haar measure, the existence of which is guaranteed by local compactness. Moreover, precompactness allows us to keep product sets under at least qualitative control, as elaborated on after this

**Definition 4.1.** Let $G$ be a locally compact group. We call a subset $S \subseteq G$ a *multiplicative set* if it is open and precompact. We say $S$ is *symmetric* if it is closed under inverses.

*Remark.* The point of this definition is that it gives us *carte blanche* to form product sets at will without worrying about their measurability, or whether their Haar measures might be zero or infinite, which would often cut short arguments involving inequalities between those measures. For instance, if $A, B, C$ are multiplicative sets, then so are $AB, AB^2C^{-1}A, A^{12}$, etc; and all their measures are finite and nonzero.

We now fix the notion of approximate group we will use for the rest of the work.

**Definition 4.2** (Approximate groups). Let $G$ be a locally compact group. A multiplicative subset $A$ of $G$ is said to be a $K$-approximate group if: (i) $1 \in A$; (ii) $A$ is symmetric; and (iii) there are finite sets $X, Y \subseteq G$, with at most $K$ elements each, such that $A \cdot A \subseteq X \cdot A$ and $A \cdot A \subseteq A \cdot Y$.

The classical results in additive combinatorics, such as Freiman’s theorem, focus on doubling rather than cover doubling, and one may reasonably ask how interchangeable they are.
They’re not completely equivalent, even in the simplest of contexts: for sufficiently large \( n \), there is a set of \( n \) integers with doubling at most 6 but cover doubling \( \Omega(\log n) \), as the next example shows. The idea is that a random subset \( A \) of \( \{1, \ldots, n\} \) satisfies the following with high probability: (i) it has small doubling; (ii) \( A + A \) contains a fixed long interval; (iii) few translates of \( A \) do not cover that long interval.

**Example 4.3.** Let \( A \) subset of \( \{1, \ldots, n\} \) chosen uniformly at random. Then \(|A| \geq \frac{1}{3}n\) with probability very close to 1 if \( n \) is large enough. Since \( A + A \subseteq \{1, \ldots, 2n\} \), the doubling of \( A \) is at most 6 with high probability.

Next we show that \( A + A \) likely contains a long interval. For any \( x \) we have \( x \not\in A + A \) iff for every \( i < x \) it is not the case that \( i \) and \( x - i \) both lie in \( A \). If \( A \) is uniformly random, the odds that \( i \) and \( x - i \) are both in \( A \) are at least \( \frac{1}{4} \). Thus

\[
\mathbb{P}[x \not\in A + A] \leq \prod_{i=1}^{x-1} \left(1 - \frac{1}{4}\right) = \left(\frac{3}{4}\right)^{x-1}.
\]

By the union bound, the probability that \( A + A \) contains every number between \( \frac{1}{2}n \) and \( \frac{3}{2}n \) is at least \( 1 - n \left(\frac{3}{4}\right)^{n/2} \); in other words, \( A + A \) contains \( J = \{\frac{1}{2}n, \ldots, \frac{3}{2}n\} \) with high probability.

Now fix \( X = \{x_1, \ldots, x_K\} \subseteq \mathbb{Z} \); we will show that \( J \subseteq X + A \) with low probability. We start by finding a large subset \( C \) of \( J \) for which the events \( c \in x_i + A, \ c' \in x_j + A \) are independent for any \( i,j \) and distinct \( c, c' \in C \). We build \( C \) inductively as follows. Start by putting \( \frac{1}{2}n \) into \( C \) and, while \( C + (X - X) \) does not contain \( J \), add to \( C \) some element of \( J \) not in \( C + (X - X) \). Since \( J \) has \( n \) elements and \( C + (X - X) \) has at most \( K^2 |C| \) elements, we can keep going until \( |C| = \frac{n}{K^2} \). Now, by construction, if \( c, c' \in C \) are distinct then \( x_i + c \neq x_j + c' \) for all \( i, j \); thus the events \( c \in x_i + A, \ c' \in x_j + A \) are independent as promised.

The probability that an individual \( c \in C \) lies in \( X + A \) is at most \( 1 - \frac{1}{2^K} \), since that is equivalent to \( c - X \) intersecting \( A \). It follows that the probability that \( C \subseteq X + A \) is at most

\[
\left(1 - \frac{1}{2^K}\right)^{n/K^2} \sim e^{-n/2^K K^2}.
\]

Of course, that is also an upper bound on the probability that \( X + A \) contains \( J \).
Finally, since \( x + A \) intersects \( J \) only if \(-\frac{1}{2}n \leq x \leq \frac{3}{2}n\), the union bound implies that the probability that \( J \subseteq X + A \) for some \( X \) is at most
\[
\left( \frac{2n}{K} \right) e^{-n/2K^2} K^2 \leq \frac{(2n)^K}{K!} e^{-n/2K^2} K^2 \leq e^{K \log n - n/2K^2}.
\]

Now it is easy to see that, if \( K = \frac{1}{2} \log_2 n \) and \( n \) is large enough, this probability is very close to zero. Hence, with high probability, a random set \( A \subseteq \{1, \ldots, n\} \) has doubling at most 6 and cover doubling greater than \( \frac{1}{2} \log_2 n \).

Still, in some looser sense sets of small doubling are closely related to approximate groups, at least in the finite case. The following results of [21] are well-known:

**Theorem 4.4.** Let \( A \) be a multiplicative set in a discrete group \( G \) with counting measure \( \mu \), and let \( K \geq 1 \).

(i) If \( \mu(A^3) \leq K \mu(A) \), then \( (A \cup \{1\} \cup A^{-1})^3 \) is an \( O(K^{O(1)}) \)-approximate group;

(ii) If \( \mu(A^2) \leq K \mu(A) \), then there exists an \( O(K^{O(1)}) \)-approximate group \( H \) such that \( \mu(H) \leq O(K^{O(1)}) \mu(A) \) and \( A \) is contained in \( O(K^{O(1)}) \) translates of \( H \).

*Remark.* The results above were actually proved in the more general context of groups with bi-invariant Haar measure; this includes not only discrete groups, but also abelian groups, compact groups, and connected nilpotent Lie groups. In Lemma 5.5 we prove a slightly weaker version of (i) that works for all locally compact groups.

Finally, we will use the following lemmas throughout this work; the proofs are elementary.

**Lemma 4.5** (Basic facts). Let \( G, G' \) be groups and \( \phi : G \rightarrow G' \) a morphism. Then the following hold:

(i) If \( A \subseteq G \) contains the identity, is symmetric, or has cover doubling \( K \), then \( \phi(A) \) has the same property;

(ii) If \( A' \subseteq G' \) contains the identity, is symmetric, or has cover doubling \( K \), then \( \phi^{-1}(A') \) has the same property;
(iii) More generally, if $A, B \subseteq G$ and $B$ can be covered by $K$ translates of $A$, then $\phi(B)$ can be covered by $K$ translates of $\phi(A)$;

(iv) If $A', B' \subseteq G'$, $B'$ can be covered by $K$ translates of $A'$, and $\phi$ is surjective, then $\phi^{-1}(B)$ can be covered by $K$ translates of $\phi^{-1}(A)$.
CHAPTER 5

Sanders-Croot-Sisask Theory

The goal of this chapter is to find, inside an approximate group $A$, “small neighborhoods of the identity”: for each fixed $m$, a set $S \subseteq A$ such that $S^m \subseteq A$ (actually $A^4$, but we regard $A$ and $A^4$ as “close”).

The way we go about this follows the strategy of Croot and Sisask in [8]; it works by a random sampling procedure, as opposed to Sanders’ density increment method of [18]. What the approaches of Sanders and Croot-Sisask have in common is that they both construe the ambient group $G$ as acting on some conveniently chosen metric space $M$, and proving that there must be a set $S \subseteq G$ of appreciable size which “almost stabilizes” certain points of $M$. The triangle inequality then implies that translations from $S^m$, being a series of $S$-translations, move these points by a controlled amount. Finally these authors show, using the specifics of each of their definitions, that this implies $S^m \subseteq A$.

The surprise here, in adapting the techniques of [8], is that (small) finite sets of points suffice to accurately sample from “continuous” sets as well as discrete. One technical point that needs some care is that many obvious facts about counting and discrete probability become less obvious when counting measure is replaced with general Haar measure.

5.1 Finding Almost-Periods

Throughout, $G$ will be a locally compact Hausdorff group, and $A, B$ will be multiplicative subsets of $G$. A fortiori, all product sets such as $A^4$, $AB^{-1}$, etc, will also be open and precompact; see Definition 4.1. This ubiquitous finiteness underwrites our applications of
Fubini-Tonelli, of which there are many, and the hypotheses for which we will not bother to check explicitly. Sets denoted by variables other than $A, B$ will have their regularity properties explicitly stated.

Last words on notation: due to conflict with notation for product sets, we will denote the $n$-fold cartesian product of a set $A$ with itself by $A^\times n$; and $A$ will usually refer to multiplicative sets obeying some approximate closure condition, most commonly sets of small doubling or approximate groups.

**Proposition 5.1** (Finding almost periods). Let $A, B \subseteq G$, let $0 < \varepsilon < 1$ and $K \geq 1$ be parameters, and suppose there is a measurable subset $S$ of $G$ such that $\mu(SA) \leq K \mu(A)$. Then there is a measurable set $T$ which is “dense” $S^{-1}$, i.e.

$$\mu(T) \geq \frac{\mu(S^{-1})}{2K^{8/\varepsilon^2}},$$

and such that every $t \in TT^{-1}$ is an “$\varepsilon$-almost period” of $1_A * 1_B$, in the sense that

$$\|1_A * 1_B(tx) - 1_A * 1_B(x)\|_2^2 < \varepsilon^2 \mu(A)^2 \mu(B).$$

**Proof.** The argument is in two steps; here is a rough sketch. Step 1: we show that a randomly chosen Dirac measure supported on a finite subset of $A$ of appropriate size “looks like $A$” as far as convolution with $1_B$ is concerned. Step 2: we combine the first part with the controlled size of $SA$ to show that many translates of a random Dirac measure supported on a finite subset of $SA$ also look like $A$. It follows that there is some Dirac measure $\delta$ with many translates that look like $A$, which is to say that many translates of $A$ look like $\delta$; thus many translates of $A$ look like each other.

To formalize the above intuitions, our notion of “looking like” will be $L^2(G)$-distance, which has the invariance properties needed for the sketch to work. The work [8] includes extensions of these results to $L^p$ norms but we will not need them here. We now flesh out step 1.

Let $C$ be a randomly chosen $k$-tuple from $A^\times k$, where $k$ is a positive integer to be specified later but depends only on $\varepsilon$. We will use $C$ to code the Dirac measures in the rough sketch.
Note that $C$ is a $k$-tuple, not a $k$-subset; this is only a minor change from [8] and simplifies some integration arguments in the continuous case which do not arise in the discrete case.

Although $C$ is a $k$-tuple, we will use the notation $\sum_{c \in C} f(c)$; it will simply mean the sum as $c$ runs over the coordinates of $C$. Similarly, $1_C(x)$ will indicate how many times $x$ appears among the coordinates of $C$.

Define $\mu_C = \frac{1}{k} \mu_A$, a sort of “Dirac-ization” of $1_A$, and put

$$\mu_C * 1_B = \frac{\mu_A}{k} \sum_{c \in C} 1_{cB}.$$  

The idea is that for many $C$ the functions $\mu_C * 1_B$ and $1_A * 1_B$ will be $L^2$-close; it is in this sense that “a random Dirac measure looks like $A$”. Our aim for the next few paragraphs will be to control the expected value of the $L^2$ distance $\|\mu_C * 1_B - 1_A * 1_B\|_2^2$.

To that end we calculate some statistics of $\mu_C * 1_B$. Fix $x \in G$, let $C = (c_1, \ldots, c_k)$ be taken randomly from $A \times k$ according to (the $k$-fold product of) the law $\mu$, and let us find the first and second moments of $\mu_C * 1_B(x)$. We have

$$\mathbb{E}_{C \in A \times k} [\mu_C * 1_B(x)] = \frac{1}{\mu(A)^k} \int_{A \times k} \mu_C * 1_B(x) \, dC$$

$$= \frac{1}{\mu(A)^k} \int_{A \times k} \frac{\mu_A}{k} \sum_{c \in C} 1_{cB}(x) \, dC$$

$$= \frac{1}{k \cdot \mu(A)^{k-1}} \int_{A} \cdots \int_{A} 1_{c_1B}(x) + \cdots + 1_{c_kB}(x) \, dc_1 \ldots dc_k$$

Now note that for each $j$ we have

$$\int_{A} \cdots \int_{A} 1_{c_jB}(x) \, dc_1 \ldots dc_k = \mu(A)^{k-1} \int_{A} 1_{c_jB}(x) \, dc_j$$

$$= \mu(A)^{k-1} \int_{A} 1_A(c_j) 1_B(c_j^{-1}x) \, dc_j$$

$$= \mu(A)^{k-1} 1_A * 1_B(x)$$

Since there are $k$ such terms in the expression for $\mathbb{E}_{C \in A \times k} [\mu_C * 1_B(x)]$ above, we obtain

$$\mathbb{E}_{C \in A \times k} [\mu_C * 1_B(x)] = \frac{1}{k \cdot \mu(A)^{k-1}} \int_{A} \cdots \int_{A} 1_{c_1B}(x) + \cdots + 1_{c_kB}(x) \, dc_1 \ldots dc_k$$

$$= \frac{1}{k \cdot \mu(A)^{k-1}} (k \cdot \mu(A)^{k-1} 1_A * 1_B(x))$$

$$= 1_A * 1_B(x).$$
The second moment is not much harder:

\[
\mathbb{E}_{C \in \mathcal{A}^k} |\mu_C * 1_B(x)|^2 = \mathbb{E}_{C \in \mathcal{A}^k} \left( \frac{\mu(A)}{k} \sum_{c \in C} 1_{cB}(x) \right)^2
\]

\[
= \frac{\mu(A)^2}{k^2} \mathbb{E}_{C \in \mathcal{A}^k} \left( \sum_{c \in C} 1_{cB}(x) \right)^2
\]

\[
= \frac{\mu(A)^2}{k^2} \mathbb{E}_{C \in \mathcal{A}^k} \left[ \sum_{c, c' \in C} 1_{cB}(x)1_{c'B}(x) \right]
\]

\[
= \frac{\mu(A)^2}{k^2} \frac{1}{\mu(A)^k} \int_{\mathcal{A}^k} \sum_{c, c' \in C} 1_{cB}(x)1_{c'B}(x) dC
\]

\[
= \frac{1}{k^2 \mu(A)^{k-2}} \int_{A} \cdots \int_{A} \sum_{1 \leq i, j \leq k} 1_{c_iB}(x)1_{c_jB}(x) dc_1 \cdots dc_k
\]

The terms in the sum are of two kinds: those where \(i = j\), and the rest. We have already dealt with the former:

\[
\int_{A} \cdots \int_{A} 1_{c_iB}(x)1_{c_iB}(x) dc_1 \cdots dc_k = \int_{A} \cdots \int_{A} 1_{c_iB}(x) dc_1 \cdots dc_k
\]

\[
= \mu(A)^{k-1}1_A * 1_B(x).
\]

Let us deal with the latter. For \(i \neq j\) we have

\[
\int_{A} \cdots \int_{A} 1_{c_iB}(x)1_{c_jB}(x) dc_1 \cdots dc_k = \mu(A)^{k-2} \int_{A} \int_{A} 1_{c_iB}(x)1_{c_jB}(x) dc_i dc_j
\]

\[
= \mu(A)^{k-2} \left( \int_{A} 1_{c_jB}(x) dc_j \right) \left( \int_{A} 1_{c_iB}(x) dc_i \right)
\]

\[
= \mu(A)^{k-2} (1_A * 1_B(x))^2.
\]

Counting the number of each type of term, we arrive at an expression for the second moment:

\[
\mathbb{E}_{C \in \mathcal{A}^k} |\mu_C * 1_B(x)|^2 = \frac{1}{k^2 \mu(A)^{k-2}} \left( k \mu(A)^{k-1}1_A * 1_B(x) + k(k - 1) \mu(A)^{k-2} (1_A * 1_B(x))^2 \right)
\]

\[
= \frac{\mu(A)}{k} 1_A * 1_B(x) + \left( 1 - \frac{1}{k} \right) (1_A * 1_B(x))^2.
\]

The first and second moments enable us to find the variance:
\[
\text{Var}_{C \in A^\times k} [\mu_C * 1_B(x)] = \mathbb{E}_{C \in A^\times k} |\mu_C * 1_B(x)|^2 - (\mathbb{E}_{C \in A^\times k} [\mu_C * 1_B(x)])^2 \\
= \frac{\mu(A)}{k} 1_A * 1_B(x) + \left(1 - \frac{1}{k}\right) (1_A * 1_B(x))^2 - (1_A * 1_B(x))^2 \\
= \frac{\mu(A)}{k} 1_A * 1_B(x) - \frac{1}{k} (1_A * 1_B(x))^2 \\
\leq \frac{\mu(A)}{k} 1_A * 1_B(x).
\]

Since the variance of \(\mu_C * 1_B(x)\) is \(\mathbb{E}_{C \in A^\times k} |\mu_C * 1_B(x) - 1_A * 1_B(x)|^2\), we obtain the inequality

\[
\mathbb{E}_{C \in A^\times k} |\mu_C * 1_B(x) - 1_A * 1_B(x)|^2 \leq \frac{\mu(A)}{k} 1_A * 1_B(x).
\]

Integrating it over \(x \in G\) gives

\[
\int_G \mathbb{E}_{C \in A^\times k} |\mu_C * 1_B(x) - 1_A * 1_B(x)|^2 \, dx \leq \frac{\mu(A)}{k} \int_G 1_A * 1_B(x) \, dx \\
= \frac{1}{k} \mu(A)^2 \mu(B).
\]

Applying Fubini-Tonelli to change the order of the integrals on the left-hand side then yields

\[
\mathbb{E}_{C \in A^\times k} \|\mu_C * 1_B - 1_A * 1_B\|^2 \leq \frac{1}{k} \mu(A)^2 \mu(B),
\]

which is the sought-after bound on the average \(L^2\) distance between \(\mu_C * 1_B\) and \(1_A * 1_B\). Say that \(C \in A^\times k\) looks like \(A\) if \(\|\mu_C * 1_B - 1_A * 1_B\|^2 < \frac{1}{k} \mu(A)^2 \mu(B)\). Since \(\mu(A^\times k)\) is finite, we may employ Markov’s inequality to conclude that

\[
\mathbb{P}_{C \in A^\times k} [C \text{ looks like } A] = \frac{\mu \{ C \in A^\times k \mid C \text{ looks like } A \}}{\mu(A^\times k)} \geq \frac{1}{2}.
\]

This concludes step 1 referred to in the rough sketch; on to step 2.

Choose \(C\) randomly from \((SA)^\times k\) according to the law \(\mu\) and fix \(t \in S^{-1}\). The translate \(tC\) is got from \(C\) by multiplying each coordinate on the left by \(t\); in other words, it shifts the support of the Dirac measure associated to \(C\), but keeping the weights fixed.

Due to left-invariance of Haar measure, we have

\[
\mathbb{P}_{C \in (SA)^\times k} [tC \text{ looks like } A] = \frac{\mu \{ C \in (SA)^\times k \mid tC \text{ looks like } A \}}{\mu((SA)^\times k)} \\
= \frac{\mu \{ C \in (tSA)^\times k \mid C \text{ looks like } A \}}{\mu((tSA)^\times k)} \\
= \mathbb{P}_{C \in (tSA)^\times k} [C \text{ looks like } A]
\]
We can lower bound the latter by $\mathbb{P}_{C \in (tSA)^{xk}} [C \in A^{xk} \text{ and } C \text{ looks like } A]$, which we can in turn lower bound using results from the first part and the fact that $A \subseteq tSA$:

\[
\mathbb{P}_{C \in (tSA)^{xk}} [C \in A^{xk} \text{ and } C \text{ looks like } A] = \\
\mathbb{P}_{C \in (tSA)^{xk}} [C \in A^{xk}] \cdot \mathbb{P}_{C \in (tSA)^{xk}} [C \text{ looks like } A \mid C \in A^{xk}] = \\
\frac{\mu(A^{xk})}{\mu((tSA)^{xk})} \cdot \mathbb{P}_{C \in (tSA)^{xk}} [C \text{ looks like } A \mid C \in A^{xk}] \geq \\
\frac{\mu(A)^k}{\mu(SA)^k} \cdot \frac{1}{2} \geq \\
\frac{1}{2K^k},
\]

where in the last line we used the hypothesis that $\mu(SA) \leq K\mu(A)$. Thus we have proved that, for any fixed $t \in S^{-1}$,

\[
\mathbb{P}_{C \in (SA)^{xk}} [tC \text{ looks like } A] \geq \frac{1}{2K^k}.
\]

Integrating the above inequality over $t \in S^{-1}$ and applying Fubini-Tonelli we get

\[
\int_{S^{-1}} \mathbb{P}_{C \in (SA)^{xk}} [tC \text{ looks like } A] \, dt = \\
\int_{S^{-1}} \frac{\mu \{ C \in (SA)^{xk} \mid tC \text{ looks like } A \}}{\mu((SA)^{xk})} \, dt \\
= \frac{1}{\mu((SA)^{xk})} \int_{S^{-1}} \int_{A^{xk}} 1_{tC \text{ looks like } A} \, dC \, dt \\
= \frac{1}{\mu((SA)^{xk})} \int_{A^{xk}} \int_{S^{-1}} 1_{tC \text{ looks like } A} \, dt \, dC \\
= \frac{1}{\mu((SA)^{xk})} \int_{A^{xk}} \mu \{ t \in S^{-1} \mid tC \text{ looks like } A \} \, dC \\
= \mathbb{E}_{C \in (SA)^{xk}} \mu \{ t \in S^{-1} \mid tC \text{ looks like } A \}.
\]

(We may apply Fubini-Tonelli here because, as $A$ is open, the set $\{ t \in G \mid tC \text{ looks like } A \}$ is open for any $C$, and therefore its intersection with $S^{-1}$ is measurable.) Thus we have shown

\[
\mathbb{E}_{C \in (SA)^{xk}} \mu \{ t \in S^{-1} \mid tC \text{ looks like } A \} \geq \frac{\mu(S^{-1})}{2K^k}.
\]

It follows that there is some $C_0 \in (SA)^{xk}$ for which the set

\[
T_{C_0} = \{ t \in S^{-1} \mid tC_0 \text{ looks like } A \}
\]

is a

31
has size at least $\mu(S^{-1})/2K^k$. In other words, $C_0$ has many translates that look like $A$, which is to say, there is a large set $T_{C_0} \subseteq S^{-1}$ of translates $t$ for which
\[
\|\mu_{C_0} \ast 1_B - 1_A \ast 1_B\|_2^2 < \frac{2}{k} \mu(A)^2 \mu(B).
\]
But since
\[
\|\mu_{C_0} \ast 1_B - 1_A \ast 1_B\|_2^2 = \int_G |\mu_{C_0} \ast 1_B(x) - 1_A \ast 1_B(x)|^2 \, dx
\]
\[
= \int_G |\mu_{C_0} \ast 1_B(t^{-1}x) - 1_A \ast 1_B(x)|^2 \, dx
\]
\[
= \int_G |\mu_{C_0} \ast 1_B(x) - 1_A \ast 1_B(tx)|^2 \, dx
\]
we have $\|\mu_{C_0} \ast 1_B(x) - 1_A \ast 1_B(tx)\|_2^2 < \frac{2}{k} \mu(A)^2 \mu(B)$. In English: many translates of $1_A \ast 1_B$ are close to $\mu_{C_0} \ast 1_B$. It will not be surprising, in view of the triangle inequality, that many translates are close to each other. Indeed, for any $t_1, t_2 \in T_{C_0}$ we have
\[
\|1_A \ast 1_B(t_1 t_2^{-1}x) - 1_A \ast 1_B(x)\|_2 = \|1_A \ast 1_B(t_1 x) - 1_A \ast 1_B(t_2 x)\|_2
\]
\[
\leq \|1_A \ast 1_B(t_1 x) - \mu_{C_0} \ast 1_B(x)\|_2
\]
\[
+ \|\mu_{C_0} \ast 1_B(x) - 1_A \ast 1_B(t_2 x)\|_2
\]
\[
< 2\sqrt{\frac{2}{k}} \mu(A) \sqrt{\mu(B)}
\]
so that $\|1_A \ast 1_B(tx) - 1_A \ast 1_B(x)\|_2^2 < \frac{8}{k} \mu(A)^2 \mu(B)$ holds for any $t \in T_{C_0} T_{C_0}^{-1}$. Setting $k = \lceil 8/\epsilon^2 \rceil$ and $T = T_{C_0}$, we are done.

5.2 Finding Small Neighborhoods

Next, we will use the preceding proposition to find iterated product sets (almost) inside sets of small doubling; we’ll need the following simple lemma on convolutions of such sets.

**Lemma 5.2** (Small doubling implies large convolution). Suppose $A \subseteq G$ satisfies $\mu(A^2) \leq K \mu(A)$. Then $\|1_A \ast 1_A\|_2^2 \geq \mu(A)^3/K$.  

32
Proof. This is a simple application of Cauchy-Schwarz:

\[ \int_G |1_A * 1_A|^2 \, dx = \int_{AA} |1_A * 1_A(x)|^2 \, dx \]
\[ \geq \frac{1}{\mu(A^2)} \left( \int_{A^2} \mu(A \cap xA^{-1}) \, dx \right)^2 \]
\[ = \frac{1}{\mu(A^2)} (\mu(A))^2 \]
\[ \geq \frac{\mu(A)^3}{K}. \]

\[ \square \]

**Theorem 5.3** (Small open neighborhoods). Suppose \( A \subseteq G \) has \( \mu(A^2) \leq K \mu(A) \). Then for each integer \( m \geq 1 \) there exists a symmetric subset \( N \subseteq A^{-1}A \) such that \( \mu(N) \geq \mu(A^{-1})/2K^{8Km^2} \) and \( N^m \subseteq A^2A^{-2} \). Moreover if \( A \) is open, then \( N \) may be taken to be open, at the cost of obtaining only the weaker containment \( N \subseteq (A^{-1}A)^2 \) and the weaker size bound \( \mu(N) \geq \mu(A^{-1})/2K^{32Km^2} \).

Proof. Apply the previous proposition with \( B = S = A \) and \( \varepsilon = 1/m\sqrt{K} \); we are assured the existence of a set \( T \subseteq A^{-1} \) of measure \( \mu(T) \geq \mu(A^{-1})/2K^{8Km^2} \) such that for every \( t \in TT^{-1} \) we have \( \|1_A * 1_A(t \cdot) - 1_A * 1_A\|_2 \leq \mu(A)^3/Km^2 \). Set \( N = TT^{-1} \subseteq A^{-1}A \); from the triangle inequality and translation invariance, we get \( \|1_A * 1_A(t'x) - 1_A * 1_A(x)\|_2 \leq \mu(A)^3/K \) for each \( t' \in N^m \). We claim this implies the supports of \( 1_A * 1_A(t'x) \) and \( 1_A * 1_A(x) \) overlap in a set of positive measure.

Indeed, if \( f, g \in L^2(G) \) have almost disjoint supports, then \( \|f - g\|_2^2 = \|f\|_2^2 + \|g\|_2^2 \). In the case at hand, if the supports of \( 1_A * 1_A(t'x) \) and \( 1_A * 1_A(x) \) were almost disjoint, we’d have

\[ \|1_A * 1_A(t'x) - 1_A * 1_A(x)\|_2 = \|1_A * 1_A(t'x)\|_2^2 + \|1_A * 1_A(x)\|_2^2 \]
\[ = 2 \|1_A * 1_A(x)\|_2^2 \]
\[ \geq \frac{2}{K} \mu(A)^3 \]

33
by the previous lemma. But this is not the case. Thus the supports of $1_A \ast 1_A(t'x)$ and $1_A \ast 1_A(x) = t'A^2$ and $A^2$, respectively — must overlap; hence $t' \in A^2A^{-2}$ and $N^m \subseteq A^2A^{-2}$ as claimed.

To find open $N$ as in the statement we use the claim already proved, with $2m$ in place of $m$, to get $N' \subseteq A^{-1}A$ such that $\mu(N) \geq \mu(A^{-1})/2K^{[32Km^2]}$ and $(N')^{2m} \subseteq A^2A^{-2}$. Since $N'$ is constructed as $TT^{-1}$ for a set $T$ of positive measure, the Steinhaus theorem implies that $N'$ contains an open neighborhood of $1$. Let $O$ be such a neighborhood, and define $N = N'O$, which is also open. Then $N' \subseteq N \subseteq (N')^2$, from which the claims of the theorem follow.

The proof of the preceding theorem found in [7] can be seen as a non-iterative version of Sanders’ density increment from [18]. In [7] the authors simply skip directly to the point where the increments would stop.

The previous theorem talks about sets of small doubling. If we upgrade the hypotheses to say that $A$ is an approximate group, we can ensure that $N$ too is an approximate group. To prove this, we need two simple lemmas, which enable us to turn small growth information (such as small quintupling) into covering information of the kind required in the definition of approximate group.

**Lemma 5.4.** Let $B, C \subseteq G$ be multiplicative sets and $c_0 \in C$. There are finite sets $X, Y \subseteq C$ such that: $C \subseteq XBB^{-1}$ and $C \subseteq BB^{-1}Y$; $c_0 \in X$ and $c_0 \in Y$; $|X| \leq \mu(CB)/\mu(B)$ and $|Y| \leq \mu(C^{-1}B)/\mu(B)$.

**Proof.** Let $X \subseteq C$ be maximal with respect to the properties “$c_0 \in X$” and “$\{xB \mid x \in X\}$ is a family of disjoint sets”. Since $XB \subseteq CB$ and $\mu(\cup_{x \in X}xB) = |X| \mu(B)$ we have $|X| \leq \mu(CB)/\mu(B)$. Moreover, for every $c \in C$ there is $x \in X$ such that $cB \cap xB \neq \emptyset$. It follows that $c \in xBB^{-1}$, whence $C \subseteq XBB^{-1}$.

Using this result with $C^{-1}$ and $c_0^{-1}$ in place of $C$ and $c_0$ we find a finite set $X \subseteq C^{-1}$, containing $c_0^{-1}$, with at most $\mu(C^{-1}B)/\mu(B)$ elements, and such that $C^{-1} \subseteq XBB^{-1}$. Taking inverses we get $C \subseteq BB^{-1}X^{-1}$, and setting $Y = X^{-1}$ we are done. \qed
Lemma 5.5. Suppose \( A \subseteq G \) is symmetric and satisfies \( \mu(A^5) \leq K \mu(A) \). Then \( A^2 \) is a \( 2K \)-approximate group.

Proof. It is clear that \( A^2 \) is symmetric and \( 1 \in A^2 \). Using the previous lemma we obtain finite sets \( X \) and \( Y \) with at most \( \mu(A^4)/\mu(A) \leq K \) elements each, and such that \( A^4 \subseteq XA^2 \) and \( A^4 \subseteq A^2Y \). Thus \( A^4 \subseteq (X \cup Y)A^2 \) and \( A^4 \subseteq A^2(X \cup Y) \), concluding the proof. \( \square \)

Corollary 5.6. If \( A \subseteq G \) is a \( K \)-approximate group and \( m \geq 1 \) is an integer, then there is a \( 2K^3 \)-approximate group \( S \subseteq A^4 \) of measure at least \( \mu(A)/2K^{[32K(2m+3)^2]} \) such that \( S^m \subseteq A^4 \).

Proof. By Theorem 5.3 there is open \( N \subseteq A^4 \) of measure at least \( \mu(A)/2K^{[32K(2m+3)^2]} \) such that \( N^{2m+3} \subseteq A^4 \). Take \( S = N^2 \), which clearly satisfies \( S^m \subseteq A^4 \). Then we have

\[
\mu(N^5) \leq \mu(N^{2m+3}) \leq \mu(A^4) \leq K^3 \mu(A)
\]

and the previous lemma shows that \( S \) is a \( 2K^3 \)-approximate group. \( \square \)

The next result we aim for, Theorem 5.10, is a strengthening of Theorem 5.3; we will find sets \( \tilde{N} \subseteq A^{-1}A \) of which the iterated products lie in \( A^2A^{-2} \) even upon conjugation by elements of \( A^2A^{-2} \). The rough idea is to look at intersections of the form \( \bigcap_{x} x^{-1}Nx \) where \( N \) is the set guaranteed by the previous result and the \( x \) are taken from \( A^4 \). (We are being deliberately vague about the values of \( x \) over which the intersection is taken, though.) This strategy is motivated by a well-known trick from group theory, where if \( H, K \) are subgroups of \( G \), one obtains a subgroup of \( K \) which is normalized by \( H \) by taking the intersection \( \bigcap_{h \in H} h^{-1}Kh \). Carrying this strategy through will require some care because we want to make sure the set \( \bigcap_{x} x^{-1}Nx \) isn’t too small. This will involve picking just a few choice \( x \), so only a few sets \( x^{-1}Nx \) are being intersected; and making sure that each additional \( x^{-1}Nx \) doesn’t destroy too much of the intersection.

Picking these few \( x \) is analogous to the fact that, in exact group theory, the intersection \( \bigcap h^{-1}Kh \) gives the same subgroup whether \( h \) runs through all of \( H \) or just through a set of coset representatives for \( K \) mod \( H \). Fortunately, this is achieved by Lemma 5.4 as we
shall see. Making sure each $x^{-1}N x$ doesn’t do too much damage will require two additional lemmas.

The first controls the behavior of the modular function: recall that this is the function $\Delta : G \to (0, \infty)$ defined by $\Delta(g) = \mu(E g)/\mu(E)$ for any $E \subseteq G$ of nonzero measure. That $\Delta$ is well-defined follows from uniqueness of Haar measure, and it is easy to see that $\Delta$ is in fact a group morphism from $G$ to the group of positive reals under multiplication. A Haar measure is called unimodular if $\Delta(G) = \{1\}$; many theorems and proofs about finite approximate groups implicitly rely on unimodularity of counting measure, a luxury we do not have, hence the need for the next lemma.

**Lemma 5.7 (Modular function bound).** If $A \subseteq G$ has $\mu(A^2) \leq K \mu(A)$, then $\Delta(a) \leq K$ for every $a \in A$. If moreover $A$ is symmetric, then $\Delta(a) \geq 1/K$.

**Proof.** Since $\mu(A) > 0$ we have

$$\Delta(a) = \frac{\mu(A a)}{\mu(A)} \leq \frac{\mu(A^2)}{\mu(A)} \leq K.$$ 

On the other hand $\Delta$ is a group morphism, so $\Delta(a^{-1}) = 1/\Delta(a) \geq 1/K$. As $a$ was arbitrary and $A = A^{-1}$, the claim follows. \qed

The second lemma furnishes the inductive step to the result we will actually require, Corollary 5.9. It concerns finding sets of the form $CC^{-1}$, with $C$ large, inside intersections of the form $AA^{-1} \cap BB^{-1}$.

**Lemma 5.8.** Suppose $A$ is symmetric, $\mu(A^2) \leq K \mu(A)$ and $A_1, A_2 \subseteq A$ are open pre-compact. Then $A_1 A_1^{-1} \cap A_2 A_2^{-1}$ contains $BB^{-1}$ for some open $B \subseteq A$ of measure at least

$$\frac{1}{K^2} \frac{\mu(A_1) \mu(A_1^{-1})}{\mu(A)}.$$ 

**Proof.** Given $x \in G$, set $B_x = A_1 \cap A_2 x$; this is an open set. Then

36
\[ B_x B^{-1}_x = (A_1 \cap A_2 x) \cdot (A_1^{-1} \cap x^{-1} A_2^{-1}) \]
\[ \subseteq (A_1 A_1^{-1}) \cap (A_2 x \cdot x^{-1} A_2^{-1}) \]
\[ = A_1 A_1^{-1} \cap A_2 A_2^{-1}. \]

It remains to show that for some \( x \) the set \( B_x \) has the required size. We achieve this by showing that the expected size for random \( x \) is large. By Fubini-Tonelli we have

\[
\int_G \mu(B_x) \, dx = \int_G \int_G 1_{A_1}(y) 1_{A_2 x}(y) \, dy \, dx
\]
\[ = \int_G 1_{A_1}(y) \int_G 1_{A_2 x}(y) \, dx \, dy
\]
\[ = \int_G 1_{A_1}(y) \int_G 1_{A_2^{-1} y}(x) \, dx \, dy
\]
\[ = \int_G 1_{A_1}(y) \mu(A_2^{-1} y) \, dy
\]

The last integrand is supported on \( A_1 \subseteq A \), and by Lemma 5.7 we have \( \mu(A_2^{-1} y) \geq \mu(A_2^{-1})/K \) for \( y \in A_1 \). Thus

\[
\int_G \mu(B_x) \, dx \geq \int_G 1_{A_1}(y) \cdot \frac{1}{K} \mu(A_2^{-1}) \, dy
\]
\[ = \frac{1}{K} \mu(A_1) \mu(A_2^{-1}).
\]

Since \( A_1 \cap A_2 x \) is non-empty only if \( x \in A_1 A_2^{-1} \subseteq A^2 \), the expected value of \( \mu(A_1 \cap A_2) x \) on \( AA \) is

\[
\frac{1}{\mu(A^2)} \int_G \mu(A_1 \cap A_2 x) \, dx \geq \frac{1}{\mu(A^2)} \frac{1}{K} \mu(A_1) \mu(A_2^{-1})
\]
\[ \geq \frac{1}{K^2} \frac{\mu(A_1) \mu(A_2^{-1})}{\mu(A)}.
\]

It follows that for some \( x \in A_1 A_2^{-1} \) we have \( \mu(B_x) \geq \mu(A_1) \mu(A_2^{-1})/K^2 \mu(A) \), as claimed. \( \square \)

**Corollary 5.9.** Suppose \( A \) is symmetric and \( A_1, \ldots, A_k \subseteq A \) are open precompact. Then \( \bigcap_{1 \leq i \leq k} A_i^{-1} \) contains \( B^2 \) for some open \( B \subseteq A \) of measure at least \( \mu(A_1) \prod_{1 < i \leq k} \frac{\mu(A_i^{-1})}{K^2 \mu(A)} \).
Proof. We induct on \( k \), the claim being trivial for \( k = 1 \). Given \( A_1, \ldots, A_{k+1} \subseteq A \) open pre-compact, by induction hypothesis there is open \( B \subseteq A \) of measure at least \( \mu(A_1) \prod_{1 \leq i \leq k} \frac{\mu(A_i^{-1})}{K^2 \mu(A)} \) such that \( BB^{-1} \subseteq \bigcap_{1 \leq i \leq k} A_i A_i^{-1} \). Lemma 5.8 tells us that there is open \( C \subseteq A \) such that

\[
CC^{-1} \subseteq BB^{-1} \cap A_{k+1} A_{k+1}^{-1} \subseteq \bigcap_{1 \leq i \leq k+1} A_i A_i^{-1}
\]

and moreover \( C \) has measure

\[
\mu(C) \geq \mu(B) \frac{\mu(A_{k+1}^{-1})}{K^2 \mu(A)} = \mu(A_1) \prod_{1 \leq i \leq k+1} \frac{\mu(A_i^{-1})}{K^2 \mu(A)}.
\]

\( \square \)

**Theorem 5.10** (Small normal neighborhoods). Let \( A \) be a \( K \)-approximate group, \( S \subseteq A^4 \) a \( K' \)-approximate group with \( \mu(S) = \delta \mu(A) \), and \( m \geq 1 \) an integer. Then there is a \( (K')^4 O_{K,m,\delta}(1) \)-approximate group \( \tilde{S} \) with \( \mu(\tilde{S}) \gg_{K,m,\delta} \frac{1}{K} \mu(A) \) such that \( \tilde{S}^m A^4 \subseteq S^4 \).

Proof. By Theorem 5.3 there is an open symmetric set \( N \subseteq S^2 \) with

\[
\mu(N) \geq \frac{\mu(S)}{2K^O(Km^2)}
\]

such that \( N^{4m+4} \subseteq S^4 \) (where the implied constants in the \( O \)-notation, here and elsewhere, are absolute). Using Lemma 5.4 find a finite set \( X \subseteq A^4 \) with at most \( \mu(A^4 N)/\mu(N) \) elements such that \( 1 \in X \) and \( A^4 \subseteq XN^2 \). We can bound \( |X| \) in terms of \( K \) only by observing that \( A^4 N \subseteq A^4 S^2 \subseteq A^4(A^4)^2 = A^{12} \) and using the fact that \( \mu(A^{12}) \leq K^{11} \mu(A) = K^{11} \mu(S)/\delta \). Hence

\[
|X| \leq \frac{2}{\delta} K^O(Km^2).
\]

Define \( P = \bigcap_{x \in X} x^{-1} N^2 x \); note that \( P \) is a symmetric open set, and that \( xP x^{-1} \subseteq N^2 \) for every \( x \in X \). (In particular, \( P \subseteq N^2 \).)

Regarding each \( x^{-1} N^2 x \) as \( A_x A_x^{-1} \) where \( A_x = x^{-1} N \) we apply Lemma 5.9 to find an open set \( B \subseteq S \) such that \( P \supseteq BB^{-1} \) and

\[
\mu(B) \geq \mu(N) \prod_{x \in X \setminus \{1\}} \frac{\mu(N x)}{K^2 \mu(S)}.
\]
Since \( x \in A^4 \), Lemma 5.7 gives the bound \( \Delta(x) \geq 1/K^4 \), thus \( \mu(Nx) \geq \mu(N)/K^4 \) and

\[
\mu(B) \geq \mu(N) \prod_{x \in X \setminus \{1\}} \frac{\mu(N)}{K^6 \mu(S)} \geq K^6 \left( \frac{1}{K^6 \frac{1}{2K^O(Km^2)}} \right)^{|X|} \mu(S)
\]

Finally, since \( B^{-1} \subseteq S^{-1} = S \) and \( \mu(SS) \leq K'\mu(S) \), Lemma 5.7 gives \( \mu(BB^{-1}) \geq \mu(B)/K' \); hence \( P \), containing \( BB^{-1} \), must be large:

\[
\mu(P) \geq \frac{1}{K'} K^6 \left( \frac{1}{2K^O(Km^2)} \right)^{|X|} \mu(S) \geq \frac{1}{K'} K^6 2^{-\frac{2}{3}K^O(Km^2)} \left( K^{-O(Km^2)} \right)^{\frac{2}{3}K^O(Km^2)} \mu(S) \geq \frac{1}{K'} K^{-O(Km^2)} \mu(S)
\]

The unwieldy nature of the latter expression justifies our shortening it to \( \frac{1}{K'} \Omega_{K,m,\delta}(\mu(S)) \).

Finally, define \( \tilde{S} = P^2 \); it too has measure at least \( \frac{1}{K'} \Omega_{K,m,\delta}(\mu(S)) \), and we claim it has the desired properties. Let us check that \( (\tilde{S}^m)^4 \subseteq S^4 \): let \( a \in A^4 \) be arbitrary, and recalling that \( X \) was chosen so that \( A^4 \subseteq N^2 X \), pick \( x \in X, t \in N^2 \) such that \( a = tx \). We have

\[
a \tilde{S}^m a^{-1} = tx \tilde{S}^m x^{-1} t^{-1} = tx P^{2m} x^{-1} t^{-1} = t(xP x^{-1})^{2m} t^{-1} \subseteq N^2 (N^2)^{2m} N^2 = N^{4m+4} \subseteq S^4.
\]

Finally, we verify that \( \tilde{S} \) is a \( (K')^4 O_{K,m,\delta}(1) \)-approximate group. It is clearly open and symmetric, and in view of Lemma 5.5 it suffices to show that \( P \) has small quintupling. This is easy:

\[
\mu(P^5) \leq \mu(N^{10}) \leq \mu(S^4) \leq (K')^3 \mu(S) \leq (K')^4 O_{K,m,\delta}(1) \mu(P).
\]

\[\square\]

Remark. It is interesting to note how the size bound on the “small neighborhood” in Theorem 5.3, which is exponential in \( K \) and \( m \), is severely worsened — to triply exponential! — by the requirement that the neighborhood be approximately normalized by \( A^4 \). Note also that the argument needs \( A \) to be an approximate group, not just a set of small doubling, because we require \( A^{12} \) to have controlled size.
CHAPTER 6

Finding a Lie Model

The aim of this chapter is to find a useful morphism from an ultraproduct of $K$-approximate groups into a Lie group.

6.1 Building a Metric

We start with a definition.

**Definition 6.1** (Ultra approximate group). An *ultra $K$-approximate group* is an ultraproduct $A = \prod_{n \rightarrow \alpha} A_n$, where each $A_n$ is a $K$-approximate group in $G_n$. An *ultra approximate group* is an ultra $K$-approximate group for some $K$.

Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of approximate groups inside lch groups $(G_n)_{n \in \mathbb{N}}$. Let $\alpha$ be a non-principal ultrafilter on $\mathbb{N}$. We are going to endow the ultra approximate group $A = \prod_{n \rightarrow \alpha} A_n$ with a (pseudo)metric that makes it a locally compact space. Roughly speaking, the smaller and smaller balls of the pseudometric will be sets whose longer and longer iterated products still lie in $A$, but which are nevertheless somewhat large; the existence of such sets is guaranteed by Sanders-Croot-Sisask theory.

Local compactness of the ultraproduct will allow us, via Gleason-Yamabe theory, to model it closely by a Lie group; this allows us to extract a more regular object from the approximate groups, something we will call a *strong approximate group*. The extra regularity will be necessary to ensure the so-called escape norm to be defined on the approximate group has certain desirable properties. The escape norm, in turn, will be used to split the classification of open precompact approximate groups in general groups into a discrete case.
and a Lie case.

To define the pseudometric with which it all begins we employ a general construction due to G. Birkhoff and Kakutani independently. The construction takes as input a sequence of subsets of a group satisfying an appropriate nesting condition, and outputs a left-invariant pseudometric with respect to which the group operations are continuous. The next lemma shows that in ultra approximate groups one can find the required sequence of subsets.

**Lemma 6.2.** Let $A$ be an ultra approximate group. There is a sequence of ultra approximate groups $A_0 \supseteq A_1 \supseteq \cdots$ such that $A_0 = A^4$ and we have the product/normalization nesting property $(A_0^2)^{i+1} \subseteq A_i$. Moreover, $A$ can be covered by $O_{K,i}(1)$ left-translates of $A_i$.

**Proof.** Let $K$ be the approximate group parameter of $A$ and let $A_1, A_2, \ldots$ be $K$-approximate groups such that $A = \prod_{n \to \alpha} A_n$. Define $S_{n,0} = A_n$ and inductively pick $S_{n,i+1} \subseteq S_{n,i}$ to be an $O_{K,i}(1)$-approximate group such that $(S_{n,i+1}^8)^{i+1} \subseteq S_{n,i}^4$; such $S_{n,i+1}$ exists by Theorem 5.10, and moreover satisfy $\mu(S_{n,i}) \gg K,i \mu(A_n)$. Since the $S_{n,i}$ all contain 1, the nesting property implies in particular that $S_{n,i} \subseteq S_{n,i+1} \subseteq S_{n,i}^4$ for each $i$, whence every $S_{n,i}$ lies in $S_{n,0}^4 = A_n^4$.

Define $A_i = \prod_{n \to \alpha} S_{n,i}^4$; then the nesting condition is satisfied:

$$(A_0^2)^{i+1} = \prod_{n \to \alpha} (S_{n,i+1}^8)^{i+1} \subseteq \prod_{n \to \alpha} S_{n,i}^4 = A_i$$

Moreover, by Ruzsa covering (Lemma 5.4) and the lower bound on the measure of $S_{n,i}$, each $A_n$ can be covered by $O_{K,i}(1)$ left-translates of $S_{n,i}^2$, and therefore of $S_{n,i}^4$. By Łoś’s theorem it follows that $A$ can be covered by $O_{K,i}(1)$ translates of $A_i$, as desired. \qed

**Remark.** The “small normal neighborhoods” Theorem 5.10 could have been used more directly on the ultra approximate group $A$, without heed to its origin as an ultraproduct of approximate groups, to obtain a sequence $H_0 \supseteq H_1 \supseteq \cdots$ satisfying $H_0 = A^4$ and $(H_0^2)^{i+1} \subseteq H_i$. The reason we don’t proceed so directly, and instead analyze $A$ “coordinatewise” as it were, is that we need the nested sequence to itself consist of ultra approximate groups. Sanders-Croot-Sisask theory, while capable of finding appropriately nested $H_i$, cannot guarantee that they are themselves ultraproducts.
Proposition 6.3. Let $G$ be a group and $A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots$ a sequence of symmetric subsets satisfying $(A_{i+1}^2)^{A_{i+1}} \subseteq A_i$. There exists a bounded left-invariant pseudometric $d$ defined on $G$ such the $A_i$ constitute a neighborhood base at $1$, specifically

$$\{ g \in G \mid d(g, 1) < 2^{-i} \} \subseteq A_i \subseteq \{ g \in G \mid d(g, 1) \leq 2 \cdot 2^{-i} \},$$

and such that the group operations are continuous on $\langle A_0 \rangle$ with respect to $d$.

Proof. To define the pseudometric we will use the weaker nesting condition $A_{i+1}^2 \subseteq A_i$; the stronger condition in the statement of the lemma is only needed to show continuity of the group operations.

For each dyadic rational $q = 2^{-i_1} + \cdots + 2^{-i_k}$, where $0 < i_1 < \cdots < i_k$, define the product set $B_q = A_{i_k} \cdots A_{i_1}$. (The order of multiplication is important.) We will show that the $B_q$ are nested, i.e. $B_q \subseteq B_{q'}$ whenever $q \leq q'$. More specifically, we show that if $q = 2^{-i_1} + \cdots + 2^{-i_k}$ and $i_{k+1} \geq i_k$ then $B_q \subseteq B_{q+2^{-i_{k+1}}}$; clearly this suffices.

First note that, if $i_{k+1} > i_k$ and $q' = q + 2^{-i_{k+1}}$, then $B_{q'} = A_{i_k+1} B_q \supseteq B_q$. On the other hand, if $q' = q + 2^{-i_k}$, then

$$B_q \subseteq A_{i_k} B_q = A_{i_k} (A_{i_k} \cdots A_{i_1}) \subseteq A_{i_k-1} \cdots A_{i_1},$$

where the last inclusion follows from $A_{i_k} A_{i_k} \subseteq A_{i_k-1}$. If there is another occurrence of $A_{i_k-1}$ in the latter product, we use the nesting condition again; we do this until the sets in the product are all distinct, each time passing to a larger set. This process “simulates” the carries that happen when one adds $q + 2^{-i_k}$ and seeks to write the result as a sum of distinct powers of two. At the end we have obtained $B_{q'}$, and shown that $B_q \subseteq B_{q'}$ in this case as well.

Putting the two cases together we see that $B_q \subseteq B_{q'}$ whenever $q \leq q'$: for if the dyadic rational $q' - q$ is not zero, let $2^{-N}$ be the smallest power of 2 in its base-2 expansion. By adding $2^{-N}$ to $q$ an integer number of times, we eventually reach $q'$, and by the preceding analysis we pass to a larger $B_q$ at each step.

In possession of the nested sets $B_q$, it is tempting to make them the balls of our pseudometric. But this turns out to not quite work, so we take a more indirect approach: we
will use the $B_q$ to map the group $G$ into the normed space $\ell^\infty(G)$, and pull back the metric there.

Define a map $\psi : G \to [0, 1]$ by

$$\psi(x) = \inf \{ q \in \mathbb{Q} \mid x \in B_q \} \cup \{ 1 \}$$

Note that $\psi$ is zero precisely on the intersection of all the $A_i$, and 1 outside $\bigcup_{0 < q < 1} B_q$. Now we can define our pseudometric:

$$d(g, h) = \sup_{x \in G} |\psi(x) - \psi(h^{-1}gx)|$$

It is clear that $d(g, g) = 0$ and that $d(ag, ah) = d(g, h)$. Symmetry follows from the variable change $x \mapsto g^{-1}hx$ inside the sup operator. For the triangle inequality, take any $g, h, k \in G$; we have

$$d(g, k) = \sup_{x \in G} |\psi(x) - \psi(k^{-1}gx)|$$

$$= \sup_{x \in G} |\psi(x) - \psi(h^{-1}gx) + \psi(h^{-1}gx) - \psi(k^{-1}gx)|$$

$$\leq \sup_{x \in G} |\psi(x) - \psi(h^{-1}gx)| + \sup_{x \in G} |\psi(h^{-1}gx) - \psi(k^{-1}gx)|$$

Applying the variable change $x \mapsto g^{-1}hx$ inside the second sup operator, it becomes

$$\sup_{x \in G} |\psi(x) - \psi(h^{-1}gx)| + \sup_{x \in G} |\psi(x) - \psi(k^{-1}hx)| = d(g, h) + d(h, k).$$

Our next task is to show that

$$\{ g \in G \mid d(g, 1) < 2^{-i} \} \subseteq A_i \subseteq \{ g \in G \mid d(g, 1) \leq 2 \cdot 2^{-i} \}.$$

Let $g \in G$ be such that $d(g, 1) < 2^{-i}$; then

$$|\psi(1) - \psi(g)| \leq \sup_{x \in G} |\psi(x) - \psi(gx)| = d(g, 1) < 2^{-i}.$$  

But $\psi(1) = 0$, whence $\psi(g) < 2^{-i}$, which means there is $q < 2^{-i}$ such that $g \in B_q$. By the nesting property of the $B_q$, we have $B_q \subseteq B_{2^{-i}} = A_i$, and that’s the first inclusion.

For the second, suppose $g \in A_i = B_{2^{-i}}$; we must show that $|\psi(x) - \psi(gx)| \leq 2 \cdot 2^{-i}$ for every $x \in G$.  

43
Suppose first that $x$ lies outside $B_{1/2}$. Then $gx$ lies outside $B_{1/2}$; for if it were the case that $gx \in B_{1/2} = A_{i-1} \cdots A_2 A_1$, we’d have $x \in A_i A_{i-1} \cdots A_2 A_1 = B_{1/2}$. Moreover, by definition of $\psi$ we have $\psi(x) \geq 1 - 2^{-i}$ and $\psi(gx) \geq 1 - 2^{-i}$. Since $\psi(x), \psi(gx) \leq 1$ we obtain $|\psi(x) - \psi(gx)| \leq 2 \cdot 2^{-i}$.

It remains to see what happens in case $x \in B_{1/2}$. Let $1 \leq m < 2^i$ be the smallest integer such that $x \in B_{m/2}$; then, by observations made during the analysis of the $B_q$, we have $gx \in A_i B_{m/2} \subseteq B_{(m+1)/2}$. Conversely, we can’t have $gx \in B_{(m-2)/2}$, or that would imply $x \in A_i B_{(m-2)/2} \subseteq B_{(m-1)/2}$, contradicting the definition of $m$.

From all this it follows that $(m - 1)2^{-i} \leq \psi(x) \leq m2^{-i}$ and $(m - 2)2^{-i} \leq \psi(gx) \leq (m + 1)2^{-i}$, whence $|\psi(x) - \psi(gx)| \leq 2 \cdot 2^{-i}$, as desired.

Finally, let us show that the group operations are continuous on $\langle A_0 \rangle$ with respect to $d$. Let $(g_n)$ be a sequence in $G$ converging to $g \in \langle A_0 \rangle$; thus $g \in A_0^N$ for some $N$. This means $d(g_n, g) \to 0$, which by left-invariance is equivalent to $d(g^{-1}g_n, 1)$ and $d(g_n^{-1}g, 1)$ going to zero. Since the $A_i$ form a neighborhood base at 1, this in turn is equivalent to the following: for every $i$ and large enough $n$ depending on $i$, both $g^{-1}g_n$ and $g_n^{-1}g$ lie in $A_i$.

Now we seek to establish that $g_n^{-1} \to g^{-1}$, i.e. that for every $i$ and large enough $n$, we have $gg_n^{-1} \in A_i$. Well, simply note that $gg_n^{-1} = g(g_n^{-1}g)g^{-1}$, so if $i \geq N$ and $g_n^{-1}g \in A_{i+1}$ then $gg_n^{-1} \in (A_{i+1})^A_{i+1} \subseteq A_i$. For continuity of the product, let $(h_n)$ be another sequence in $G$ converging to $h \in \langle A_0 \rangle$, say $h \in A_0^M$; our task is to show that $g_nh_n \to gh$. To that end, write $(gh)^{-1}g_nh_n = h^{-1}(g^{-1}g_n)(h_nh^{-1})h$; it is clear that, if $i \geq M$ and $g^{-1}g_n, h_nh^{-1} \in A_{i+1}$, then $(gh)^{-1}g_nh_n \in (A_{i+1})^A_{i+1} \subseteq A_i$, and we are done. □

### 6.2 Local Compactness

The next proposition establishes the last important fact about the pseudometric constructed from the nested sequence given by Sanders-Croot-Sisask theory.

**Proposition 6.4.** Let $A = \prod_{n \to \alpha} A_n$ be an ultra approximate group, $(A_i)$ a sequence such as in Lemma 6.2, and $d$ a pseudometric defined from it as in Proposition 6.3. Then $\langle A \rangle$
Proof. In view of the preceding results, and the fact that $A^4$ is a neighborhood of 1, it suffices to prove that $\langle A \rangle$ is complete and $A^4$ is totally bounded.

Total boundedness is simple. Let $\varepsilon > 0$ be arbitrary and $B(1, \varepsilon)$ the $d$-ball of radius $\varepsilon$ centered at 1. Pick $k$ large enough that $2^{-k} < \varepsilon/2$, so that \( \{ g \in \langle A \rangle \mid d(g, 1) \leq 2 \cdot 2^{-k} \} \subseteq B(1, \varepsilon) \). The inclusions proved in Proposition 6.3 imply $A_k \subseteq B(1, \varepsilon)$, and by left-invariance we have $x A_k \subseteq B(x, \varepsilon)$ for any $x \in \langle A \rangle$. Thus it suffices to show that $A^4$ can be covered by finitely many left-translates of $A_k$. But that was established in Lemma 6.2.

It remains to prove completeness. In a sense, completeness is the one property that does not depend on the fact that $A$ is an ultraproduct of $K$-approximate groups; rather it is a manifestation of a more general phenomenon, the so-called countable saturation of ultraproducts. Roughly speaking, if it is possible to define an object $X$ by its relation to a countable number of other objects, and it is possible for $X$ to exist in some model, then $X$ exists in an ultraproduct. In our case, the limit of a Cauchy sequence can be defined by its relationship to the (countably many) terms of the sequence; and it certainly exists in some model, namely the metric completion. Therefore it already exists in the ultraproduct. Here are the details.

Let $(x_i)$ be a Cauchy sequence in $\langle A \rangle$, and, passing to a subsequence if necessary, assume it satisfies $d(x_i, x_j) < 2^{-i}$ whenever $i \leq j$; this means $x_j \in x_i A_i$ for all $i \geq j$, and in particular that for every integer $M$ we have $\bigcap_{i \leq M} x_i A_i \neq \emptyset$. Let $y_M = (y_{M,n})_{n \in \mathbb{N}}$ be an element of this intersection, where $y_{M,n} \in A_n$. The $y_M$ are approximations to the limit we seek; in the language of model theory, they are witnesses to the fact that “limit of the Cauchy sequence $(x_i)$” is a consistent type. To extract the actual limit from the approximations, which is a sequence of sequences, we do the usual thing: diagonalize.

The statement $y_M \in \bigcap_{i \leq M} x_i A_i$ is equivalent to: $y_{M,n} \in \bigcap_{i \leq M} x_{i,n} A_{i,n}$ for all $n$ in some $\alpha$-large set $\Sigma_M$. By replacing each $\Sigma_M$ with $\Sigma_M \setminus \{M\}$ (which preserves largeness since $\alpha$ is non-principal) and after that replacing $\Sigma_M$ with $\Sigma_1 \cap \cdots \cap \Sigma_M$, we may assume that
\( \Sigma_1 \supseteq \Sigma_2 \supseteq \cdots \) and that no integer is in infinitely many \( \Sigma_M \). Thus for each \( n \) there is a smallest index \( M(n) \) such that \( n \notin \Sigma_{M(n)+1} \). Now define \( x \in \langle A \rangle \) coordinatewise by \( x_n = y_{M(n),n} \) in case \( M(n) \geq 1 \), and arbitrarily if \( M(n) = 0 \). We claim that \( x \in \bigcap_{i \in \mathbb{N}} x_i A_i \); to establish this, we must show that \( x \in x_j A_j \) for every \( j \), which is to say, that for \( \alpha \)-many \( n \) it is the case that \( x_n \in x_{j,n} A_{j,n} \). The set of such \( n \) actually includes all of \( \Sigma_j \), and is therefore \( \alpha \)-large. Indeed, if \( n \in \Sigma_j \) then \( M(n) \geq j \), whence \( x_n = y_{M(n),n} \in \bigcap_{i \leq M(n)} x_{i,n} A_{i,n} \subseteq x_{j,n} A_{j,n} \).

Finally, from \( x \in \bigcap_{i \in \mathbb{N}} x_i A_i \) it follows that \( x^{-1} x_i \in A_i \) for each \( i \), and using the inclusions between the sets \( A_i \) and \( d \)-balls from Proposition 6.3 we have \( d(x_{i+1}, x) \leq 2^{-i} \). Thus \( x = \lim x_i \) and \( \langle A \rangle \) is complete. \( \square \)

6.3 Local Compactness Gives Lie Model

After the technical lemmas in the previous section, let us stop and take stock. We started with a sequence \((A_n)\) of \( K \)-approximate groups and took its ultraproduct, \( A \), which we call an ultra \( K \)-approximate group. Using Sanders-Croot-Sisask theory we constructed a sequence \((A_i)\) of appropriately nested ultra approximate subgroups of \( A \), where moreover each \( A_i \) is large. The Birkhoff-Kakutani construction then turned this sequence into a left-invariant pseudometric on \( \langle A \rangle \), with respect to which the \( A_i \) form a neighborhood base at 1 and the group operations are continuous; thus \( \langle A \rangle \) is made into a topological group. Moreover, the largeness of the \( A_i \) makes the group locally compact.

Recall that one of our main goals is to construct a morphism from \( \langle A \rangle \) to a Lie group that is faithful in several ways. That goal is within our grasp; we’ve made \( \langle A \rangle \) a locally compact group, and Gleason-Yamabe theory can turn locally compact groups into Lie groups. Let us now be precise about the properties we require of the “faithful morphism”, and construct it.

**Definition 6.5** (Good models). Let \( A = \prod_{n \to \alpha} A_n \) be an ultra approximate group, \( G \) a topological group, and \( \pi : \langle A \rangle \to G \) a surjective group morphism. We say \( \pi \) is a **good model** of \( A \) if the following hold:

46
(i) There is an open neighborhood of the identity $U_0 \subseteq G$ such that $\pi^{-1}(U_0) \subseteq A$ (and consequently $U_0 \subseteq \pi(A)$);

(ii) $\pi(A)$ is contained in a compact set;

(iii) If $F \subseteq U \subseteq G$, with $U$ open and $F$ compact contained in the image of $\pi$, then there is $A' = \prod_{n \to \alpha} A'_n$ with $A'_n \subseteq A_n$ such that each $A'_n$ is open and $\pi^{-1}(F) \subseteq A' \subseteq \pi^{-1}(U)$.

In the remainder of this section we will find good models with lots of extra structure.

Theorem 6.6 (Locally Compact Model). Let $A$ be an ultra approximate group. There is a locally compact metric group $G$ and a good model $\pi : \langle A \rangle \to G$ of $A^4$.

Proof. Let $(A_i)$ be a nested sequence as in Lemma 6.2 and endow $\langle A \rangle$ with a pseudometric $d$ as in Proposition 6.3. By Proposition 6.4 it makes $\langle A \rangle$ a locally compact group. To obtain the locally compact metric group $G$ we simply quotient out by the (normal) subgroup of elements at distance zero from $1$. Recall that this subgroup is exactly the intersection of all the $A_i$.

We now check that this quotienting map, call it $\pi$, satisfies the definition of good model; obviously it is surjective.

For the open neighborhood $U_0$ required by condition (i) we may take any open ball around $1$ of radius less than $1/2$; indeed, the inclusions proved in Proposition 6.3 imply that for such $U_0$ we have $\pi^{-1}(U_0) \subseteq A_1$, and the latter in turn is contained in $A^4$.

Condition (ii) requires precompactness of $\pi(A^4)$. We show that its (topological) closure is compact. Since $A^4$ is totally bounded and $\pi$ is an isometry, $\pi(A^4)$ is totally bounded as well; and it is a simple exercise to see that if a set is totally bounded, so is its closure. Moreover $G$ is complete, so the closure of $\pi(A^4)$ is complete.

Condition (iii) requires that, if $F$ is compact, $U$ is open, and $F \subseteq U$, there is $A' = \prod_{n \to \alpha} A'_n$ for some $A'_n \subseteq A^4_n$ such that $\pi^{-1}(F) \subseteq A' \subseteq \pi^{-1}(U)$. This is where we will take advantage of the fact that the sequence $(A_i)$ consists of ultra approximate groups.
First, let \( r = d(F, G \setminus U) \), which is strictly positive by compactness of \( F \). Let \( U' \) be the open ball of radius \( r/2 \) centered at 1; then \( F(U')^2 \subseteq U \). In fact, \( d(F(U')^2, G \setminus U) > 0 \); for any \( x \in F \), \( u, u' \in U' \), and \( y \in G \setminus U \), we have

\[
d(xuu', y) \geq d(x, y) - d(x, xu) - d(xu, xu') \geq r - d(1, u) - d(1, u') > r - \frac{r}{2} - \frac{r}{2} = 0.
\]

Note that \( U' \), like any ball centered at 1, is symmetric. Since \( \pi^{-1}(U') \) is an open neighborhood of 1 in \( \langle A \rangle \) and the \( A_i \) form a neighborhood base at 1, there is an integer \( k \) such that \( A_k \subseteq \pi^{-1}(U') \). We claim that the required intermediate set \( A' \) can be taken to be a finite union of translates of \( A_k \).

Since \( \pi \) is an isometry, it maps open balls to open balls, i.e. it is an open map. Thus \( \pi(A_k) \) has non-empty interior, from which it follows that a finite number of its translates suffice to cover \( F \), say \( F \subseteq \bigcup_{x \in X} x \cdot \pi(A_k) \) for a finite set \( X \). We may assume that only translates which intersect \( F \) are included in the union (otherwise discard them), and hence that each shift \( x \) lies in \( F \cdot \pi(A_k)^{-1} \subseteq F(U')^{-1} = FU' \subseteq F(U')^2 \subseteq U \subseteq U_0 \subseteq \pi(A^4) \). Thus each \( x \) is of the form \( \pi(a) \) for some \( a \in A^4 \), and we can write \( F \subseteq \bigcup_{a \in Y} \pi(aA_k) \) for some finite set \( Y \). Taking inverse \( \pi \) images we obtain \( \pi^{-1}(F) \subseteq \bigcup_{a \in Y} aA_k \).

Finally, note that \( \pi(A_k) \subseteq U' \) and \( x \in FU' \) together imply that \( x \cdot \pi(A_k) \subseteq F(U')^2 \subseteq U \), and, again taking inverse images, \( \bigcup_{a \in Y} aA_k \subseteq \pi^{-1}(U) \).

Thus \( A' = \bigcup_{a \in Y} aA_k \) satisfies the necessary inclusions, and is also an ultraproduct of open sets, namely \( A' = \prod_{n \to \alpha} \left( \bigcup_{a \in Y} a_n A_{k,n} \right) \).

The final goal of this section is to obtain a good Lie model for an ultra approximate group.

The next three lemmas provide ways to get new good models from old without increasing the covering parameter \( K \) too much.

**Lemma 6.7.** Let \( G \) be a group, \( A \subseteq G \) a \( K \)-approximate group, and \( G' \leq G \) an open subgroup. Then \( A' = A^2 \cap G' \) is a \( K^3 \)-approximate group and \( A^4 \cap G' \) can be covered by at most \( K^3 \) left translates of \( A' \).
Proof. It is clear that $A'$ is open, symmetric and contains 1, and moreover that $(A')^2 \subseteq A^4 \cap G'$; thus it suffices to prove the last claim.

We know that $A^4$ can be covered by $K^3$ left translates of $A$. Now, if such a translate $xA$ does not intersect $G'$, discard it. If it does intersect $G'$, say at $x'$, then there is $a \in A$ such that $xa = x'$; thus $x = x'a^{-1}$ and $xA \subseteq x'A^2$. From this it follows that $xA \cap G' \subseteq x'(A^2 \cap G') = x'A'$. Doing this for each of the $K^3$ translates that cover $A^4$ we see that $A^4 \cap G'$ is contained in at most $K^3$ translates of $A'$.

**Lemma 6.8 (Good model into subgroup).** Let $A$ be an ultra $K$-approximate group, $G$ a locally compact metric group, and $\pi : \langle A \rangle \to G$ a good model of $A^4$. Let $G'$ be an open subgroup of $G$. Then there is an open subgroup $G'' \subseteq G'$ and a large ultra $K^6$-approximate group $\tilde{A} \subseteq A^4$ such that the restriction $\pi : \langle \tilde{A} \rangle \to G''$ is a good model of $\tilde{A}$.

Proof. Let $\tilde{A} = A^4 \cap \pi^{-1}(G')$ and $G'' \subseteq G'$ be the subgroup generated by $\pi(\tilde{A})$; the preceding lemma, applied to $A^2$ (which is a $K^2$-approximate group), says that $\tilde{A}$ is a $K^6$-approximate group.

Let $U_0$ be a neighborhood of 1 in $G$ such that $\pi^{-1}(U_0) \subseteq A^4$, as required by condition (i) of $\pi : \langle A \rangle \to G$ being a good model. Define $U_0''$ to be $U_0 \cap G'$; then $\pi^{-1}(U_0'') = \pi^{-1}(U_0) \cap \pi^{-1}(G') \subseteq \tilde{A}$, establishing condition (i) of good models. Moreover, surjectivity of $\pi$ implies that $U_0'' \subseteq \pi(\tilde{A})$, so the latter has non-empty interior, which means $G''$ is open.

Condition (ii) is straightforward: $\pi(\tilde{A}) \subseteq \pi(A^4)$; the latter is contained in a compact subset $K$ of $G$; $K \cap G''$ is compact in $G''$; and $\pi(\tilde{A}) \subseteq K \cap G''$.

For condition (iii), take $F \subseteq U \subseteq G''$ with $F$ compact and $U$ open. By condition (iii) applied to $\pi$ as good model of $A^4$ we can find an ultraproduct $A' \subseteq A^4$ such that $\pi^{-1}(F) \subseteq A' \subseteq \pi^{-1}(U)$. But the latter inclusion implies $A' \subseteq \tilde{A}$.

Next we claim that $A$ can be covered by finitely many translates of $\tilde{A}$. Indeed, since $\pi(A)$ is contained in a compact set we can find a finite set $Y \subseteq G'$ such that $\pi(A) \subseteq \bigcup_{y \in Y} y \cdot U_0''$. Moreover, as $\pi$ is surjective, each $y \in Y$ is of the form $\pi(x)$ for some $x \in \langle A \rangle$. Choose
one such $x$ for each $y \in Y$ and collect them into a set $X \subseteq \langle A \rangle$, which lets us write $\pi(A) \subseteq \bigcup_{x \in X} \pi(x)U''_0$.

Taking $\pi$-inverse images we obtain

$$A \cdot \ker \pi \subseteq \bigcup_{x \in X} \pi^{-1}(\pi(x)U''_0) = \bigcup_{x \in X} x \cdot \pi^{-1}(U''_0) \subseteq \bigcup_{x \in X} x \cdot \tilde{A}$$

so $\tilde{A}$ is large, as claimed.

Finally, we need to show that $\tilde{A}$ is actually an ultra approximate group, i.e. that it is an ultraproduct. Let $F$ be a compact set such that $\pi(A^4) \subseteq F$, and hence $A^4 \subseteq \pi^{-1}(F)$. The subgroup $G'$ is open and therefore closed, so $F \cap G'$ is compact. By condition (iii) applied to $\pi$ as a good model of $A^4$ there is an ultraproduct $A'$ such that $\pi^{-1}(F \cap G') \subseteq A' \subseteq \pi^{-1}(G')$.

It follows that $\tilde{A} = A^4 \cap A'$, and we are done.

**Lemma 6.9 (Good model into quotient).** Let $A$ be an ultra approximate group, $G$ a locally compact metric group, and $\pi : \langle A \rangle \to G$ a good model of $A$. Let $U_0$ be the neighborhood of 1 in $G$ guaranteed by condition (i) and suppose $H$ is a compact normal subgroup of $G$ contained in $U_0$. Denote by $q$ the quotient map $G \to G/H$. Then $q \circ \pi : \langle A \rangle \to G/H$ is also a a good model of $A$.

**Proof.** First we show that $q$ is an open map. Let $U \subseteq G$ be open; the preimage of $q(U)$ is $UH$, which is a $q$-saturated open set; thus $q(UH) = q(U)$ is open in $G/H$. Now choose $U_1$ a neighborhood of 1 in $G$ such that $U_1H \subseteq U_0$ (say a small enough ball, as in the proof of Theorem 6.6). Then $q(U_1)$ works as the neighborhood of condition (i) in $G/H$. Indeed, we have the inclusions

$$\pi^{-1} \circ q^{-1}(q(U_1)) = \pi^{-1}(U_1H) \subseteq \pi^{-1}(U_0) \subseteq A.$$ 

In the other direction, from $U_1H \subseteq U_0$ and $U_0 \subseteq \pi(A)$ we see that $q(U_1) = q(U_1H) \subseteq q(U_0) \subseteq q \circ \pi(A)$.

Continuity of $q$ immediately implies that $q \circ \pi(A)$ is contained in a compact set, given that $\pi(A)$ is; this takes care of condition (ii).
To verify condition (iii), let \( F \subseteq U \subseteq G/H \) be arbitrary with \( F \) compact and \( U \) open in \( G/H \). Then \( q^{-1}(F) \subseteq q^{-1}(U) \) with \( q^{-1}(U) \) open and \( q^{-1}(F) \) closed. We claim that \( q^{-1}(F) \) is moreover compact.

Let \( K \) be a compact neighborhood of 1 in \( G \); then \( q(K) \) is a compact neighborhood of 1 in \( G/H \). Since \( F \) is compact, we can cover it with finitely many left translates of \( q(K) \); and \( q \) being surjective, any translate \( y \cdot q(K) \) is of the form \( q(xK) \) for some \( x \in G \). So we may write \( F \subseteq \bigcup_{i=1}^{N} q(x_i K) \). Taking inverse images we see that \( q^{-1}(F) \subseteq \bigcup_{i=1}^{N} x_i KH \). Now, \( KH \) is compact, and so is any finite union of its translates; \( q^{-1}(F) \) is then a closed subset of a compact set, and thus compact itself.

We are now in a position to invoke condition (iii) on the nested sequence \( q^{-1}(F) \subseteq q^{-1}(U) \subseteq G \): it says there is an ultraproduct \( A' \) such that \( q^{-1}(F) \subseteq \pi(A') \subseteq q^{-1}(U) \). Since \( q \) is surjective, it follows that \( F \subseteq q \circ \pi(A') \subseteq U \), and we are done. \( \square \)

Obtaining a good Lie model for an ultra approximate group is now a simple application of Gleason-Yamabe theory. We recall the main result [10, 23] here:

**Theorem 6.10 (Gleason-Yamabe).** Let \( G \) be a locally compact group. There exists an open subgroup \( G' \) of \( G \) with the following property: for any neighborhood of the identity \( U \subseteq G' \), there is a compact normal subgroup \( H \subseteq U \) such that \( G'/H \) is a connected Lie group.

Combining this with Theorem 6.6, and Lemmas 6.8 and 6.9, we obtain the main result of this section.

**Theorem 6.11 (Lie Model).** Let \( A \) be an ultra \( K \)-approximate group. There exists a large ultra \( K^6 \)-approximate group \( \hat{A} \subseteq A^4 \), a connected Lie group \( L \), and a morphism \( \pi : \langle \hat{A} \rangle \to L \) which is a good model for \( \hat{A} \).

**Proof.** Let \( \pi : \langle A \rangle \to G \) be a good model of \( A^4 \) as given by Theorem 6.6, with \( G \) locally compact. Let \( G' \) be an open subgroup of \( G \) as guaranteed by the Gleason-Yamabe result.

By Lemma 6.8 we can find an open subgroup \( G'' \subseteq G' \) and a large ultra \( K^6 \)-approximate group \( \hat{A} \subseteq A^4 \) such that the restriction \( \pi : \langle \hat{A} \rangle \to G'' \) is a good model of \( \hat{A} \).
Let $\tilde{U}_0$ be a neighborhood of 1 in $G''$ as required by condition (i) of good models, and again by Gleason-Yamabe let $H \subseteq \tilde{U}_0$ be a compact subgroup of $G''$ such that $L = G''/H$ is a connected Lie group. Let $q : G'' \rightarrow G''/H$ be the quotient map. By Lemma 6.9, $q \circ \pi : \langle \tilde{A} \rangle \rightarrow L$ is a good model of $\tilde{A}$. 

Remark. In the case where $A$ is an ultraproduct of finite $K$-approximate groups and $L$ a connected Lie model of it, [7] shows that $L$ is necessarily nilpotent. From that they derive upper bounds on the dimension of (certain quotients of) $L$ in terms of the covering parameter $K$. Unfortunately we don't have that luxury, because the Lie model of an ultraproduct of open precompact $K$-approximate groups need not be nilpotent; for instance, consider a small ball $A \subseteq \text{SL}(2, \mathbb{R})$, which is a 1000-approximate group by Example 2.3. Its ultrapower $A$ is naturally a subset of $\text{SL}(2, \mathbb{R}^*)$. The standard part map $\text{st} : \mathbb{R}^* \rightarrow \mathbb{R}$, applied entrywise, gives a natural morphism of $\langle A \rangle$ into $\text{SL}(2, \mathbb{R})$, which turns out to be a good model of $A$. But $\text{SL}(2, \mathbb{R})$ is not nilpotent; it is not even solvable.
CHAPTER 7

Strong Approximate Groups and Good Escape Norms

In this chapter we will use good Lie models to extract, from ultra approximate groups, an object which is “closer to linear” in some sense. Thus it is no surprise that Lie algebras play a role. We will also define a norm-like function and show that it has good properties on these linear-like objects.

7.1 Extracting Strong Approximate Groups

**Definition 7.1.** A $K$-approximate group $A$ is said to be a strong $K$-approximate group if there is a symmetric $S \subseteq A$ such that

(i) (Smallness) $(S^{A^3})^{10^4 K^3} \subseteq A$;

(ii) (Macro trapping condition) If $g, g^2, \ldots, g^{1000} \in A^{100}$ then $g \in A$;

(iii) (Micro trapping condition) If $g, g^2, \ldots, g^{(100K)^3} \in A$ then $g \in S$.

We call such an $S$ a core of $A$; more than one core may exist. Note that, if we pass from $S$ to a subset $S' \subseteq S$, smallness and macro trapping are unaffected. Thus, if $S' = \{g \in S \mid g, g^2, \ldots, g^{(100K)^3} \in A\}$, then $S'$ is also a core of $A$, and moreover it is open, since $A$ is open. In other words, we can always assume the core is open if necessary.

As usual, an ultra strong approximate group is an ultraproduct $A = \prod_{n \to \alpha} A_n$ where each $A_n$ is a strong $K$-approximate group for some $K$ independent of $n$. In the sequel we will often abbreviate “strong ($K$-)approximate group” to “($K$)-s.a.g.”.
Remark. Given an ultra s.a.g. \( A = \prod_{n \to \alpha} A_n \), if one chooses a core \( S_n \) for each \( A_n \), then their ultraproduct \( S = \prod_{n \to \alpha} S_n \) is a core for \( A \). Conversely, an ultra approximate group \( A \) which is a (not necessarily ultra) s.a.g. is in fact an ultra s.a.g. if it admits a core which is an ultraproduct.

An obvious example of a strong \( 5^d \)-approximate group is an open convex body \( B \) in \( \mathbb{R}^d \) that is symmetric about the origin; its core may be taken to be \( \frac{1}{10^5 5^3} B \). The next lemma turns this example into more general ones.

It is worth noting that, while \( K \)-approximate groups are also \( K' \)-approximate groups for all \( K' \geq K \), the same is not true for strong \( K \)-approximate groups. Specifically, the “smallness” property of the core may not hold for \( K' > K \). Moreover, there does not seem to be a simple fix for this (e.g. passing to a smaller core via Sanders-Croot-Sisask theory), because shrinking the core may break the micro trapping condition.

Even if \( A \) is a \( K \)-s.a.g., \( A^2 \) and homomorphic images of \( A \) may fail to be \( K^{O(1)} \)-s.a.g.; here the macro trapping condition also fails.

All in all, the notion of strong approximate group is quite subtle, which is why the arguments in this section are technical and delicate. In particular, to go from the ultraproduct setting to the standard setting we use a compactness argument, but for it to go through we need a fairly intricate setup; it is not enough that there are strong approximate groups at the ultraproduct level.

Lemma 7.2 (Small balls are strong approximate groups). Let \( G \) be any group, \( L \) a Lie group, \( \phi : G \to L \) a morphism and \( B \) an bounded convex open neighborhood of 0 in the Lie algebra \( l \) of \( L \). Then there is \( r_0 > 0 \) such that, for all \( 0 < r < r_0 \) and \( K \geq 10^{\text{dim}(L)} \), any set \( A \subseteq G \) satisfying

\[
\phi^{-1}(\exp(rB)) \subseteq A \subseteq \phi^{-1}(\exp(2rB))
\]

is a (possibly not open or precompact) \( K \)-s.a.g.; and for the core we may take any set \( S \) with

\[
\phi^{-1}(\exp(\frac{r}{10^5 K^3} B)) \subseteq S \subseteq \phi^{-1}(\exp(\frac{r}{2 \cdot 10^5 K^3} B))
\]
Proof. Let $K \geq 10^{\dim(L)}$ be arbitrary and let $C_\lambda$ denote the set $\phi^{-1}(\exp(\lambda B))$. By Example 2.3 and Lemma 4.5 the latter is a (possibly not open or precompact) $K$-approximate group for sufficiently small $\lambda$. Now for given $r > 0$ let $C_r \subseteq A \subseteq C_{2r}$, $C_r/10^5K^3 \subseteq S \subseteq C_{r/2.10^4K^3}$ be arbitrary. To show that $A$ is a s.a.g. with core $S$ for small enough $r$, it suffices to establish the following “overestimates” for small $r$:

(i) $((C_r/2.10^4K^3)^{C_{2r}})^{10^4K^3} \subseteq C_r$;
(ii) if $g, \ldots, g^{1000} \in (C_{2r})^{100}$ then $g \in C_r$; and
(iii) if $g, \ldots, g^{(100K)^3} \in C_{2r}$ then $g \in C_{r/10^5K^3}$.

These in turn follow, by application of $\phi^{-1}$, from the Lie-side overestimates

(i) $(\exp(\frac{r}{2.10^4K^3}B)^{10^4K^3})^{\exp(2rB)^4} \subseteq \exp(rB)$;
(ii) if $g, \ldots, g^{1000} \in \exp(2rB)^{100}$ then $g \in \exp(rB)$;
(iii) if $g, \ldots, g^{(100K)^3} \in \exp(2rB)$ then $g \in \exp(\frac{r}{10^5K^3}B)$.

But the latter are simple consequences of the Baker-Campbell-Hausdorff formula for small enough $r_0$. (Note in particular that one can choose $r_0$ independent of $K$.)

\[ \square \]

**Theorem 7.3** (Extracting ultra strong approximate groups). Let $A$ be an ultra approximate group, $L$ a Lie group and $\pi : \langle A \rangle \to L$ a good model of $A$. Let $B$ be an open bounded convex subset of the Lie algebra $L$ of $L$. Then there is $r_0 > 0$ such that, for all $0 < r < r_0$ and $K \geq 10^{\dim(L)}$, any ultraproduct of open sets $\tilde{A}$ satisfying $\pi^{-1}(\exp(rB)) \subseteq \tilde{A} \subseteq \pi^{-1}(\exp(2rB))$ is a large ultra $K$-s.a.g. inside $A$.

In particular, if $A$ is an ultra approximate group, then there exists $M \in \mathbb{N}$ and an $M$-s.a.g. $\tilde{A} \subseteq A$ such that $A$ can be covered by $M$ left-translates of $\tilde{A}$.

**Proof.** Let $K \geq 10^{\dim(L)}$ be arbitrary and choose $r_0$ as in Lemma 7.2; if necessary, shrink it until $\exp(2r_0B)$ is contained in the neighborhood $U_0$ required by condition (i) in the definition of good models, and also $\exp$ is a diffeomorphism from $2r_0B$ to $\exp(2r_0B)$. Then
for $0 < r < r_0$, a set $\tilde{A}$ as in the statement is a $K$-s.a.g. by that same lemma. We need to show that it is an ultra s.a.g., and that it is large.

Largeness is simple: $\pi(A)$ is precompact, so may be covered by finitely many translates of $\exp(rB)$; pulling back along $\pi$, we cover $A$ by finitely many translates of $\pi^{-1}(\exp(rB))$, hence of $\tilde{A}$.

Next, the “ultra” property. Lemma 7.2 says that any set $S$ with $\pi^{-1}(\exp(\frac{r}{10^2K^2}B)) \subseteq S \subseteq \pi^{-1}(\exp(\frac{r}{10^2K^2}B))$ works as a core for $\tilde{A}$. If we can find such an $S$ which is an ultraproduct, we will have shown that $\tilde{A}$ is an ultra s.a.g., as per the remark after definition 7.1. But this is a straightforward consequence of condition (iii) in the definition of good models: let $F$ be the closure of $\exp(\frac{r}{10^2K^2}B)$, which is compact; let $U$ be $\exp(\frac{r}{10^2K^2}B)$, which is open; and note that by choice of $r_0$ we have $F \subseteq U \subseteq U_0$. Thus there is an ultraproduct $S$ that satisfies $F \subseteq S \subseteq U$ and therefore works as a core for $\tilde{A}$.

To establish the last claim, and make this theorem non-vacuous, we must show that there is at least one such ultraproduct $\tilde{A}$. We know that: $\pi$ is a good model; $\exp(2rB)$ is open, since $2rB$ is open; and the closure of $\exp(rB)$, being a closed subset of the precompact set $U_0$, is a compact set contained in the image of $A$. Thus we may apply condition (iii) in the definition of good models to find the required $\tilde{A}$. Let $N$ be the number of translates of $\tilde{A}$ needed to cover $A$; then the last claim in the theorem holds with $M = \max\{10^{\dim(L)}, N\}$. \lozenge

Let us record here a consequence at the level of plain approximate groups, rather than ultra approximate groups.

**Corollary 7.4** (Extracting strong approximate groups). Suppose $A$ is a $K$-approximate group. Then there is a strong $O_K(1)$-approximate group $\tilde{A} \subseteq A$ such that $A$ can be covered by $O_K(1)$ left-translates of $\tilde{A}$.

*Proof.* Suppose the statement is false. Then, carefully negating all the quantifiers, we see that for some $K$ there is a sequence of $K$-approximate groups $(A_n)$ such that, for any $n$-s.a.g. $\tilde{A}_n \subseteq A_n$, more than $n$ translates of $\tilde{A}_n$ are needed to cover $A_n$. 

56
Consider the ultraproduct $A = \prod_{n \to \alpha} A_n$. By the previous theorem there is $N_0 \in \mathbb{N}$ and an ultra $N_0$-s.a.g. $\tilde{A} = \prod_{n \to \alpha} \tilde{A}_n$, contained in $A^4$, such that $A$ can be covered by $N_0$ translates of $\tilde{A}$.

Then, by Łoś’s theorem, there are arbitrarily large $n$ such that $N_0$ translates of the $N_0$-s.a.g. $\tilde{A}_n$ suffice to cover $A_n$. But this is impossible for $n > N_0$. This contradiction establishes the corollary.

7.2 Strong Approximate Groups Have Good Escape Norm

Now that we have extracted strong approximate groups from plain approximate groups, we put the former to the main use they find in this paper.

**Definition 7.5** (Escape norm). Let $G$ be a group and $X$ any subset containing 1. The escape norm relative to $X$ is:

$$\|g\|_{e,X} = \inf \left\{ \frac{1}{n+1} \mid n \in \mathbb{N}; \ g^i \in X \text{ for all } 0 \leq i \leq n \right\}$$

**Remark.** Since $g^0 = 1 \in X$ for every $g \in G$, this quantity is well-defined everywhere. We have $\|g\|_{e,X} = 1$ precisely when $g \notin X$, and, at the other extreme, $\|g\|_{e,X} = 0$ precisely when $\langle g \rangle \subseteq X$. Note also that, if $X$ is symmetric, then $\|g\|_{e,X} = \|g^{-1}\|_{e,X}$.

In the sequel we will define a number of functions from the escape norm, so we need to establish some degree of regularity for it. Since its codomain is a discrete set, the escape norm cannot be continuous in general, so the next result is the best we can hope for.

**Lemma 7.6.** Let $G$ be a group, $A$ a strong approximate group in $G$, and $\|\cdot\|_{e,A}$ the associated escape norm. Then $\|\cdot\|_{e,A}$ is a Borel measurable function on $G$. More precisely, for every $r \in \mathbb{R}$ the set of $g \in G$ such that $\|g\|_{e,A} < r$ is open.

**Proof.** If $r$ is a real number, let $E_r$ be the set of $g$ such that $\|g\|_{e,A} < r$. We have $E_r = \emptyset$ for $r \leq 0$ and $E_r = G$ for $r > 1$. Suppose $0 < r \leq 1$ remains and let $n$ be the largest integer such that $r \leq \frac{1}{n+1}$; then $E_r = E_{\frac{1}{n+1}}$. We show that the latter is open.

57
For \( m \) a non-negative integer, let \( A_m \) be the set of \( g \) such that \( g^m \in A \). For fixed \( m \) the map \( g \mapsto g^m \) is continuous, so each \( A_m \) is open. Now note that \( E_{\frac{1}{m+1}} \) is none other than \( A_0 \cap \cdots \cap A_{n+1} \), which is open.

The escape norm has especially useful properties when \( X \) is taken to be a strong approximate group, as the next theorem shows.

**Theorem 7.7.** Let \( A \) be a strong \( K \)-approximate group. There are absolute constants \( C_0, C_1 > 0 \) such that the escape norm relative to \( A \) satisfies:

1. If \( g \in \langle A \rangle \) and \( h \in A^{49} \) then \( \| hgh^{-1} \|_{e,A} \leq 1000 \| g \|_{e,A} \);
2. If \( g_1, \ldots, g_n \in \langle A \rangle \) then \( \| g_1 \cdots g_n \|_{e,A} \leq C_0 K^4 (\| g_1 \|_{e,A} + \cdots + \| g_n \|_{e,A}) \);
3. If \( g \in \langle A \rangle \) and \( h \in A^{49} \) then \( \|[g,h]\|_{e,A} \leq C_1 K^{11} \| g \|_{e,A} \| h \|_{e,A} \).

**Remark.** We have not tried to optimize \( C_0 \) or \( C_1 \). The proof below obtains \( C_0 = 2 \cdot 10^{10} \) and \( C_1 = 16 \cdot 10^{26} \).

**Proof.** The first property is simple. If \( \| g \|_{e,A} \geq 1/1000 \) there is nothing to prove, so suppose \( \| g \|_{e,A} < 1/1000 \). Now, if \( g, g^2, \ldots, g^N \in A \) (with \( N \geq 1000 \)) then each of the conjugates \( hgh^{-1}, (hgh^{-1})^2, \ldots, (hgh^{-1})^N \) lie in \( hAh^{-1} \), which is contained in \( A^{99} \). In particular, for any \( m \leq N/1000 \), the geometric progression

\[
(hgh^{-1})^m, ((hgh^{-1})^m)^2, \ldots, ((hgh^{-1})^m)^{1000}
\]

lies in \( A^{99} \). Thus, by the macro trapping condition, \( (hgh^{-1})^m \in A \) for every \( 0 \leq m \leq N/1000 \).

In other words, we have proved that: if \( \| g \|_{e,A} \leq \frac{1}{N+1} \), then \( \| hgh^{-1} \|_{e,A} \leq \frac{1000}{N+1000} \). Claim (i) follows.

To establish the other two claims, the strategy is to relate the escape norm to another norm, for which the claims are easy. The latter norm is of the kind introduced in the proof of Proposition 6.3, where one has a group \( G \) act by translation on its space of bounded
functions, and pulls back the $\ell^\infty$ norm. More precisely, if we fix a bounded function $\Psi$ on $G$ and define its “derivative” $\partial_g \Psi$ by $\partial_g \Psi(x) = \Psi(gx) - \Psi(x)$, we have

$$\| \partial_{gh} \Psi \|_\infty \leq \| \partial_{gh} \Psi - \partial_h \Psi \|_\infty + \| \partial_h \Psi \|_\infty = \| \partial_g \Psi \|_\infty + \| \partial_h \Psi \|_\infty. \quad (7.1)$$

So, defining $\| \cdot \|_\Psi = \| \partial_g \Psi \|_\infty$, we get $\| gh \|_\Psi \leq \| g \|_\Psi + \| h \|_\Psi$. The main technical challenge in this proof is to find a function $\Psi$ for which we can establish the inequalities necessary to transfer this property of $\| \cdot \|_\Psi$ to $\| \cdot \|_{e,A}$, namely

$$C^{-1} K^{-4} \| g \|_\Psi \leq \| g \|_{e,A} \leq C K^4 \| g \|_\Psi. \quad (7.2)$$

for some absolute constant $C > 0$. In the end we will not prove these directly, but will use $\| \cdot \|_\Psi$ heavily in the proof of (ii).

We start with the latter inequality, which is simpler, and for which we only require that $\Psi$ be supported on $A^{100}$ and $\Psi(1) \geq 1/2$. As it turns out, we can prove a stronger version of the inequality, without the $K^4$ factor on the right-hand side, and we can take $C = 2000$.

The reasoning is similar to that in Theorem 5.3. By definition of $\| \cdot \|_\Psi$ and the triangle inequality (applied to the appropriate telescoping sum) we have $\| \partial_{g^i} \Psi \|_\infty \leq i \| g \|_\Psi$. If $0 \leq i < 1/2 \| g \|_\Psi$, we have in particular $|\Psi(g^i) - \Psi(1)| < 1/2$, whence $\Psi(g^i) > 0$ and $g^i \in A^{100}$. The macro trapping condition for $A$ then implies that $g^i \in A$ for $0 \leq i < 1/2000 \| g \|_\Psi$, which is to say,

$$\| g \|_{e,A} \leq 2000 \| g \|_\Psi \quad (7.3)$$

It remains to prove the left-hand inequality, i.e. that $\Psi$ can be found which is supported on $A^{100}$, and such that $\Psi(1) \geq 1/2$ and $\| g \|_\Psi \leq C K^4 \| g \|_{e,A}$.

One way to view the construction is as follows. We are trying to make $\Psi$ “Lipschitz” with respect to the escape norm, and moreover have the Lipschitz constant depend only on $K$. We do this by defining and convolving two functions: $\phi$, which is Lipschitz but whose Lipschitz constant we can’t control in terms of $K$; and $\psi$, which isn’t Lipschitz but whose fine-scale
variation we can bound in terms of $K$ only. It is a familiar idea that convolution products inherit the regularity properties of each factor; perhaps that can serve as motivation for what follows, although this specific instance is not standard. Neither $\phi$ nor $\psi$ is hard to construct; the difficulty will be showing that their convolution inherits the required properties.

We proceed to the details. Let $\varepsilon \geq 0$ be a parameter. (Note that for now we allow for $\varepsilon$ to be zero.) Put $\|g\|^{(e)} = \|g\|_{e,A} + \varepsilon$ and define

$$d^{(e)}(g) = \inf \left\{ \sum_{i=1}^{n} \|g_i\|^{(e)}_{e,A} \mid n \geq 1, \ g_1 \cdots g_n = g \right\},$$

a sort of ($\varepsilon$-fuzzed) shortest-path distance to the identity. We will also use the following variants: $d^{(e)}(g, h)$, understood to mean $d^{(e)}(gh^{-1})$, or equivalently $d^{(e)}(hg^{-1})$; and $d^{(e)}(g, B)$, where $B$ is a set, understood to mean $\inf_{b \in B} d^{(e)}(g, b)$. It is clear that $d^{(e)}(g) \leq \|g\|^{(e)}_{e,A}$, that $d^{(e)}(g_1 \cdots g_n) \leq d^{(e)}(g_1) + \cdots + d^{(e)}(g_n)$, and that therefore $d^{(e)}(g_1, g_3) \leq d^{(e)}(g_1, g_2) + d^{(e)}(g_2, g_3)$.

In the sequel we will integrate several functions defined from $d^{(e)}$, so we need to establish some regularity for it.

**Lemma 7.8.** For any $\varepsilon \geq 0$ the function $d^{(e)}$ is Borel measurable. More precisely, for any $r \in \mathbb{R}$ the set of $g$ such that $d^{(e)}(g) < r$ is open.

**Proof.** For a real number $r$, let $D_r$ be the set of $g$ with $d^{(e)}(g) < r$, and $E_r$ the set of $g$ such that $\|g\|_{e,A} < r$. By Lemma 7.6 the $E_r$ are open.

If $r \leq \varepsilon$ or $r > 1 + \varepsilon$ then $D_r$ is empty or all of $G$ respectively, so assume $\varepsilon < r \leq 1 + \varepsilon$. We will show that for every $g \in D_r$ there is a neighborhood of $g$ contained in $D_r$.

First, note that $g \in D_r$ if and only if it is possible to write $g$ as a product $g_1 \cdots g_n$ with $\sum_{i=1}^{n} \|g_i\|^{(e)}_{e,A} < r$. Now, for each $i$, define $r_i$ as follows. If $\|g_i\|_{e,A} = \frac{1}{m_i+1}$ put $r_i = \frac{1}{m_i}$; we know $n_i > 0$ because $\|g_i\|^{(e)}_{e,A} < 1 + \varepsilon$. On the other hand, if $\|g_i\|_{e,A} = 0$, put $r_i = \frac{1}{m_i}(r - \sum_{i=1}^{n} \|g_i\|^{(e)}_{e,A})$.

Finally, we claim that the open set $E_{r_1} \cdots E_{r_n}$, which includes $g$, is contained in $D_r$. Indeed, its elements are of the form $g'_1 \cdots g'_n$; and we have ensured that, where $\|g_i\|_{e,A} > 0$,
also $\|g'_i\|_{e,A} \leq \|g_i\|_{e,A}$; and where $\|g_i\|_{e,A} = 0$, the corresponding $\|g'_i\|_{e,A}$ are not large enough to make $\sum_{i=1}^{n} \|g'_i\|_{e,A}^{(e)}$ go over $r$.

\[\square\]

**Corollary 7.9.** For any set $B$ and $\varepsilon \geq 0$, the function $d^{(e)}(\cdot, B)$ is Borel measurable. More precisely, for any $r \in \mathbb{R}$ the set of $g$ such that $d^{(e)}(g, B) < r$ is open.

**Proof.** For any set $B$ and $r \in \mathbb{R}$, we have $d^{(e)}(g, B) < r$ if and only if there is $b \in B$ such that $d^{(e)}(gb^{-1}) < r$. In other words,

$$\{ g \in \langle A \rangle \mid d^{(e)}(g, B) < r \} = \bigcup_{b \in B} \{ g \in \langle A \rangle \mid d^{(e)}(gb^{-1}) < r \} = \bigcup_{b \in B} b \cdot \{ g \in \langle A \rangle \mid d^{(e)}(g) < r \}$$

By the preceding lemma each set in that union is open, which concludes the proof. $\square$

The plan is to show that

$$\|g\|^{(e)}_{e,A} \leq CK^4 d^{(e)}(g)$$

(7.4)

for an absolute constant $C > 0$; it will then follow that

$$\|g_1 \cdots g_n\|^{(e)}_{e,A} \leq CK^4 d^{(e)}(g_1 \cdots g_n) \leq CK^4(d^{(e)}(g_1) + \cdots + d^{(e)}(g_n)) \leq CK^4(\|g_1\|^{(e)}_{e,A} + \cdots + \|g_n\|^{(e)}_{e,A}).$$

The $\|\cdot\|_{\psi}$ norm mentioned earlier will serve as a stepping stone to inequality (7.4). Letting $\varepsilon \to 0$ then yields property (ii) in the theorem statement.

Let us now define the two functions $\phi$ and $\psi$, and note a few of their properties:

**Lemma 7.10 (Auxiliary functions).** For any $\varepsilon > 0$ there are functions $\phi^{(e)}, \psi : \langle A \rangle \to [0, 1]$ such that:

(i) $\phi^{(e)}$ and $\psi$ are both identically 1 on $A$;

(ii) $\phi^{(e)}$ and $\psi$ are supported on $A^2$;

(iii) For all $g \in \langle A \rangle$ we have

$$\|\partial_g \phi^{(e)}\|_{\infty} \leq \frac{d^{(e)}(g)}{d^{(e)}(1, \langle A \rangle \setminus A)}.$$
\( (\text{i}) \) For all \( h \in S \) and \( y \in A^4 \) we have
\[
\| \partial_{ghy^{-1}} \psi \|_\infty \leq \frac{1}{10^4 K^3}.
\]

Proof. We start with \( \phi^{(e)} \); define
\[
\phi^{(e)}(x) = \max \left\{ 0 , 1 - \frac{d^{(e)}(x, A)}{d^{(e)}(1, \langle A \rangle \setminus A)} \right\}
\]
It’s clear that \( 0 \leq \phi^{(e)}(x) \leq 1 \) for all \( x \). For any \( y \notin A \), we have either \( d^{(e)}(y) = 1 + \varepsilon \), in case the infimum in the definition of \( d^{(e)} \) is achieved with a one-term product; or \( d^{(e)}(y) \geq 2\varepsilon \), in case two or more terms yield a value closer to the infimum. Since \( \varepsilon \) is assumed small, we have \( d^{(e)}(y) \geq 2\varepsilon \). Now, if \( x \in A \) then \( d^{(e)}(x, A) = \varepsilon \), and in that case it is plain that \( \phi^{(e)}(x) \geq 1/2 \).

On the other hand, if \( \phi^{(e)}(x) > 0 \) then it must be the case that \( d^{(e)}(x, A) < d^{(e)}(1, \langle A \rangle \setminus A) \); in other words, \( d^{(e)}(1, Ax^{-1}) < d^{(e)}(1, \langle A \rangle \setminus A) \). Hence \( Ax^{-1} \) contains an element, say \( ax^{-1} \), which is not in \( \langle A \rangle \setminus A \), i.e. lies in \( A \). Therefore \( x \in A^2 \).

Lastly, for any \( x, g \in \langle A \rangle \) and \( a \in A \) we have \( d^{(e)}(gx, a) \leq d^{(e)}(gx, x) + d^{(e)}(x, a) \). Hence \( d^{(e)}(gx, A) \leq d^{(e)}(g) + d^{(e)}(x, A) \) and \( d^{(e)}(gx, A) - d^{(e)}(x, A) \leq d^{(e)}(g) \). Replacing \( x \) with \( gx \) and \( g \) with \( g^{-1} \) in the latter inequality, and using the symmetry of \( d^{(e)} \), we obtain \( d^{(e)}(x, A) - d^{(e)}(gx, A) \leq d^{(e)}(g) \), and thus \( |d^{(e)}(gx, A) - d^{(e)}(x, A)| \leq d^{(e)}(g) \). From this property (iii) follows easily.

We now define and analyze \( \psi \). Recall that \( S \) is the core of the s.a.g. and put \( Q = S^{A^4} \). We will make \( \psi \) a sort of step function, with level sets \( A, QA \setminus A, Q^2A \setminus QA \). Formally, we define \( \psi(g) = 1 \) if \( g \in A \); \( \psi(g) = 1 - i/10^4 K^3 \) if \( g \in Q^{i+1}A \setminus Q^iA \) for \( 0 \leq i < 10^4 K^3 \); and \( \psi(g) = 0 \) otherwise. Property (i) is clear; property (ii) follows from the fact that \( Q^{10^4 K^3} \subseteq A \); and property (iv) is true by construction.

We are now ready to define the previously discussed function \( \Psi \in \ell^\infty(G) \), and to establish some preliminary properties.

**Lemma 7.11.** For any \( \varepsilon > 0 \) the function \( \Psi : \langle A \rangle \to [0, \infty) \) defined by
\[
\Psi(x) = \frac{1}{\mu(A)} \int_{A^2} \phi^{(e)}(y)\psi(y^{-1}x) \, dy
\]
is bounded. Moreover,

(i) $\Psi$ is supported on $A^4$ and $\Psi(1) \geq 1$;

(ii) $\|\Psi\|_{\infty} \leq K$; and

(iii) For any $g \in \langle A \rangle$ we have $\|\partial_g \Psi\|_{\infty} \leq \frac{K}{2\delta} d^{(e)}(g)$.

Proof. We start with (i). That $\Psi$ is zero outside $A^4$ follows directly from $\phi^{(e)}$ and $\psi$ being supported on $A^2$; and

$$
\Psi(1) = \frac{1}{\mu(A)} \int_{A^2} \phi^{(e)}(y) \psi(y^{-1}) \, dy
\geq \frac{1}{\mu(A)} \int_{A} \phi^{(e)}(y) \psi(y^{-1}) \, dy
\geq \frac{1}{\mu(A)} \int_{A} 1 \, dy
= 1.
$$

Now (ii). Since $\phi^{(e)}$ and $\psi$ are non-negative and both bounded by 1, for every $x$ we have

$$
|\Psi(x)| = \frac{1}{\mu(A)} \int_{A^2} \phi^{(e)}(y) \psi(y^{-1}x) \, dy
\leq \frac{1}{\mu(A)} \int_{A^2} 1 \, dy
\leq K,
$$

We turn to (iii). Although $\Psi$ is defined by an integral over $A^2$, it makes no difference if that integral is over all of $\langle A \rangle$, since $\phi^{(e)}$ and $\psi$ are supported on $A^2$. This observation makes the following calculation go a bit more smoothly. For $g \in \langle A \rangle$ we have

$$
\partial_g \Psi(x) = \frac{1}{\mu(A)} \int_{\langle A \rangle} \phi^{(e)}(gy) \psi(y^{-1}x) \, dy - \frac{1}{\mu(A)} \int_{\langle A \rangle} \phi^{(e)}(y) \psi(y^{-1}x) \, dy
= \frac{1}{\mu(A)} \left( \int_{\langle A \rangle} \phi^{(e)}(gy) \psi(y^{-1}x) \, dy - \int_{\langle A \rangle} \phi^{(e)}(y) \psi(y^{-1}x) \, dy \right)
= \frac{1}{\mu(A)} \int_{\langle A \rangle} \partial_g \phi^{(e)}(y) \psi(y^{-1}x) \, dy.
$$
Still bearing in mind that $\psi$ is supported on $A^2$ and $|\psi| \leq 1$, this gives

\[
|\partial g \Psi(x)| \leq \frac{1}{\mu(A)} \int_{\langle A \rangle} |\partial_y \phi^{(e)}(y)| \cdot |\psi(y^{-1}x)| \, dy.
\]

\[
\leq \|\partial_y \phi^{(e)}\|_\infty \frac{1}{\mu(A)} \int_{\langle A \rangle} |\psi(y^{-1}x)| \, dy
\]

\[
\leq \|\partial_y \phi^{(e)}\|_\infty \frac{1}{\mu(A)} \int_{\langle A \rangle} 1_{x A^2}(y) \, dy
\]

\[
= \|\partial_y \phi^{(e)}\|_\infty \frac{1}{\mu(A)} \mu(A^2)
\]

\[
\leq K \|\partial_y \phi^{(e)}\|_\infty.
\]

Now, since $1 \in A$, we have $d^{(e)}(1,\langle A \rangle \backslash A) \geq 2\varepsilon$, so by item (iii) in Lemma 7.10 we have $\|\partial_y \phi^{(e)}\|_\infty \leq \frac{1}{2\varepsilon} d^{(e)}(g)$, and we are done.

\[\square\]

Remark. Property (i) is (stronger than) what we needed in order to prove the second of the inequalities in (7.2).

For $X \geq 0$, let $P(X)$ stand for the inequality $\|\partial_y \Psi\|_\infty \leq X d^{(e)}(g)$. We have just shown $P(K/2\varepsilon)$, and will now upgrade that to $P(\text{O}(K^4))$.

First, we note that a Lipschitz bound on $\Psi$ implies one for $\phi^{(e)}$. For any $g \notin A$ we have $\|g\|_{e,A} = 1$, which, by equation (7.3), implies $\|\partial_y \Psi\|_\infty \geq 1/2000$; so, if $P(X)$ holds, we get $d^{(e)}(g) \geq 1/2000X$. Since $g \notin A$ was arbitrary, we get $d^{(e)}(1,\langle A \rangle \backslash A) \geq 1/2000X$, and item (iii) in Lemma 7.10 now gives

\[
\|\partial_y \phi^{(e)}\|_\infty \leq 2000X \cdot d^{(e)}(g)
\]

(7.6)

Next, notice that

\[
\partial g^n \Psi = n \partial_g \Psi + \sum_{i=0}^{n-1} \partial_g^i \partial_g \Psi
\]

or, rearranging,

\[
\partial_g \Psi = \frac{1}{n} \partial_g^n \Psi - \frac{1}{n} \sum_{i=0}^{n-1} \partial_g^i \partial_g \Psi.
\]

64
Using the triangle inequality and the bound $\|\Psi\|_{\infty} \leq K$, this yields

$$\|\partial_g \Psi\|_{\infty} \leq \frac{1}{n} \|\partial_g^n \Psi\|_{\infty} + \frac{1}{n} \sum_{i=0}^{n-1} \|\partial_g^i \partial_g \Psi\|_{\infty}$$

$$\leq \frac{2K}{n} + \frac{1}{n} \sum_{i=0}^{n-1} \|\partial_g^i \partial_g \Psi\|_{\infty}$$

$$\leq \frac{2K}{n} + \frac{1}{n} \sum_{i=0}^{n-1} \|\partial_g^i \partial_g \Psi\|_{\infty}.$$ 

To bound the second term, we use the following identity, valid for $g \in A$ and $h \in S$:

$$\partial_h \partial_g \Psi(x) = \frac{1}{\mu(A)} \int_{A^4} (\partial_g \phi^{(e)})(y) \cdot \partial_y^{-1} h_y \psi(y^{-1} x) \, dy$$

together with item (iv) in Lemma 7.10 and the bound (7.6) to obtain

$$\|\partial_h \partial_g \Psi\|_{\infty} \leq \frac{1}{\mu(A)} \int_{A^4} \|\partial_g \phi^{(e)}(y)\|_{\infty} \|\partial_y^{-1} h_y + \psi(y^{-1} x)\|_{\infty} \, dy$$

$$\leq \frac{1}{\mu(A)} \int_{A^4} 2000X \cdot d^{(e)}(g) \cdot \frac{1}{10^4 K^3} \, dy$$

$$= \frac{1}{\mu(A)} 2000X \cdot d^{(e)}(g) \frac{1}{10^4 K^3} \mu(A^4)$$

$$\leq \frac{X}{5} d^{(e)}(g).$$

for $h \in S$ and $g \in \langle A \rangle$. Plugging this back into the estimate for $\|\partial_g \Psi\|_{\infty}$ we have, for $g, \ldots, g^{n-1} \in S$:

$$\|\partial_g \Psi\|_{\infty} \leq \frac{2K}{n} \frac{X}{5} d^{(e)}(g).$$

By definition of escape norm, we may take $n = 1/(100K)^3 \|g\|_{e,A}$; this gives

$$\|\partial_g \Psi\|_{\infty} \leq 2K(100K)^3 \|g\|_{e,A} + \frac{X}{5} d^{(e)}(g) \leq X' \|g\|^{(e)}_{e,A}$$

where $X' = 2 \cdot 10^6 K^4 + \frac{1}{5} X$ and $g \in S$. If $g \notin S$, the inequality holds for a different reason: $\|g\|^{(e)}_{e,A} > 1/(100K)^3$ in that case, and

$$\|\partial_g \Psi\|_{\infty} = \sup_{x \in \langle A \rangle} |\Psi(gx) - \Psi(x)| \leq \sup_{x \in \langle A \rangle} |\Psi(gx)| + \sup_{x \in \langle A \rangle} |\Psi(x)| \leq 2K$$

65
by item (ii) in Lemma 7.11.

To upgrade \(|\partial_g \Psi|_\infty \leq X'g|_{e,A}^{(e)}|\) to our goal, which is \(|\partial_g \Psi|_\infty \leq X'd(g)|\), we use the "triangle inequality" \(|\partial_{gh} \Psi|_\infty \leq |\partial_g \Psi|_\infty + |\partial_h \Psi|_\infty|\) noted in (7.1). For any \(g_1, \ldots, g_n\) such that \(g_1 \cdots g_n = g\), it implies that

\[
|\partial_g \Psi|_\infty \leq |\partial_{g_1} \Psi|_\infty + \cdots + |\partial_{g_n} \Psi|_\infty
\]

Taking the infimum over all such \(g_1, \ldots, g_n\) yields \(|\partial_g \Psi|_\infty \leq X'd(g)|\), or \(P(X')\). Thus we have shown that \(P(X)\) implies \(P(2 \cdot 10^6 K^4 + \frac{K}{2\varepsilon} X)\). Starting with \(P(K/2\varepsilon)\) (item (iii) in Lemma 7.11) and running this improvement \(\lceil \log_{5/2} \varepsilon \rceil\) times we obtain (something stronger than) \(P(10^7 K^4)\).

Finally, we can prove (ii). For each \(\varepsilon > 0\) we have

\[
|g_1 \cdots g_n|_{e,A} \leq 2000|g_1 \cdots g_n|_{\Psi} \quad \text{(by inequality (7.3))}
\]

\[
\leq 2000(|g_1|_{\Psi} + \cdots + |g_n|_{\Psi}) \quad \text{(by inequality (7.1))}
\]

\[
\leq 2 \cdot 10^{10} K^4 (d_{g_1}^{(e)} + \cdots + d_{g_n}^{(e)})
\]

\[
\leq 2 \cdot 10^{10} K^4 (|g_1|_{e,A}^{(e)} + \cdots + |g_n|_{e,A}^{(e)}).
\]

Letting \(\varepsilon \to 0\) gives the claim.

It remains to prove (iii). Most of the argument is similar to that of (ii), but it turns out to be convenient to tackle separately the case of "large" \(h\), i.e. \(h \in A_{49} \setminus A\). Recalling that \(C_0\) is the constant appearing in the product inequality (ii), by (i) we have

\[
|ghg^{-1}h^{-1}|_{e,A} \leq C_0 K^4 (|g|_{e,A} + |hg^{-1}h^{-1}|_{e,A})
\]

\[
\leq C_0 K^4 (|g|_{e,A} + 1000|g^{-1}|_{e,A})
\]

\[
= 1001C_0 K^4 |g|_{e,A}
\]

\[
\leq 1001C_0 K^4 |g|_{e,A} |h|_{e,A}
\]

since \(|h|_{e,A} \leq 1\). Thus in this case (iii) holds with \(C_1 = 3 \cdot 10^{13}\).
Next we turn to “small” $g, h$, i.e. $g, h \in A$, and borrow substantially from the proof of (ii). Our task is made somewhat simpler by having established (ii), which allows us to do without the $\varepsilon$-fuzzed version of quantities of interest. In analogy to the preceding, define
\[
    d(g) = \inf \left\{ \sum_{i=1}^{n} \|g_i\|_{e,A} \mid n \geq 1, \ g_1 \cdots g_n = g \right\}
\]
As before, we define $d(g, h)$ to be $d(gh^{-1})$ and $d(g, A)$ as $\inf_{a \in A} d(ga^{-1})$. But now that the $\varepsilon$ terms are no longer present, it’s clear that for all $g, h$ we have
\[
    d(gh) \leq d(g) + d(h)
\]
and therefore
\[
    d(x, z) \leq d(x, y) + d(y, z).
\]
Moreover $d(g) = d(g^{-1})$, which means $d$ is a pseudometric. Now define
\[
    \phi(g) = \max \left\{ 0, 1 - \frac{d(g, A)}{d(1, \langle A \rangle \setminus A)} \right\}
\]
The function $\phi$ is well-defined because the product inequality (ii) keeps $d(1, \langle A \rangle \setminus A)$ from vanishing: if $x \notin A$ then $\|x\|_{e,A} = 1$, and since $\|x\|_{e,A} \leq C_0 K^4 d(x)$ we get $d(1, x) = d(x) \geq C_0^{-1} K^{-4}$. Thus $d(1, \langle A \rangle \setminus A) \geq C_0^{-1} K^{-4}$. At the other extreme, if $g \in A$ then $d(g, A) = 0$, so $\phi$ is identically 1 on $A$.

This lower bound on $d(1, \langle A \rangle \setminus A)$ implies a Lipschitz bound on $\phi$ as follows. Since $d$ is a pseudometric, for every $g, x$ we have the inequalities
\[
    d(gx, A) \leq d(gx, x) + d(x, A) = d(g) + d(x, A)
\]
\[
    d(x, A) \leq d(x, gx) + d(gx, A) = d(g^{-1}) + d(gx, A)
\]
and as a consequence
\[
    |d(gx, A) - d(x, A)| \leq d(g).
\]
The latter readily implies that
\[
    \|\partial_g \phi\|_{\infty} \leq C_0 K^4 d(g) \leq C_0 K^4 \|g\|_{e,A}. \tag{7.7}
\]
By Lemma 7.8 and its corollary, both $d$ and $\phi$ are Borel measurable. Define $\Phi$ to be the convolution of $\phi$ with itself, i.e.

$$\Phi(x) = \frac{1}{\mu(A)} \int_{A^2} \phi(y)\phi(y^{-1}x) \, dy$$

Note that it vanishes outside of $A^4$, and that $\Phi(1) \geq 1$:

$$\Phi(1) = \frac{1}{\mu(A)} \int_{A^2} \phi(y)\phi(y^{-1}) \, dy \geq \frac{1}{\mu(A)} \int_{A} \phi(y)\phi(y^{-1}) \, dy = \frac{1}{\mu(A)} \int_{A} 1 \, dy = 1.$$  

This allows $\Phi$ to play in the proof of (iii) the same bridge role that $\Psi$ played in the proof of (ii): by the argument used to establish inequality (7.2) we have $\|g\|_{e,A} \leq 2000\|g\|_{\phi}$, so if we can show that

$$\|[g,h]\|_{\phi} = \|\partial_{[g,h]}\Phi\|_{\infty} \leq CK^{11}\|g\|_{e,A}\|h\|_{e,A}$$  \hspace{1cm} (7.8)

then we can quickly prove (iii):

$$\|[g,h]\|_{e,A} \leq 2000\|[g,h]\|_{\phi} \leq 2000CK^{11}\|g\|_{e,A}\|h\|_{e,A}.$$  

So we focus on establishing (7.8).

It is straightforward to check that

$$(\partial_{[g,h]}\Phi)(hxg) = \Phi(hgx) - \Phi(hgx) = \partial_h\partial_g\Phi(x) - \partial_g\partial_h\Phi(x)$$

whence

$$\|\partial_{[g,h]}\Phi(x)\|_{\infty} \leq \|\partial_h\partial_g\Phi(x)\|_{\infty} + \|\partial_g\partial_h\Phi(x)\|_{\infty}.$$  

To bound the terms on the right, we use the identity

$$\partial_h\partial_g\Phi(x) = \frac{1}{\mu(A)} \int_{A^4} \partial_y\phi(y)(\partial_{y^{-1}hy}\phi)(y^{-1}x) \, dy$$

which is established in the same way as equation 7.5. (This is where we need the hypothesis that $g, h \in A$.) From it we get

$$\|\partial_{[g,h]}\Phi\|_{\infty} \leq \frac{2}{\mu(A)} \int_{A^4} \|\partial_y\phi\|_{\infty} \cdot \|\partial_{y^{-1}hy}\phi\|_{\infty} \, dy$$

$$\leq \frac{2}{\mu(A)} \|\partial_y\phi\|_{\infty} \sup_{y \in A^4} \|\partial_{y^{-1}hy}\phi\|_{\infty} \mu(A^4).$$
Now, by (7.6) we have \( \| \partial \phi \|_\infty \leq C_0 K^4 \| g \|_{e,A} \) and

\[
\sup_{y \in A^4} \| \partial_{y^{-1} h y} \phi \|_\infty \leq C_0 K^4 \sup_{y \in A^4} d(y^{-1} h y) \\
\leq C_0 K^4 \sup_{y \in A^4} \| y^{-1} h y \|_{e,A} \\
\leq 1000 C_0 K^4 \| h \|_{e,A}.
\]

Plugging these estimates back into the inequality for \( \| \partial_{[g,h]} \Phi \|_\infty \) we obtain

\[
\| \partial_{[g,h]} \Phi \|_\infty \leq 2000 C_0^2 K^{11} \| g \|_{e,A} \| h \|_{e,A}
\]

which lets us complete the proof:

\[
\| [g, h] \|_{e,A} \leq 2000 \| \partial_{[g,h]} \Phi \|_\infty \leq 4 \cdot 10^6 C_0^2 K^{11} \| g \|_{e,A} \| h \|_{e,A}.
\]
CHAPTER 8

Reduction to Lie Case in Bounded Dimension

8.1 Good Escape Norm Reduces to Lie Case

In this section we use the escape norm to find, for a $K$-strong approximate group $A$, a compact normal subgroup $H \subseteq A$ such that $\langle A \rangle / H$ is a Lie group. Since continuous homomorphic images of $K$-approximate groups are themselves $K$-approximate groups, this allows us to reduce to the case where $A$ is a $K$-approximate Lie group.

**Proposition 8.1.** Let $A$ be a strong $K$-approximate group and $H$ the set of all $x \in \langle A \rangle$ with $\|x\|_{e,A} = 0$. Then $H$ is a compact normal subgroup of $\langle A \rangle$ contained in $A$.

**Proof.** In the remark after Definition 7.1 we noted that $\|x\|_{e,A} = 1$ when $x \notin A$, and that $\|x\|_{e,A} = \|x^{-1}\|_{e,A}$ if $A$ is symmetric. It follows immediately that $H \subseteq A$ and that $H$ is closed under inverses. Closure under products follows from the second Gleason lemma: if $x, y \in H$ then $\|xy\|_{e,A} \leq C_0 K^4 (\|x\|_{e,A} + \|y\|_{e,A}) = 0$, whence $xy \in H$. Normality of $H$ follows from the first Gleason lemma: if $x \in H$ and $g \in \langle A \rangle$, then $\|gxg^{-1}\|_{e,A} \leq 1000 \|x\|_{e,A} = 0$, so $gxg^{-1} \in H$.

Lastly we check that the complement of $H$ is open. Let $x \in \langle A \rangle$ be of nonzero escape norm and $U$ a neighborhood of 1 such that $U^{[C_0 K^4/\|x\|_{e,A}]} \subseteq A$, so that every $u \in U$ has $\|u\|_{e,A} < \|x\|_{e,A}/C_0 K^4$. We claim $xU \subseteq \langle A \rangle \setminus H$.

Indeed, if $u \in U$ then by Theorem 7.7 we have

$$\|x\|_{e,A} = \|xu \cdot u^{-1}\|_{e,A} \leq C_0 K^4 (\|xu\|_{e,A} + \|u^{-1}\|_{e,A})$$
whence
\[ \| xu \|_{e,A} \geq \frac{1}{C_0 K_T} \| x \|_{e,A} - \| u^{-1} \|_{e,A} \]
\[ = \frac{1}{C_0 K_T} \| x \|_{e,A} - \| u \|_{e,A} \]
\[ > 0. \]

Since $H$ is a closed subset of the precompact set $A$, it is itself compact. \hfill \Box

Remark. We will often refer to elements of escape norm zero as “non-escaping elements”.

**Proposition 8.2.** Let $A$ be a strong $K$-approximate group, $H$ the subgroup of non-escaping elements, and $\pi : \langle A \rangle \to \langle A \rangle / H$ the quotient map. Then $\langle A \rangle / H$ has the “no small subgroups” property. More specifically, $\pi(A^{99})$ has non-empty interior and does not contain any non-trivial subgroups.

**Proof.** Since $H \subseteq A$, $A^{99}$ contains the $\pi$-saturated set $A^{98} H$. Thus $\pi(A^{99})$ contains the open set $\pi(A^{98} H)$.

Now suppose $\pi(A^{99})$ contains a subgroup $K$, and let $\pi(x) \in K$ for an arbitrary $x \in A^{99}$. We will show that $x \in H$, hence $\pi(x) = 1$ — thereby establishing that $\pi(A^{99})$ contains only the trivial subgroup.

Indeed, we have $\langle \pi(x) \rangle \subseteq K \subseteq \pi(A^{99})$, whence $\langle x \rangle \subseteq A^{99} H \subseteq A^{100}$. The macro trapping condition implies that $\langle x \rangle \subseteq A$, thus $x \in H$ as claimed. \hfill \Box

**Corollary 8.3.** Let $A$ be a strong $K$-approximate group and $H$ the subgroup of non-escaping elements. Then $\langle A \rangle / H$ is a Lie group.

**Proof.** Since $\langle A \rangle$ is a locally compact group, so is its continuous image $\langle A \rangle / H$. The latter is also a Hausdorff space since $H$ is closed. The claim now follows from Lemma 8.2 and the Gleason-Yamabe theorem. \hfill \Box
8.2 A Continuous Escape Norm

In the last section we proved that, if $A$ is a strong $K$-approximate group, $H$ is the normal subgroup of non-escaping elements, and $\pi: \langle A \rangle \to \langle A \rangle / H$ is the quotient map, then $\pi(A)$ is a $K$-approximate Lie group which does not contain any non-trivial subgroups.

By Corollary 7.4 we can find an $O_K(1)$-strong approximate subgroup $A' \subseteq \pi(A)$ which is $O_K(1)$-equivalent to $A$. For this reason, we will focus on strong approximate subgroups of Lie groups, which do not contain non-trivial subgroups. In the sequel, unless noted otherwise, $A$ denotes a $K$-strong approximate group inside a Lie group $\langle A \rangle$, with Lie algebra $\mathfrak{a}$, and $A$ does not contain non-trivial subgroups.

We now introduce a new norm that measures escape along continuous paths rather than discrete ones. Intuitively, the old escape norm of $g \in \langle A \rangle$ looks at the first time the image of a morphism $\phi: \mathbb{Z} \to \langle A \rangle$ defined by $\phi(1) = g$ lies outside $A$; in order to extend the analysis to Lie groups, it is natural to examine when the image of morphisms $\phi: \mathbb{R} \to \langle A \rangle$ leave $A$. A morphism from $\mathbb{R}$, unlike one from $\mathbb{Z}$, isn’t uniquely defined by the image of $1$; but it is uniquely defined by the image of “infinitesimal elements”, i.e. by its tangent vector at $0$. Thus the natural home for a continuous escape norm is the Lie algebra $\mathfrak{a}$ of $\langle A \rangle$.

It will turn out that this continuous escape norm satisfies analogues of the Gleason lemmas, and allows us to both bound the dimension of a $K$-approximate Lie group, and to give a more precise description of its structure in terms of finite approximate groups and “convex sets”.

Definition 8.4. Let $A$ be a strong $K$-approximate group inside the Lie group $\langle A \rangle$. For any $X \in \mathfrak{a}$ define the path norm (relative to $A$) of $X$ by

$$\|X\|_{\text{path}, A} = 1/\inf\{t > 0 \mid \exp(tX) \notin A\}.$$ 

(We adopt the convention $\inf\emptyset = \infty$.)

We elucidate the difference between the escape norm and the path norm with the following
Example 8.5. In the ambient group $\mathbb{R}$, let $B = (-1, 1)$, $P = \{10i \mid i \in \mathbb{Z}, \, |i| \leq N\}$, and $A = P + B$. Since $A + A = (P + P) + (B + B) = \{-20N, \ldots, -10, 0, 10, \ldots, 20N\} + (-2, 2)$, we see that $A$ is a 6-approximate group.

Intuitively, $A$ is an arithmetic progression of intervals of length 2 centered at multiples of 10, and the elements of small escape norm in $A$ are those situated reasonably close to a reasonably small multiple of 10. More precisely, if $n$ is an integer between $-N$ and $N$, and $-1 < \delta < 1$, then $\|10n + \delta\|_{e,A}$ is the larger among $\|10n\|_{e,P}$ and $\|\delta\|_{e,B}$. Thus, for a given $\varepsilon > 0$, the set of elements of escape norm at most $\varepsilon$ is a shorter arithmetic progression of smaller intervals.

Now we look at the path norm. The Lie algebra of $\mathbb{R}$ is $\mathbb{R}$ itself, and the exponential map is simply the identity. Then the path norm of $10n + \delta$ is $1/\lambda$, where $\lambda$ is the smallest real multiple of $10n + \delta$ that lies outside of $A$. Thus the set of elements of path norm at most $\varepsilon$ is simply the interval $(-\varepsilon, \varepsilon)$.

It is good to keep in mind this image of an approximate group as the set sum of a progression and a convex set, with the convex set picked out as a small ball in the path norm.

Lemma 8.6. The path norm is well-defined and homogeneous.

Proof. Since $A$ is open and $\exp : a \rightarrow \langle A \rangle$ is a continuous map such that $\exp(0) = 1$, the infimum of the set $\{t > 0 \mid \exp(tX) \notin A\}$ cannot be zero.

For homogeneity, note that for $\lambda > 0$ we have

$$\{t > 0 \mid \exp(t \cdot \lambda X) \notin A\} = \frac{1}{\lambda} \{t > 0 \mid \exp(tX) \notin A\}$$

so that

$$\|\lambda X\|_{\text{path}, A} = \frac{1}{\lambda} \inf \frac{1}{\lambda} \{t > 0 \mid \exp(tX) \notin A\} = \frac{\lambda}{\inf \{t > 0 \mid \exp(tX) \notin A\}} = \lambda \|X\|_{\text{path}, A}.$$ 

The case where $\lambda < 0$ now follows from the fact that $A = A^{-1}$.

Lemma 8.7. The path norm is upper semicontinuous.
Proof. Let $X_0 \in \mathfrak{a}$ and $\varepsilon > 0$ be arbitrary. Suppose $\|X_0\|_{\text{path},A} \leq 1/t_0$, i.e. $\exp(tX_0) \in A$ for all $t \in [0, t_0)$. The map $\phi(t, X) = \exp(tX)$ is continuous on $\mathbb{R} \times \mathfrak{a}$, hence the inverse image $\phi^{-1}(A)$ is an open set containing $[0, t_0) \times \{X_0\}$. It follows that, for some open set $U$ containing $X_0$, we have $\phi([0, t_0) \times U) \subseteq A$; in other words, $\exp(tX) \in A$ for every $X \in U$ and $t \in [0, t_0)$. This means $\|X\|_{\text{path},A} \leq 1/t_0$ for every $X \in U$. Since $1/t_0$ was an arbitrary upper bound on the escape norm of $X_0$, the claim follows. \qed

Lemma 8.8. The escape norms relative to the sets $A, A^2, \ldots, A^{100}$ obey the inequalities

$$\|\cdot\|_{e,A} \geq \|\cdot\|_{e,A^2} \geq \cdots \geq \|\cdot\|_{e,A^{100}} \geq \frac{1}{1000} \|\cdot\|_{e,A}$$

and likewise for the path norm.

Proof. Since $A \subseteq A^2 \subseteq \cdots \subseteq A^{100}$, all inequalities are obvious for all norms, except the last. We start with the original escape norm. Let $x \in \langle A \rangle$ and $N$ be the least positive integer such that $x^N \notin A^{100}$, so that $\|x\|_{e,A^{100}} = 1/N$. If $N \leq 1000$ there is nothing to prove. Otherwise $x, \ldots, x^{N-1} \in A^{100}$, and the macro trapping condition implies that $x, \ldots, x^{[(N-1)/1000]} \in A$, so

$$\|x\|_{e,A} \leq 1/((N - 1)/1000) + 1) \leq 1000\|x\|_{e,A^{100}}.$$

We turn to the path norm. Let $X \in \mathfrak{a}$ and let $t_0$ be the infimum in the definition of $\|X\|_{\text{path}, A^{100}}$, so that $\exp(tX) \in A^{100}$ for all $0 \leq t < t_0$. By the macro trapping condition, $\exp(tX) \in A$ for all $0 \leq t < t_0/1000$, and the claim follows. \qed

Lemma 8.9. For any $X \in \mathfrak{a}$ we have $\|\exp(X)\|_{e,A} \leq \|X\|_{\text{path},A}$. If $X \in \mathfrak{a}$ is such that $\|X\|_{\text{path},A} \leq 1$, then also $\|X\|_{\text{path},A} \leq 1000\|\exp(X)\|_{e,A}$.

Proof. For the first inequality, note that $\|\exp(X)\|_{e,A}$ is $1/n$ where $n$ is the least positive integer such that $\exp(X)^n \notin A$. Since $\exp(X)^n = \exp(nX)$, this same $n$ appears in $\{t > 0 \mid \exp(tX) \notin A\}$. If $t_0$ is the infimum of this set, it follows that $t_0 \leq n$, and hence the claim.
For the second inequality, we start from the observation that for any $t > 0$ we have

$$\exp(tX) = \exp([t]X) \cdot \exp((t - [t])X)$$

Since $\|X\|_{\text{path}, A} \leq 1$ and $0 \leq t - [t] < 1$ we have $\exp((t - [t])X) \in A$, and therefore $\exp(tX) \in A^2$ whenever $\exp([t]X) \in A$. Thus $\|X\|_{\text{path}, A^2} \leq \|\exp(X)\|_{e, A}$, and now the result follows by Lemma 8.8.

Having established basic regularity and some bounds for the path norm, we now focus on more “norm-like” properties: non-degeneracy and a kind of triangle inequality. (Recall that we have already proved homogeneity.)

**Lemma 8.10.** Let $A$ be a strong $K$-approximate subgroup of a Lie group $\langle A \rangle$ with Lie algebra $\mathfrak{a}$, and suppose $A$ contains no nontrivial subgroup. If $X \in \mathfrak{a}$ has path norm zero, then $X = 0$.

**Proof.** If $\|X\|_{\text{path}, A} = 0$, then $\exp(tX) \in A$ for all $t \in \mathbb{R}$. Since $\{\exp(tX) \mid t \in \mathbb{R}\}$ is a subgroup, it must be trivial, i.e. $\exp(tX) = 1$ for all $t$. But $\exp$ is a homeomorphism near the origin in $\mathfrak{a}$; hence $X = 0$.

**Proposition 8.11.** There is an absolute constant $C_2$ such that for every $X, Y \in \mathfrak{a}$ we have

$$\|X + Y\|_{\text{path}, A} \leq C_2 K^4 (\|X\|_{\text{path}, A} + \|Y\|_{\text{path}, A})$$

**Proof.** The proof will show that we may in fact take $C_2 = 1000C_0$, with $C_0$ from the Gleason lemmas, Theorem 7.7.

Let $X, Y \in \mathfrak{a}$ and pick $t$ be small enough that Lemma 8.9 holds for $X + Y$:

$$\|t(X + Y)\|_{\text{path}, A} \leq 1000\|\exp(t(X + Y))\|_{e, A}$$

Shrinking $t$ further if necessary, we may assume that the BCH formula holds:

$$\exp(t(X + Y)) = \exp(tX) \exp(tY) \exp(t^2 Z(t))$$
where $Z(t) \in \mathfrak{a}$ is a smooth function of $t$ in a neighborhood of 0. By the Gleason triangle inequality and Lemma 8.9 we have

$$\|\exp(t(X + Y))\|_{e,A} \leq C_0 K^4 \left( \|\exp(tX)\|_{e,A} + \|\exp(tY)\|_{e,A} + \|\exp(t^2 Z(t))\|_{e,A} \right)$$

$$\leq C_0 K^4 \left( \|tX\|_{\text{path},A} + \|tY\|_{\text{path},A} + \|t^2 Z(t)\|_{\text{path},A} \right).$$

Putting these together we obtain

$$\|t(X + Y)\|_{\text{path},A} \leq 1000 C_0 K^4 \left( \|tX\|_{\text{path},A} + \|tY\|_{\text{path},A} + \|t^2 Z(t)\|_{\text{path},A} \right)$$

and by homogeneity of the path norm

$$\|X + Y\|_{\text{path},A} \leq 1000 C_0 K^4 \left( \|X\|_{\text{path},A} + \|Y\|_{\text{path},A} + |t| \|Z(t)\|_{\text{path},A} \right)$$

Since $Z(t)$ converges as $t \to 0$ and the path norm is upper semicontinuous, we may take limits as $t \to 0$ and obtain the claim. \qed

As a side note, we establish an analogue of the Gleason commutator estimate for the path norm.

**Proposition 8.12.** There is an absolute constant $C_3$ such that for every $X, Y \in \mathfrak{a}$ we have

$$\|[X, Y]\|_{\text{path},A} \leq C_3 K^{15} \|X\|_{\text{path},A} \|Y\|_{\text{path},A}$$

**Proof.** Let $X, Y \in \mathfrak{a}$ and pick $t$ small enough that Lemma 8.9 holds for $[X, Y]$:

$$\|t [X, Y]\|_{\text{path},A} \leq 1000 \|\exp(t [X, Y])\|_{e,A}$$

Shrinking $t$ further if necessary, we may assume the BCH formula holds:

$$\exp(t [X, Y]) = [\exp(tX), \exp(tY)] \exp(t^2 Z(t))$$

where $Z(t)$ is a smooth function of $t$ on a neighborhood of 0. By the Gleason inequalities and Lemma 8.9 we have
\[ \| \exp(t [X,Y]) \|_{e,A} \leq C_0 K^4 \left( \| \exp(tX) \exp(tY) \|_{e,A} + \| \exp(t^2 Z(t)) \|_{e,A} \right) \]
\[ \leq C_0 C_1 K^{15} \| tX \|_{\text{path},A} \| tY \|_{\text{path},A} + C_0 K^4 \| t^2 Z(t) \|_{\text{path},A} \]

Putting these together yields
\[ \| t [X,Y] \|_{\text{path},A} \leq C_0 C_1 K^{15} \| tX \|_{\text{path},A} \| tY \|_{\text{path},A} + C_0 K^4 \| t^2 Z(t) \|_{\text{path},A} \]

By homogeneity of path norm,
\[ \| [X,Y] \|_{\text{path},A} \leq C_0 C_1 K^{15} \| X \|_{\text{path},A} \| Y \|_{\text{path},A} + C_0 K^4 \| t \| \| Z(t) \|_{\text{path},A} \]

As before, the convergence of \( Z(t) \) as \( t \to 0 \) and upper semicontinuity of the path norm allow us to take limits as \( t \to 0 \) and complete the proof. \( \square \)

For convenience we gather the preceding facts about the path norm under one

**Proposition 8.13.** Let \( A \) be a strong \( K \)-approximate group inside a Lie group \( \langle A \rangle \), which does not contain non-trivial subgroups. There exist absolute constants \( C_2, C_3 \) such that the path norm on the Lie algebra \( a \) satisfies:

(i) \( \| \exp(X) \|_{e,A} \leq \| X \|_{\text{path},A} \), and if \( \| X \|_{\text{path},A} = 1 \) then \( \| X \|_{\text{path},A} \leq 100 \| \exp(X) \|_{e,A} \);

(ii) \( \| X \|_{\text{path},A} = 0 \) if and only if \( X = 0 \);

(iii) \( \| X + Y \|_{\text{path},A} \leq C_2 K^4 (\| X \|_{\text{path},A} + \| Y \|_{\text{path},A}) \);

(iv) \( \| [X,Y] \|_{\text{path},A} \leq C_3 K^{15} \| X \|_{\text{path},A} \| Y \|_{\text{path},A} \).

### 8.3 A Dimension Bound

The Gleason-type properties of the path norm established in the previous section let us bound the dimension of a strong \( K \)-approximate group \( A \) inside a Lie group, provided \( A \) does not
contain non-trivial subgroups. The overall strategy has three steps: (i) find a large piece \( A' \) of \( A \) on which the map \( x \mapsto x^2 \) is injective; (ii) lower bound the Jacobian determinant of \( x \mapsto x^2 \) on a large piece \( A'' \) of \( A' \), in terms of the dimension of \( A \), giving a lower bound on the measure of the image of \( A'' \) under the square map; and (iii) combine that lower bound with the fact that the image of \( x \mapsto x^2 \) lies in \( A^2 \), which is only \( K \) times larger than \( A \), to obtain an upper bound on \( \dim A \) in terms of \( K \) only.

Throughout this section, \( A \) is a strong \( K \)-approximate group containing no non-trivial subgroups. We start with step (i).

**Lemma 8.14.** For every \( \varepsilon > 0 \) there is a \( 2K^3 \)-approximate group \( T \subseteq A \) of measure at least \( \mu(A)/2K^{O(K/\varepsilon^2)} \) such that every \( t \in T \) has \( \|t\|_{e,A} < \varepsilon \).

**Proof.** Let \( m = \lceil 1000/\varepsilon \rceil \). By Corollary 5.6 we may find a \( 2K^3 \)-approximate group \( T \subseteq A^4 \) of measure \( \mu(A)/2K^{O(K/\varepsilon^2)} \) such that \( T^m \subseteq A^4 \). In particular, for any \( t \in T \) it is the case that \( t, t^2, \ldots, t^m \in A^4 \). By the macro trapping condition we have \( t, t^2, \ldots, t^{m/1000} \in A \), so that \( \|t\|_{e,A} < \varepsilon \). \( \square \)

**Lemma 8.15.** For every \( g \in A \) we have \( \frac{1}{2}\|g^2\|_{e,A} \leq \|g\|_{e,A} \leq 500\|g^2\|_{e,A} \).

**Proof.** If one of \( \|g\|_{e,A} \) or \( \|g^2\|_{e,A} \) is zero then so is the other, by the macro trapping condition, so we are done.

On the other hand, suppose \( \|g^2\|_{e,A} = \frac{1}{n+1} \), i.e. \( g^2, g^4, \ldots, g^{2n} \in A \) but \( g^{2n+2} \notin A \). Then \( \|g\|_{e,A} \geq \frac{1}{2n+2} = \frac{1}{2}\|g^2\|_{e,A} \), which is the first inequality.

Moreover \( g, g^2, \ldots, g^{2n+1} \in A^2 \), so that \( \|g\|_{e,A^2} \leq \frac{1}{2}\|g^2\|_{e,A} \). But by the macro trapping property \( \|g\|_{e,A} \leq 1000\|g\|_{e,A^2} \), which gives the second inequality. \( \square \)

**Lemma 8.16.** Let \( A_\varepsilon \subseteq A \) consist of those \( a \in A \) such that \( \|a\|_{e,A} < \varepsilon \). Then for some \( \varepsilon \gg 1/K^{15} \), the map \( x \mapsto x^2 \) is injective on \( A_\varepsilon \).

**Proof.** Let \( a, b \in A_\varepsilon \) and suppose \( a = gb \), with \( g \neq 1 \). Then \( a^2b^{-2} = g^2[g^{-1}, b] \), and by Theorem 7.7 we have
\|a^2 b^{-2}\|_{e,A} \geq \frac{1}{C_0K^4} \|g\|_{e,A}^2 - \|[g^{-1},b]\|_{e,A} \\
\geq \frac{1}{C_0K^4} \|g\|_{e,A}^2 - C_1 K^{11} \|g\|_{e,A} \|b\|_{e,A} \\
\geq \frac{1}{500C_0K^4} \|g\|_{e,A} - C_1 K^{11} \|g\|_{e,A} \|b\|_{e,A} \\
= \|g\|_{e,A} \left( \frac{1}{500C_0K^4} - C_1 K^{11} \|b\|_{e,A} \right).

The passage from the second to the third line is justified by the preceding lemma. But now it's clear that if \( \varepsilon < 1/500C_0C_1 K^{15} \) the escape norm of \(a^2 b^{-2}\) is positive, and hence \(a^2 \neq b^2\). \(\square\)

We move on to step (ii). To lower bound the Jacobian determinant, we will need to control the eigenvalues of the adjoint action of elements of \(A\) on the Lie algebra. Recall that the adjoint action of a Lie group \(G\) on its Lie algebra \(\mathfrak{g}\) associates to each \(g \in G\) an automorphism \(\text{Ad}_g : \mathfrak{g} \to \mathfrak{g}\) defined by \(g \cdot \exp(tX) \cdot g^{-1} = \exp(t \text{Ad}_g X)\).

The strategy is to show the eigenvalues of \(\text{Ad}_g\) are close to 1 by bounding the operator norm of \(\text{Ad}_g - I\), since the latter upper bounds the spectral radius of \(\text{Ad}_g - I\). One may think of these properties — all eigenvalues of \(\text{Ad}_g\) being close to 1, the operator norm of \(\text{Ad}_g - I\) being small — as continuous/infinitesimal versions of “approximate nilpotence”. This is not surprising, since nilpotent-type behavior is a crucial feature of discrete approximate groups.

**Lemma 8.17.** The operator norm of \(\text{Ad}_g - I\), relative to the path norm, is \(O(K^{15} \|g\|_{e,A})\).

**Proof.** Let \(X \in \mathfrak{a}\) and choose \(t\) small enough for the BCH formula to hold:

\[\exp((\text{Ad}_g - I)tX) = \exp(t \text{Ad}_g X) \exp(-tX) \exp(t^2Z(t))\]
where $Z(t)$ is a smooth function defined on a neighborhood of 0. Then

$$
\|(Ad_g - I)tX\|_{\text{path,}A} \leq \|\exp((Ad_g - I)tX)\|_{e,A}
= \|\exp(tAd_gX)\exp(-tX)\exp(t^2Z(t))\|_{e,A}
= \|g\exp(tX)g^{-1}\exp(-tX)\exp(t^2Z(t))\|_{e,A}
= \|[g,\exp(tX)]\exp(t^2Z(t))\|_{e,A}
\leq C_0K^4 \left(\|[g,\exp(tX)]\|_{e,A} + \|\exp(t^2Z(t))\|_{e,A}\right)
\leq C_0K^4 \left(C_1K^{11}\|g\|_{e,A}\|tX\|_{\text{path,}A} + \|t^2Z(t)\|_{\text{path,}A}\right).
$$

Shrinking $t$ if necessary we may use Lemma 8.9 to write

$$
\|(Ad_g - I)tX\|_{\text{path,}A} \leq 1000C_0K^4 \left(C_1K^{11}\|g\|_{e,A}\|tX\|_{\text{path,}A} + \|t^2Z(t)\|_{\text{path,}A}\right).
$$

Then, by homogeneity of path norm, we obtain

$$
\|(Ad_g - I)X\|_{\text{path,}A} \leq 1000C_0K^4 \left(C_1K^{11}\|g\|_{e,A}\|X\|_{\text{path,}A} + |t|\|Z(t)\|_{\text{path,}A}\right).
$$

Since $Z(t)$ converges as $t \to 0$ and the path norm is upper semicontinuous, we may take limits as $t \to 0$ and obtain

$$
\|(Ad_g - I)X\|_{\text{path,}A} \leq 1000C_0C_1K^{15}\|g\|_{e,A}\|X\|_{\text{path,}A}.
$$

Having bounded the operator norm of $Ad_g - I$ relative to the path norm, one might think to use Gelfand’s formula to bound the spectral radius:

**Theorem 8.18** (Gelfand’s Formula [14]). If $\|\cdot\|$ is any norm on a vector space, and $\|\cdot\|_{\text{op}}$ the operator norm induced from it, then for any operator $T : V \to V$ the spectral radius of $T$ is given by $\lim_{k \to \infty}\|T^k\|^{1/k}_{\text{op}}$.

However, the path norm is not technically a norm, since it only obeys the triangle inequality up to constants. That motivates the introduction of a “geodesic” version of the path norm, which is actually a norm:

80
Definition 8.19. For $X \in \mathfrak{a}$ define

$$\|X\|_{\text{geo},A} = \inf \{ \|X_1\|_{\text{path},A} + \cdots + \|X_n\|_{\text{path},A} \mid X_1 + \cdots + X_n = X \}$$

Lemma 8.20. The geodesic norm is a genuine norm on $\mathfrak{a}$, and satisfies

$$\|X\|_{\text{geo},A} \leq \|X\|_{\text{path},A} \leq C_0 K^4 \|X\|_{\text{geo},A}.$$ 

Proof. From the definition it is clear that $\|X\|_{\text{geo},A} \leq \|X\|_{\text{path},A}$. Moreover, Proposition 8.13 says

$$\|X\|_{\text{path},A} \leq C_0 K^4 (\|X_1\|_{\text{path},A} + \cdots + \|X_n\|_{\text{path},A})$$

for any $X_1, \ldots, X_n$ such that $X_1 + \cdots + X_n = X$. Taking infima we obtain $\|X\|_{\text{path},A} \leq C_0 K^4 \|X\|_{\text{geo},A}$. This establishes the inequalities and non-degeneracy of $\|\cdot\|_{\text{geo},A}$.

Homogeneity of the geodesic norm follows easily from that of the path norm, and the triangle inequality for $\|\cdot\|_{\text{geo},A}$ is a simple consequence of its definition. \qed

Proposition 8.21. For every $g \in A$ the eigenvalues of $\text{Ad}_g$ are within $O(K^{15} \|g\|_{e,A})$ of 1.

Proof. We will show that the spectral radius of $\text{Ad}_g - I$ is at most $1000C_0 C_1 K^{15} \|g\|_{e,A}$, from which the result follows.

By Gelfand’s formula, the spectral radius in question is $\lim_{k \to \infty} \|(\text{Ad}_g - I)^k\|_{\text{op}}^{1/k}$, where the operator norm is relative to the geodesic norm. Using Lemma 8.20 we bound $\|(\text{Ad}_g - I)^k\|_{\text{op}}$ as follows: for any $X \in \mathfrak{a}$ we have

$$\|(\text{Ad}_g - I)^k X\|_{\text{geo},A} \leq \|(\text{Ad}_g - I)^k X\|_{\text{path},A} \leq (1000C_0 C_1 K^{15} \|g\|_{e,A})^k \|X\|_{\text{path},A} \leq C_0 K^4 (1000C_0 C_1 K^{15} \|g\|_{e,A})^k \|X\|_{\text{geo},A}.$$ 

Thus $\|(\text{Ad}_g - I)^k\|_{\text{op}}^{1/k} \leq (C_0 K^4)^{1/k} 1000C_0 C_1 K^{15} \|g\|_{e,A}$, and taking limits as $k \to \infty$ the result follows. \qed

Proposition 8.22. Let $A$ be a strong $K$-approximate Lie group containing no non-trivial subgroups. Then $\langle A \rangle$ has dimension $O(K^{31} \log K)$. 

81
Proof. Define \( f : \langle A \rangle \to \langle A \rangle \) by \( f(x) = x^2 \). By Lemma 8.16 and Proposition 8.21 there is \( \varepsilon \gg 1/K^{15} \) such that \( f \) is injective on \( A_\varepsilon \) and for every \( g \in A_\varepsilon \) every eigenvalue of \( \text{Ad}_g \) is within distance \( \frac{1}{2} \) of 1. Moreover, by Lemma 8.14, with this choice of \( \varepsilon \) we have \( \mu(A_\varepsilon) \geq \mu(A)/2K^{O(K^{31})} \). Since \( Df_g = I + \text{Ad}_g^{-1} \) (and \( A_\varepsilon \) is symmetric), the Jacobian determinant of \( f \) is everywhere at least \((1 + \frac{1}{2})^{\dim L}\).

It follows, by the change of variables formula, that \( \mu(f(A_\varepsilon)) \geq \left(\frac{3}{2}\right)^{\dim(A)} \mu(A_\varepsilon) \). On the other hand, \( f(A_\varepsilon) \subseteq A^{2}_\varepsilon \subseteq A^2 \), so \( \mu(f(A_\varepsilon)) \leq K\mu(A) \leq 2K^{O(K^{31})}\mu(A_\varepsilon) \). Putting these inequalities together yields

\[
\left(\frac{3}{2}\right)^{\dim(A)} \leq 2K^{O(K^{31})}
\]

and taking logarithms yields the result. \( \square \)
CHAPTER 9

A “Finite Plus Convex” Decomposition

Throughout this section, $A$ will be a strong $K$-approximate Lie group which does not contain nontrivial subgroups.

**Lemma 9.1.** Let $G$ be a local group and $H$ a normal sublocal group with reference neighborhood $V$. Let $W^6 \subseteq V$. If $H$ is open then $W/H$ is discrete.

*Proof.* We need only to show that $\{1\}$ is open in $W/H$; by definition of quotient topology, it suffices to check that $\pi^{-1}\{1\}$ is open in $W$. But $\pi^{-1}\{1\}$ in $W$ is just $H \cap W$, which is open.

For all $x, y \in H$ such that $xy$ is defined and lies in $V$, we should have $xy \in H$; for all $g \in V$, $x \in H$ such that $gxg^{-1}$ is defined and lies in $V$, we should have $gxg^{-1} \in H$.

**Proposition 9.2.** Let $A$ be a strong $K$-approximate subgroup of a Lie group $\langle A \rangle$, with Lie algebra $a$. For $X \in a$ let $\text{ad}_X : a \to a$ be the linear operator $Y \mapsto [X,Y]$. Then the operator norm of $\text{ad}_X$ relative to the geodesic norm is $O(K^{19}\|X\|_{\text{path},A})$. Its spectral radius is $O(K^{15}\|X\|_{\text{path},A})$.

*Proof.* The geodesic operator norm bound is a simple consequence of Proposition 8.13 and Lemma 8.20:

$$
\|\text{ad}_X(Y)\|_{\text{geo},A} \leq \|\text{ad}_X(Y)\|_{\text{path},A} \\
\leq C_3 K^{15}\|X\|_{\text{path},A}\|Y\|_{\text{path},A} \\
\leq C_0 C_3 K^{19}\|X\|_{\text{path},A} \cdot \|Y\|_{\text{geo},A}.
$$
To bound the spectral radius we use Proposition 8.13 to bound $\|ad^k_X\|_{op}$, and finish with Gelfand’s formula:

$$\|ad^k_X(Y)\|_{geo,A} \leq \|ad^k_X(Y)\|_{path,A} \leq (C_3K^{15}\|X\|_{path,A})^k\|Y\|_{path,A} \leq C_0K^4(C_3K^{15}\|X\|_{path,A})^k\|Y\|_{geo,A}$$

and therefore

$$\lim_{k \to \infty} \|ad^k_X\|_{op}^{1/k} \leq \lim_{k \to \infty} (C_0K^4)^{1/k}C_3K^{15}\|X\|_{path,A} = C_3K^{15}\|X\|_{path,A}.$$


**Lemma 9.3.** If $V$ is a normed vector space and $T : V \to V$ has operator norm less than 1, then $\|\exp(T) - I\|_{op} < 2$.

**Proof.** We have

$$\|\exp(T) - I\|_{op} = \|T + \frac{1}{2!}T^2 + \frac{1}{3!}T^3 \cdots \|_{op} \leq \|T\|_{op} \left(\|I\|_{op} + \frac{1}{2!}\|T\|_{op} + \frac{1}{3!}\|T\|_{op} + \cdots \right) \leq \|T\|_{op}(e - 1) < 2\|T\|_{op}.$$


**Proposition 9.4.** There is $\delta \gg 1/K^{19}$ such that, for any $X, Y \in \mathfrak{a}$ of path norm less than $\delta$, the product $\exp(X)\exp(Y)$ lies in the image of the exponential map.

**Proof.** For $X, Y \in \mathfrak{a}$, the Baker-Campbell-Hausdorff-Dynkin formula says that

$$\log(\exp(X)\exp(Y)) = X + \int_0^1 F(\exp(ad_X)\exp(ad_Y))Y \, dt,$$
as long as $X,Y$ are sufficiently small, where $F(z) = z \log z/(z-1)$ and the exponentials on the right-hand side are matrix exponentials. However, it is known that every Lie group supports an analytic structure, and since both sides of the equation above are analytic functions, the formula holds on any connected subset of $\mathfrak{a} \times \mathfrak{a}$ containing $(0,0)$ on which the integral is well-defined. Therefore it suffices to show that the integral converges for $X,Y$ of path norm up to some $\delta \gg 1/K^{19}$.

Expanding the logarithm as a Taylor series around 1, we obtain

$$F(z) = 1 - \sum_{n=1}^{\infty} \frac{(-1)^n}{n(n+1)}(z - 1)^n,$$

which converges for $|z - 1| < 1$; so it is enough to show that, relative to the geodesic norm, the operator norm of $\exp(\text{ad}_X) \exp(\text{ad}_Y) - I$ is less than 1 for all $0 \leq t \leq 1$. We have

$$\| \exp(\text{ad}_X) \exp(\text{ad}_Y) - I \|_{\text{op}} \leq \| \exp(\text{ad}_X) (\exp(\text{ad}_Y) - I) + \exp(\text{ad}_X) - I \|_{\text{op}}$$

$$\leq \| \exp(\text{ad}_X) \|_{\text{op}} \| I - \exp(\text{ad}_Y) \|_{\text{op}} + \| \exp(\text{ad}_X) - I \|_{\text{op}}$$

$$\leq (\| \exp(\text{ad}_X) - I \|_{\text{op}} + \| I \|_{\text{op}}) \| I - \exp(\text{ad}_Y) \|_{\text{op}}$$

$$+ \| \exp(\text{ad}_X) - I \|_{\text{op}}$$

$$\leq (1 + 2\| \text{ad}_X \|_{\text{op}}) \cdot 2\| \text{ad}_Y \|_{\text{op}} + 2\| \text{ad}_X \|_{\text{op}}$$

$$= 2(\| \text{ad}_Y \|_{\text{op}} + 2\| \text{ad}_X \|_{\text{op}}) + 4\| \text{ad}_X \|_{\text{op}}\| \text{ad}_Y \|_{\text{op}}$$

$$\ll K^{19} (|t| \| Y \|_{\text{path},A} + \| X \|_{\text{path},A}) + |t| K^{38} \| X \|_{\text{path},A} \| Y \|_{\text{path},A}$$

$$\ll K^{19} (\| X \|_{\text{path},A} + \| Y \|_{\text{path},A}) + K^{38} \| X \|_{\text{path},A} \| Y \|_{\text{path},A}.$$

Hence there exists $\delta \gg 1/K^{19}$ such the integral converges whenever $X,Y$ have path norm less than $\delta$, as claimed. \qed

**Theorem 9.5.** Let $A$ be a strong $K$-approximate Lie group. There are $\delta, \varepsilon \gg 1/K^{19}$ such that, if

$$A_\varepsilon = \{ a \in A | \| a \|_{e,A} < \varepsilon \}$$

$$B_\delta = \{ X \in \mathfrak{a} | \| X \|_{\text{path},A} < \delta \}$$

85
then \( \exp(B_\delta) \cap A_\varepsilon \) is a normal sublocal group of \( A_\varepsilon \) (with reference neighborhood \( A_\varepsilon \) itself) and \( A_{\varepsilon/6C_0K^4}/\exp(B_\delta) \) is a discrete local \( K \)-approximate group.

**Proof.** Most of the work will be to show that \( \exp(B_\delta) \cap A_\varepsilon \) is indeed a sublocal normal group; once that is established, the Gleason lemmas (Theorem 7.7) readily imply that \( (A_{\varepsilon/6C_0K^4})^6 \subseteq A_\varepsilon \), and, since \( B_\delta \) is open, Lemma 9.1 shows that \( A_{\varepsilon/6C_0K^4}/\exp(B_\delta) \) is discrete.

For \( \delta \) we choose the number given by Proposition 9.4, and set \( \varepsilon = \delta/1000 \). Due to Proposition 9.4, for any \( X,Y \in B_\delta \) we have \( \exp(X) \exp(Y) = \exp(Z) \) for some \( Z \). Moreover, as in Proposition 9.2,

\[
Z = X + \int_0^1 F(\exp(\text{ad}_X) \exp(\text{ad}_Y))Y \, dt
\]

and therefore, reasoning as in the preceding Proposition,

\[
\|Z\|_{\text{geo},A} \leq \|X\|_{\text{geo},A} + \int_0^1 \left( 1 + \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left\| \exp(\text{ad}_X) \exp(\text{ad}_Y) - 1 \right\|_{\text{op}} \right) \|Y\|_{\text{geo},A} \, dt
\]

\[
\leq \|X\|_{\text{geo},A} + \int_0^1 \left( 1 + \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left[ t \left( O(K^{19}\|Y\|_{\text{path},A})
+ O(K^{38}\|X\|_{\text{path},A}\|Y\|_{\text{path},A}) \right)
+ O(K^{19}\|X\|_{\text{path},A})^n \right) \right] \|Y\|_{\text{geo},A} \, dt
\]

\[
\leq \|X\|_{\text{geo},A} + \left( 1 + \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left( O(K^{19}(\|X\|_{\text{path},A} + \|Y\|_{\text{path},A}))
+ O(K^{38}\|X\|_{\text{path},A}\|Y\|_{\text{path},A})^n \right) \right) \|Y\|_{\text{geo},A}
\]

\[
\leq 2 (\|X\|_{\text{geo},A} + \|Y\|_{\text{geo},A})
\]

\[
\leq 2 (\|X\|_{\text{path},A} + \|Y\|_{\text{path},A}.)
\]

It follows that \( \|Z\|_{\text{path},A} \leq O(K^4(\|X\|_{\text{path},A} + \|Y\|_{\text{path},A})) \), and by our choice of \( \delta \) this is less than 1. Now, by Lemma 8.9, if \( \exp(Z) \in A_\varepsilon \) then

\[
\|Z\|_{\text{path},A} \leq 1000 \|\exp(Z)\|_{e,A} < \delta
\]

and thus \( Z \in B_\delta \).
Now consider \( g \in A_\varepsilon \) and \( \exp(X) \in B_\delta \); we want to show that, if \( g \cdot \exp(X) \cdot g^{-1} \) lies in \( A_\varepsilon \), then it lies in \( B_\delta \). We have

\[
g \cdot \exp(X) \cdot g^{-1} = \exp(\text{Ad}_g X)
\]

and \( \| (\text{Ad}_g - I) X \|_{\text{path}, A} = O(K^{15}\|g\|_{e,A}) \) by Lemma 8.17. Using Lemma 8.13 we then get

\[
\| \text{Ad}_g X \|_{\text{path}, A} \leq C_0 K^4 \| X \|_{\text{path}, A} + O(K^{19}\|g\|_{e,A})
\]

and by our choice of \( \varepsilon \) and \( \delta \) this is less than 1. On the other hand, the Gleason lemmas (Theorem 7.7) tell us that \( \| g \cdot \exp(X) \cdot g^{-1} \|_{e,A} \leq 1000 \| \exp(X) \|_{e,A} \); thus if \( g \cdot \exp(X) \cdot g^{-1} \) lies in \( A_\varepsilon \) we may invoke Lemma 8.9 again to obtain

\[
\| \text{Ad}_g X \|_{\text{path}, A} \leq 1000 \| \exp(\text{Ad}_g X) \|_{e,A} < 1000\varepsilon \leq \delta
\]

so \( \exp(\text{Ad}_g X) \in \exp(B_\delta) \) as required.

Finally, recall how we gradually refined our object of study across this work. We started with a \( K \)-approximate group \( A \); passed to a strong \( O_K(1) \)-approximate group \( \tilde{A} \subseteq A^4 \); passed to a quotient \( \tilde{A}/H \) which is an \( O_K(1) \)-approximate Lie group; passed to a strong \( O_K(1) \)-approximate Lie group \( A' \); and, given \( \varepsilon > 0 \), passed to \( A'_\varepsilon \), the set of elements of \( A' \) of escape norm less than \( \varepsilon \). In each of these passages, \( O_K(1) \) left-translates of the next object suffice to cover the previous object. Thus, by Lemma 4.5, our Main Theorem follows.
REFERENCES


