Chapter 2

Group theory

2.1. Torsors

Given a (multiplicative) group G, a (left) *G-space* is a space X of states, together with an *action* of the group G that allows each group element $g \in G$ to transform any given state $x \in X$ to another state $gx \in X$, in a manner compatible with the group law (in particular, $ex = x$ for the group identity e, and $(gh)x = g(hx)$ for group elements g, h). One often also imposes additional compatibility conditions with other structures on the space (e.g., topological, differential, or algebraic structure).

A special case of a G-space is a *principal G-homogeneous space* or a *G-torsor*, defined as a G-space which is *uniquely transitive*, i.e., given any two states $x, y \in X$ there is a unique group element $g \in G$ such that $gx = y$; inspired by this equation, one can write g as y/x. A G-torsor can be viewed as a copy of the original group G, but one that does not necessarily have a preferred identity element[1].

Many natural concepts in mathematics and physics are more naturally torsor elements than group elements. Consider, for instance, the concept of *length*. In mathematics, one often ignores issues of units, and regards the length of a line segment as taking values in the non-negative real line $\mathbf{R}^+$; but in the absence of a preferred unit length scale, it is actually more natural to view length as taking values in some $\mathbf{R}^+$-torsor, say $\mathcal{L}$. To extract a non-negative real number for the length $|AB|$ of a line segment AB, one has to divide by some unit length $U \in \mathcal{L}$, such as a unit foot or a unit yard. For instance, if AB is 30 feet long, and U is a unit foot, then $|AB|/U = 30$.

[1] If there is a preferred identity or origin element O, then one can place the G-torsor in one-to-one correspondence with G by identifying gO with g for every group element g.

Observe that changing units is a passive transformation (see Section 2.2 below) rather than an active one, and as such behaves in the inverse manner to what one might naively expect. For instance, if one changes the unit of length U from feet to yards, which is a unit that is three times larger, then the numerical length $|AB|/U$ of AB *shrinks* by a factor of 3: AB is 30 feet long, but is only 10 yards long. Thus, while a unit yard is three times longer than a unit foot, the yard coordinate (the dual coordinate to the unit yard, which converts lengths to positive real numbers) is one third of the foot coordinate. (See Section 6.3 for further discussion.)

More generally, one can use torsors to rigorously set up the physical concepts of *units* and *dimensional analysis*. The product of two lengths in $\mathcal{L}$ is not another length, but instead takes values in another torsor, the torsor $\mathcal{L}^2 = \mathcal{L} \otimes_{\mathbf{R}^+} \mathcal{L}$ of areas. One can use the square U^2 of the unit length as a unit area. The assertion that physical laws have to be dimensionally consistent is then equivalent to the assertion that they are invariant with respect to the passive transformation of changing the units.

Much as dimensional units such as length or mass are torsors for the non-negative reals, points in space are torsors for the translation group $\mathbf{R}^3$, and (oriented) spatial coordinate frames are torsors for the linear group[2] $SL_3(\mathbf{R})$ (if the origin is fixed) or $SL_3(\mathbf{R}) \ltimes \mathbf{R}^3$ (otherwise), and so forth. Indeed, one can view basis vectors and coordinate systems as higher-dimensional analogues of units and unit measurement coordinates, respectively.

If one works with spacetime coordinate frames rather than spatial coordinate frames, then the situation is similar, but the structure group will be different (e.g., the Galilean group for Galilean relativity, the Poincaré or Lorentz group for special relativity, or the diffeomorphism group for general relativity).

Viewing group elements as quotients of torsors is sometimes helpful when trying to visualise operations such as conjugation $h \to ghg^{-1}$; one can interpret this operation as that of moving both the observer and the object by g. For instance, consider the *lamplighter group* $\mathbf{Z}/2\mathbf{Z} \wr \mathbf{Z}$, the *wreath product* of $\mathbf{Z}/2\mathbf{Z}$ with $\mathbf{Z}$. One can define this group by using as a state space X the configuration space of a doubly infinite sequence of lamps (indexed by the integers), with each lamp being either "on" or "off", and with at most finitely many of the lamps being "on", together with the position of a lamplighter, located at one of the lamps; more formally, we have $X := (\mathbf{Z}/2\mathbf{Z})_0^{\mathbf{Z}} \times \mathbf{Z}$, where $(\mathbf{Z}/2\mathbf{Z})_0^{\mathbf{Z}}$ is the space of compactly supported sequences from $\mathbf{Z}$ to $\mathbf{Z}/2\mathbf{Z}$. The lamplighter has the ability to toggle the lamp on and off at his or her current location, and also has the ability to move left or right. The

[2] If one insists on the coordinate frames being orthogonal, then the relevant group becomes $SO_3(\mathbf{R})$ or the Euclidean group $SO_3(\mathbf{R}) \ltimes R^3$, as appropriate.

lamplighter group G is then the group of transformations on the state space X that is generated by the following operations:

- e: Move the lamplighter one unit to the right.
- e^{-1}: Move the lamplighter one unit to the left.
- f: Toggle the lamp at the current location of the lamplighter.

It is not hard to show that X then becomes a G-torsor.

One way to describe a group element of G is then to describe an initial state A in X and a final state B in X, and then define B/A to be the unique group element that transforms A to B; one can view B/A as a "program" (e.g. made up of a string of e's, e^{-1}'s, and f's, or perhaps expressed in a more "high-level" language) that one could give to a lamplighter that is currently in the system A that would then transform it to B. Note that multiple programs can give the same group element, for instance, $fefe^{-1}$ is the same element of G as $efe^{-1}f$. Also, multiple pairs A, B can give rise to the same element[3] B/A.

One can express any such "program" B/A in a canonical form as that of "change the set S of lights, as described using one's current location, and then move n steps to the right (or left)". This expresses the lamplighter group as a *semi-direct product* of $(\mathbf{Z}/2\mathbf{Z})_0^{\mathbf{Z}}$ and $\mathbf{Z}$. If the lamplighter position does not change between A and B, then the program is simply that of changing a set of lights, and B/A now lives in the abelian subgroup $(\mathbf{Z}/2\mathbf{Z})_0^{\mathbf{Z}}$ of the lamplighter group.

If one conjugates a group element B/A by another group element g, one obtains the new group element $g(B/A)g^{-1} = (gB)/(gA)$. A little thought then reveals that the program needed to execute $(gB)/(gA)$ is similar to that for B/A, except that the set of lights S that one needs to change has been modified. As such, we see that the commutator $[g, B/A] = ((gB)/(gA))/(B/A)$ is an element of the abelian subgroup $(Z/2Z)_0^Z$, making the lamplighter group *metabelian* and thus *solvable*.

I found this sort of torsor-oriented perspective useful when thinking about such concepts as that of a harmonic function on a group G (something that comes up, for instance, in modern proofs of Gromov's theorem regarding groups of polynomial growth; see Section 2.5). One can instead think about a harmonic function on a G-torsor X, defined as an "energy functional" on such a space with the property that the energy of any state is equal to the average energy of the neighbouring states (at least if the group G is discrete; for continuous groups, one has to neglect higher order terms). If the group G is not commutative, then actively transforming the

[3] This perspective is a generalisation of the standard way of visualising a spatial vector as an arrow from one spatial point A to another spatial point B.

states can destroy the harmonicity property; but passively transforming the states does not. It is because of this that the space of (right)-harmonic functions still has a left G-action, and vice versa.

2.2. Active and passive transformations

Consider the following (somewhat informally posed) questions:

Question 2.2.1. *Let T be equilateral triangle in a plane whose vertices are labeled 1, 2, 3 in clockwise order. Define the following two operations on this triangle:*

- *F: Flip the triangle to swap the vertices 2 and 3, while keeping 1 fixed.*
- *R: Rotate the triangle clockwise in the plane by 120 degrees.*

Do the operations F and R commute, i.e., does $F \circ R = R \circ F$?

Question 2.2.2. *Suppose one is viewing some text on a computer screen. The text is so long that one cannot display all of it at once on a screen: currently, only a middle portion of the text is visible. To see more of the text, we press the "up" arrow key (or click the "up" button). When one does so, does the text on the screen move up or down?*

We discuss Question 2.2.2 first. The answer depends on the user interface model. Most such models are *passive transformations*; the "up" command moves the *observer* up, and one's view of the text then moves *down* as a consequence. A minority of models (such as "hand" tools in various pieces of software) are instead *active transformations*; dragging a hand tool upwards causes the text to move upward (keeping the observer position fixed).

In some cases, the model used may be ambiguous at first. If one is viewing a map, does the "+" key cause the map image on the screen to enlarge, or shrink? Somewhat confusingly, two different pieces of mapping software can respond in opposite ways to such a command; some use active transformation models ("+" makes the world bigger, so that less of the world remains inside the viewscreen), and others use passive transformation models ("+" makes the *observer* bigger, so that more of the world can now fit inside the viewscreen).

This question is a special case of the double action of a group G on itself (or more generally, on a left G-torsor, as discussed in Section 2.1). Imagine that inside a group G (or left G-torsor) one has an observer O and an object X; there exists a unique group element $g = X/O$ such that $X = gO$, and in that case we say that X has an *apparent position* of g from the perspective of the observer O.

We can then change this apparent position in two different ways. First, we may apply an active transformation, and shift the object X by a group element h to move it to $hX = hgO$; the apparent position of the object then shifts from g to hg, and so the active transformation corresponds to the *left* action of the group G on itself.

Or, we may apply a passive transformation, and shift the *observer* O by a group element h to move it to hO; since $X = gO = gh^{-1}(hO)$, we see that the apparent position shifts from g to gh^{-1}. Thus the passive transformation corresponds to the *right* action of G on itself.

Note the presence of the inverse in the passive transformation; it is this inverse which is the source of the confusion mentioned above.

Even if the group G is non-commutative, the left-action of G and the right-action of G will still commute with each other, as moving an object and moving the observer can be done in either order without any difference to the final state of the system.

Now we can answer Question 2.2.1. The answer is that it depends on whether one interprets the operations F and R as active operations or passive operations; the phrasing of the question makes it ambiguous.

For instance, we can treat the flip operation F as an active transformation, by viewing the labels 1, 2, 3 as being attached[4] to the triangle object being manipulated. As one applies F, the labels 1 and 2 physically change locations.

Or, one can view the flip operation F as a passive transformation, caused by motion of the observer rather than the object. Here, the labels 1, 2, 3 are attached to the observer rather than to the object (and are usually displayed *outside* the triangle). The operation F flips the triangle, but the labels 1, 2, 3 remain where they are.

The difference between the two interpretations of F becomes apparent once the object's value of 3 moves away from the observer's value of 3, as they are then flipping the triangle across different axes.

Similarly, one can view the rotation operation R as an active rotation, in which the triangle is physically rotated in the direction which is clockwise in its own orientation, or as a passive rotation in which the triangle is rotated in the direction which is clockwise in the observer's orientation (or equivalently, the observer is rotated in a *counterclockwise* direction).

The difference between the two interpretations of R becomes apparent once the triangle is flipped over, so that its orientation is the opposite of that of the observer.

[4] This is usually drawn by placing the labels 1, 2, 3 *inside* the triangle.

If F is active and R is passive (or vice versa), then the transformations commute. But if F and R are both active or both passive, then the transformations do not commute.

In order to fully remove all ambiguity from the system, one needs to label the vertices of the triangle *twice*: first by an "external" labeling (e.g., A, B, C) which is not affected by any of the transformations, and second, by an "internal" labeling (e.g., 1, 2, 3) which moves with the operations being applied. Traditionally, the external labels are displayed outside the triangle, and the internal vertices are displayed inside the triangle. Similarly, one needs to display an external orientation (e.g., a counterclockwise arrow, displayed outside the triangle) that is not affected by the operations being applied, and also an internal orientation (e.g., another counterclockwise arrow, displayed inside the triangle) that can get flipped over by the operations being applied. There are then four operations of interest:

(1) Active-F: Flip the triangle across the axis given by the vertex internally labeled 3, thus swapping the vertices internally labeled 1, 2, and also reversing the internal orientation.

(2) Passive-F: Flip the triangle across the axis given by the vertex externally labeled C, thus swapping the internal labels of the vertices that are externally labeled A, B, and also reversing the internal orientation.

(3) Active-R: Rotate the triangle by 120 degrees in the internally clockwise direction, moving the internal labels appropriately.

(4) Passive-R: Rotate the triangle by 120 degrees in the externally clockwise direction, moving the internal labels appropriately.

Then the two passive transformations commute with the two active transformations, but the two passive transformations do not commute with each other, and neither do the two active transformations. (It is instructive to work this out on paper, physically cutting out a triangle if necessary.)

2.3. Cayley graphs and the geometry of groups

In most undergraduate courses, groups are first introduced as a primarily *algebraic* concept—a set equipped with a number of algebraic operations (group multiplication, multiplicative inverse, and multiplicative identity) and obeying a number of rules of algebra (most notably the associative law). It is only somewhat later that one learns that groups are not solely an algebraic object, but can also be equipped with the structure of a manifold (giving rise to *Lie groups*) or a topological space (giving rise to *topological groups*). (See also [**Ta2010b**, §1.14] for a number of other ways to think about groups.)

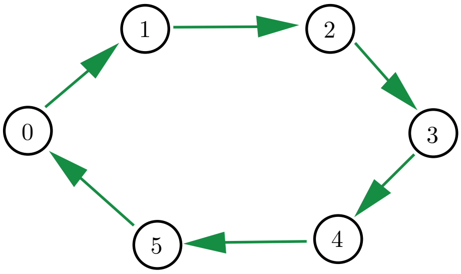

Figure 1. Cayley graph of $\mathbf{Z}/6\mathbf{Z}$ with generator 1 (in green).

Another important way to enrich the structure of a group G is to give it some *geometry*. A fundamental way to provide such a geometric structure is to specify a list of generators S of the group G. Let us call such a pair (G, S) a *generated group*; in many important cases the set of generators S is finite, leading to a *finitely generated group*. A generated group (G, S) gives rise to the *word metric* $d : G \times G \to \mathbf{N}$ on G, defined to be the maximal metric for which $d(x, sx) \leq 1$ for all $x \in G$ and $s \in S$ (or more explicitly, $d(x, y)$ is the least m for which $y = s_1^{\epsilon_1} \dots s_m^{\epsilon_m} x$ for some $s_1, \dots, s_m \in S$ and $\epsilon_1, \dots, \epsilon_m \in \{-1, +1\}$). This metric then generates the balls $B_S(R) := \{x \in G : d(x, \mathrm{id}) \leq R\}$. In the finitely generated case, the $B_S(R)$ are finite sets, and the rate at which the cardinality of these sets grow in R is an important topic in the field of *geometric group theory*. The idea of studying a finitely generated group via the geometry of its metric goes back at least to the work of Dehn [**De1912**].

One way to visualise the geometry of a generated group is to look at the (labeled) *Cayley colour graph* (or *Cayley graph*, for short) of the generated group (G, S). This is a directed coloured graph, with edges coloured by the elements of S, and vertices labeled by elements of G, with a directed edge of colour s from x to sx for each $x \in G$ and $s \in S$. The word metric then corresponds to the graph metric of the Cayley graph. See, for instance, Figure 1 and Figure 2.

We can thus see that the same group can have somewhat different geometry if one changes the set of generators. For instance, in a large cyclic group $\mathbf{Z}/N\mathbf{Z}$, with a single generator $S = \{1\}$ the Cayley graph "looks one-dimensional", and balls $B_S(R)$ grow linearly in R until they saturate the entire group, whereas with two generators $S = \{s_1, s_2\}$ chosen at random, the Cayley graph "looks two-dimensional", and the balls $B_S(R)$ typically grow quadratically until they saturate the entire group.

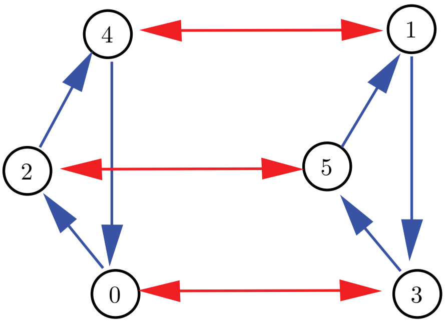

Figure 2. Cayley graph of $\mathbf{Z}/6\mathbf{Z}$ with generators 2 (in blue) and 3 (in red).

Cayley graphs have three distinguishing properties:

- (Regularity) For each colour $s \in S$, every vertex x has a single s-edge leading out of x, and a single s-edge leading into x.
- (Connectedness) The graph is connected.
- (Homogeneity) For every pair of vertices x, y, there is a unique coloured graph isomorphism that maps x to y.

It is easy to verify that a directed coloured graph is a Cayley graph (up to relabeling) if and only if it obeys the above three properties. Indeed, given a graph (V, E) with the above properties, one sets G to equal the (coloured) automorphism group of the graph (V, E); arbitrarily designating one of the vertices of V to be the identity element id, we can then identify all the other vertices in V with a group element. One then identifies each colour $s \in S$ with the vertex that one reaches from id by an s-coloured edge. Conversely, every Cayley graph of a generated group (G, S) is clearly regular, is connected because S generates G, and has isomorphisms given by right multiplication $x \mapsto xg$ for all $g \in G$. (The regularity and connectedness properties already ensure the uniqueness component of the homogeneity property.)

From the above equivalence, we see that we do not really need the vertex labels on the Cayley graph in order to describe a generated group, and so we will now drop these labels and work solely with *unlabeled* Cayley graphs, in which the vertex set is not already identified with the group. As we saw above, one just needs to designate a marked vertex of the graph as the "identity" or "origin" in order to turn an unlabeled Cayley graph into a labeled Cayley graph; but from homogeneity we see that all vertices of an unlabeled Cayley graph "look the same" and there is no canonical preference for choosing one vertex as the identity over another. I prefer here to keep the graphs unlabeled to emphasise the homogeneous nature of the graph.

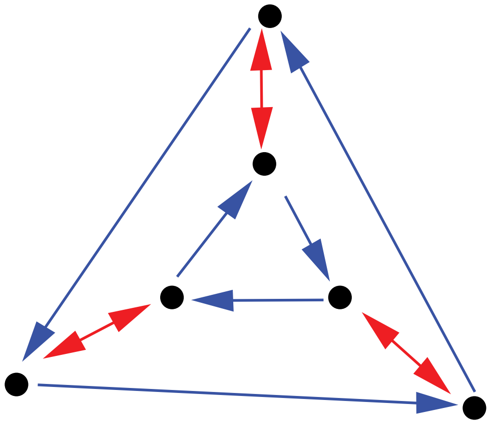

Figure 3. Cayley graph of S_3.

It is instructive to revisit the basic concepts of group theory using the language of (unlabeled) Cayley graphs, and to see how geometric many of these concepts are. In order to facilitate the drawing of pictures, I work here only with small finite groups (or Cayley graphs), but the discussion certainly is applicable to large or infinite groups (or Cayley graphs) also.

For instance, in this setting, the concept of *abelianness* is analogous to that of a *flat* (zero-curvature) geometry: given any two colours s_1, s_2, a directed path with colours $s_1, s_2, s_1^{-1}, s_2^{-1}$ (adopting the obvious convention that the reversal of an s-coloured directed edge is considered an s^{-1}-coloured directed edge) returns to where it started[5]. Thus, for instance, the two depictions of $\mathbf{Z}/6\mathbf{Z}$ above are abelian, whereas the group S_3, which is also the dihedral group of the triangle, and thus admits the Cayley graph depicted in Figure 3, is not abelian.

A subgroup (G', S') of a generated group (G, S) can be easily described in Cayley graph language if the generators S' of G' happen to be a subset of the generators S of G. In that case, if one begins with the Cayley graph of (G, S) and erases all colours except for those colours in S', then the graph *foliates* into connected components, each of which is isomorphic to the Cayley graph of (G', S'). For instance, in the above Cayley graph depiction of S_3, erasing the blue colour leads to three copies of the red Cayley graph (which has $\mathbf{Z}/2\mathbf{Z}$ as its structure group), while erasing the red colour leads to two copies of the blue Cayley graph (which as $A_3 \equiv \mathbf{Z}/3\mathbf{Z}$ as its structure group). If S' is not contained in S, then one has to first “change basis” and add or remove some coloured edges to the original Cayley graph before one can obtain this formulation (thus, for instance, S_3 contains two more

[5] Note that a generated group (G, S) is abelian if and only if the generators in S pairwise commute with each other.

subgroups of order two that are not immediately apparent with this choice of generators). Nevertheless, the geometric intuition that subgroups are analogous to foliations is still quite a good one.

We saw that a subgroup (G', S') of a generated group (G, S) with $S' \subset S$ foliates the larger Cayley graph into S'-connected components, each of which is a copy of the smaller Cayley graph. The remaining colours in S then join those S'-components to each other. In some cases, each colour $s \in S \backslash S'$ will connect an S'-component to exactly one other S'-component; this is the case for instance when one splits S_3 into two blue components. In other cases, a colour s can connect a S'-component to multiple S'-components; this is the case for instance when one splits S_3 into three red components. The former case occurs precisely when[6] the subgroup G' is *normal.* We can then *quotient out* the (G', S') Cayley graph from (G, S), leading to a quotient Cayley graph $(G/G', S\backslash S')$ whose vertices are the S'-connected components of (G, S), and the edges are projected from (G, S) in the obvious manner. We can then view the original graph (G, S) as a *bundle* of (G', S')-graphs over a base $(G/G', S\backslash S')$-graph (or equivalently, an *extension* of the base graph $(G/G', S\backslash S')$ by the fibre graph (G', S')); for instance, S_3 can be viewed as a bundle of the blue graph A_3 over the red graph $\mathbf{Z}/2\mathbf{Z}$, but not conversely. We thus see that the geometric analogue of the concept of a normal subgroup is that of a *bundle.* The generators in $S\backslash S'$ can be viewed as describing a *connection* on that bundle.

Note, though, that the structure group of this connection is not simply G', unless G' is a *central* subgroup; instead, it is the larger group $G' \rtimes \mathrm{Aut}(G')$, the semi-direct product of G' with its automorphism group. This is because a non-central subgroup G' can be "twisted around" by operations such as conjugation $g' \mapsto sg's^{-1}$ by a generator $s \in S$. So central subgroups are analogous to the geometric notion of a *principal bundle.* For instance, Figure 4 depicts the Heisenberg group

$$\begin{pmatrix} 1 & \mathbf{F}_2 & \mathbf{F}_2 \\ 0 & 1 & \mathbf{F}_2 \\ 0 & 0 & 1 \end{pmatrix}$$

over the field $\mathbf{F}_2$ of two elements, which one can view as a central extension of $\mathbf{F}_2^2$ (the blue and green edges, after quotienting) by $\mathbf{F}_2$ (the red edges). Note how close this group is to being abelian; more generally, one can think of nilpotent groups as being a slight perturbation of abelian groups.

In the case of S_3 (viewed as a bundle of the blue graph A_3 over the red graph $\mathbf{Z}/2\mathbf{Z}$), the base graph $\mathbf{Z}/2\mathbf{Z}$ is in fact embedded (three times) into the large graph S_3. More generally, the base graph $(G/G', S\backslash S')$ can be lifted

[6] Note that a subgroup G' of a generated group (G, S) is normal if and only if left-multiplication by a generator of S maps right-cosets of G' to right-cosets of G'.

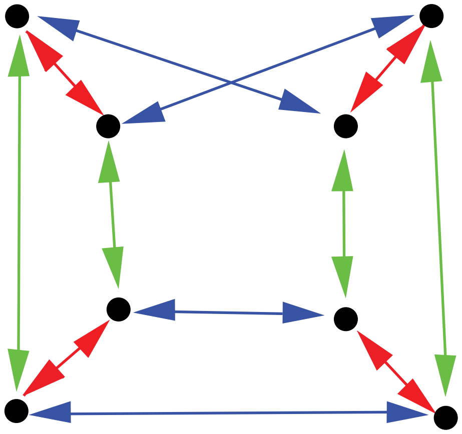

Figure 4. The Heisenberg group over $\mathbf{F}_2$.

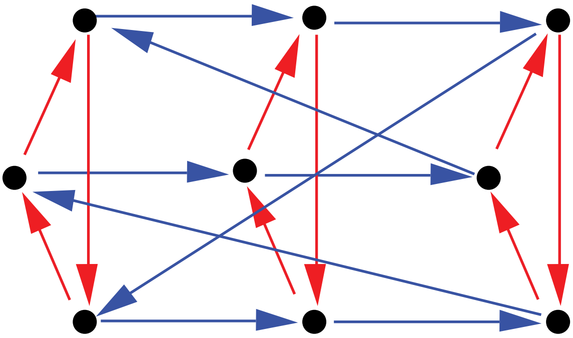

Figure 5. The group $\mathbf{Z}/9\mathbf{Z}$, with generators 1 (in blue) and 3 (in red).

back into the extension (G, S) if and only if the short exact sequence $0 \to G' \to G \to G/G' \to 0$ *splits*, in which case G becomes a *semi-direct product* $G \equiv G' \rtimes H$ of G' and a lifted copy H of G/G'. Not all bundles can be split in this fashion. For instance, consider the group $\mathbf{Z}/9\mathbf{Z}$ depicted in Figure 5. This is a $\mathbf{Z}/3\mathbf{Z}$-bundle over $\mathbf{Z}/3\mathbf{Z}$ that does not split; the blue Cayley graph of $\mathbf{Z}/3\mathbf{Z}$ is not visible in the $\mathbf{Z}/9\mathbf{Z}$ graph directly, but only after one quotients out the red fibre subgraph. The notion of a splitting in group theory is analogous to the geometric notion of a *global gauge* (see [**Ta2009b**, §1.4]). The existence of such a splitting or gauge, and the relationship between two such splittings or gauges, are controlled by the *group cohomology* of the sequence $0 \to G' \to G \to G/G' \to 0$.

Even when one has a splitting, the bundle need not be completely trivial, because the bundle is not principal, and the connection can still twist the fibres around. For instance, S_3 when viewed as a bundle over $\mathbf{Z}/2\mathbf{Z}$ with fibres A_3 splits, but observe that if one uses the red generator of this splitting

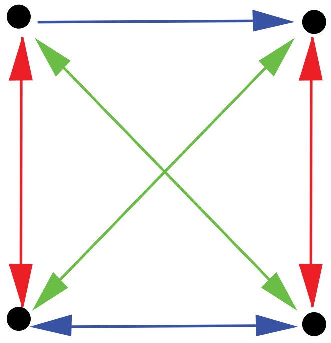

Figure 6. The Klein four-group $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$..

to move from one copy of the blue A_3 graph to the other, that the orientation of the graph changes. The bundle is trivialisable if and only if G' is a *direct summand* of G, i.e., G splits as a direct product $G = G' \times H$ of a lifted copy H of G/G'. Thus we see that the geometric analogue of a direct summand is that of a trivialisable bundle (and that trivial bundles are then the analogue of direct products). Note that there can be more than one way to trivialise a bundle. For instance, with the Klein four-group $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ depicted in Figure 6, the red fibre $\mathbf{Z}/2\mathbf{Z}$ is a direct summand, but one can use either the blue lift of $\mathbf{Z}/2\mathbf{Z}$ or the green lift of $\mathbf{Z}/2\mathbf{Z}$ as the complementary factor.

2.4. Group extensions

In mathematics, one frequently starts with some space X and wishes to *extend* it to a larger space Y. Generally speaking, there are two ways in which one can extend a space X:

- By *embedding* X into a space Y that has X (or at least an isomorphic copy of X) as a *subspace*.
- By *covering* X by a space Y that has X (or an isomorphic copy thereof) as a *quotient*.

For many important categories of interest (such as *abelian categories*), the former type of extension can be represented by the *exact sequence*,

$$0 \to X \to Y$$

and the latter type of extension can be represented by the exact sequence

$$Y \to X \to 0.$$

In some cases, X can be both embedded in, and covered by, Y, in a consistent fashion; in such cases we sometimes say that the above exact sequences *split*.

An analogy would be to that of digital images. When a computer represents an image, it is limited both by the scope of the image (what it is picturing), and by the resolution of an image (how much physical space is represented by a given pixel). To make the image "larger", one could either *embed* the image in an image of larger scope but equal resolution (e.g., embedding a picture of a 200×200 pixel image of a person's face into a 800×800 pixel image that covers a region of space that is four times larger in both dimensions, e.g., the person's upper body) or *cover* the image with an image of higher resolution but of equal scope (e.g., enhancing a 200×200 pixel picture of a face to a 800×800 pixel of the same face). In the former case, the original image is a *sub-image* (or *cropped image*) of the extension, but in the latter case the original image is a *quotient* (or a *pixelation*) of the extension. In the former case, each pixel in the original image can be identified with a pixel in the extension, but not every pixel in the extension is covered. In the latter case, every pixel in the original image is *covered* by several pixels in the extension, but the pixel in the original image is not canonically identified with any particular pixel in the extension that covers it; it "loses its identity" by dispersing into higher resolution pixels.

Remark 2.4.1. Note that "zooming in" the visual representation of an image by making each pixel occupy a larger region of the screen neither increases the scope or the resolution; in this language, a zoomed-in version of an image is merely an *isomorphic copy* of the original image; it carries the same amount of information as the original image, but has been represented in a new *coordinate system* which may make it easier to view.

In the study of a given category of spaces (e.g., topological spaces, manifolds, groups, fields, etc.), embedding and coverings are both important; this is particularly true in the more topological areas of mathematics, such as manifold theory. But typically, the term *extension* is reserved for just one of these two operations. For instance, in the category of fields, coverings are quite trivial; if one covers a field k by a field l, the kernel of the covering map $\pi : l \to k$ is necessarily trivial and so k, l are in fact isomorphic. So in field theory, a *field extension* refers to an embedding of a field, rather than a covering of a field. Similarly, in the theory of metric spaces, there are no non-trivial isometric coverings of a metric space, and so the only useful notion of an extension of a metric space is the one given by embedding the original space in the extension.

On the other hand, in group theory (and in group-like theories, such as the theory of dynamical systems, which studies group actions), the term "extension" is reserved for coverings, rather than for embeddings. I think one of the main reasons for this is that coverings of groups automatically generate a special type of embedding (a *normal* embedding), whereas most embeddings don't generate coverings. More precisely, given a group extension G of a base group H,

$$G \to H \to 0,$$

one can form the *kernel* $K = \ker(\phi)$ of the covering map $\pi : G \to H$, which is a normal subgroup of G, and we thus can extend the above sequence canonically to a *short exact sequence*

$$0 \to K \to G \to H \to 0.$$

On the other hand, an embedding of K into G,

$$0 \to K \to G$$

does not similarly extend to a short exact sequence unless the the embedding is normal.

Another reason for the notion of extension varying between embeddings and coverings from subject to subject is that there are various natural *duality operations* (and more generally, *contravariant functors*) which turn embeddings into coverings and vice versa. For instance, an embedding of one vector space V into another W induces a covering of the *dual space* V^* by the dual space W^*, and conversely; similarly, an embedding of a locally compact abelian group H in another G induces a covering of the *Pontryagin dual* $\hat{H}$ by the Pontryagin dual $\hat{G}$. In the language of images, embedding an image in an image of larger scope is largely equivalent to covering the Fourier transform of that image by a transform of higher resolution, and conversely; this is ultimately a manifestation of the basic fact that frequency is inversely proportional to wavelength.

Similarly, a common duality operation arises in many areas of mathematics by starting with a space X and then considering a space $C(X)$ of functions on that space (e.g., continuous real-valued functions, if X was a topological space, or in more algebraic settings one could consider homomorphisms from X to some fixed space). Embedding X into Y then induces a covering of $C(X)$ by $C(Y)$, and conversely, a covering of X by Y induces an embedding of $C(X)$ into $C(Y)$. Returning again to the analogy with images, if one looks at the collection of *all* images of a fixed scope and resolution, rather than just a single image, then increasing the available resolution causes an *embedding* of the space of low-resolution images into the space of high-resolution images (since of course every low-resolution image is an example of a high-resolution image), whereas increasing the available scope causes a *covering* of the space of narrow-scope images by the space of wide-scope images (since every wide-scope image can be *cropped* into a narrow-scope image). Note in the case of images, that these extensions can be split: not only can a low-resolution image be viewed as a special case of a high-resolution image, but any high-resolution image can be *pixelated* into a low-resolution one. Similarly, not only can any wide-scope image be cropped into a narrow-scope image, a narrow-scope image can be extended to a wide-scope image simply by filling in all the new areas of scope with black (or by using more advanced image processing tools to create a more visually pleasing extension). In the category of sets, the statement that every covering can be split is precisely the *axiom of choice*.

I've recently found myself having to deal quite a bit with group extensions in my research, so I have decided to make some notes on the basic theory of such extensions. This is utterly elementary material for a group theorist, but I found this task useful for organising my own thoughts on this topic, and also in pinning down some of the jargon in this field.

2.4.1. Basic concepts.

Definition 2.4.2 (Group extension)**.** An *extension* of a group H is a group G, together with a surjective *projection map* (or *covering map*) $\pi : G \to H$. If the kernel of π can be identified with (i.e. is isomorphic to) a group K, we say that G is an *extension* of H by K, and we have the short exact sequence

$$0 \to K \to G \to H \to 0.$$

If the group K has some property $\mathcal{P}$, we say that G is a $\mathcal{P}$ extension of H. Thus, for instance, if K is abelian, G is an abelian extension of H; if K is central (in G), G is a central extension of H; and so forth. We refer to H as the *base* of the extension, and K as the *fibre*, and refer to H and K collectively as *factors* of G.

If K has some property $\mathcal{P}$, and H has some property $\mathcal{Q}$, then we say that[7] G is *$\mathcal{P}$-by-$\mathcal{Q}$*. Thus, for instance, G is abelian-by-finite if K is abelian and H is finite, but finite-by-abelian if K is finite and H is abelian.

One can think of a K-by-H group as a group that looks like H "at large scales" and like K "at small scales"; one can also view this group as a principal K-bundle over H.

There are several ways to generate a group extension $G \to H \to 0$. First, given any homomorphism $\pi : G \to G'$ from one group G to another, the *homomorphism theorem* tells us that G is an extension of the image $\pi(G)$, with kernel $\ker(\pi)$:

$$0 \to \ker(\pi) \to G \to \pi(G) \to 0.$$

Conversely, every group extension arises in this manner.

[7] I have no idea why the order is traditionally arranged in this way; I would have thought that extending a $\mathcal{Q}$ group by a $\mathcal{P}$ group would give a $\mathcal{P}$-by-$\mathcal{Q}$ group, rather than the other way around; perhaps at one point the idea of a normal embedding was considered more important than a group extension. Nevertheless, the notation seems to be entrenched by now.

A group extension $\pi : G \to H$ *splits* if there is a homomorphism $\phi : H \to G$ such that $\pi(\phi(h)) = h$ for all $h \in H$. In this case, H acts on the kernel K by conjugation (after identifying H with $\phi(H)$); denoting this action by ρ (thus $\rho(h)k := \phi(h)k\phi(h)^{-1}$), we can then canonically identify G with the *semi-direct product* $K \rtimes_\rho H$, defined as the set of pairs (k, h) with $k \in K$, $h \in H$, with the group law $(k, h)(k', h') := (k\rho(h)(k'), hh')$, by identifying (k, h) with $k\phi(h)$. Conversely, every semi-direct product $K \rtimes_\rho H$ is a group extension of H by K which splits. If the conjugation action ρ is trivial, then the semi-direct product simplifies to the *direct product* $K \times H$. In particular, any semi-direct product which is a central extension is of this form.

Note that, in general, an extension of H by K is a different concept from an extension of K by H, because one can have H as a normal subgroup but not as a quotient, or vice versa. For instance, S_3 has A_3 as a normal subgroup, but not as a quotient; S_3 is an extension of $\mathbf{Z}/2\mathbf{Z}$ by A_3, but not vice versa. To put it another way, the operator "-by-" is not commutative: H-by-K is a different concept from K-by-H.

A subgroup L of an K-by-H group G is automatically an K'-by-H' group for some subgroups H', K' of H, K respectively; this is essentially *Goursat's lemma.* Indeed, one can take $K' := K \cap L$ and $H' := \pi(L)$, where $\pi : G \to H$ is the projection map. Furthermore, the index of the subgroup is the product of the index of H' in H, and the index of K' in K.

Some standard notions in group theory can be defined using group extensions:

(1) A *metabelian group* is the same thing as an abelian-by-abelian group, i.e., an abelian extension of an abelian group.

(2) A *metacyclic group* is the same thing as an cyclic-by-cyclic group, i.e., a cyclic extension of a cyclic group.

(3) A *polycyclic group* is defined recursively by declaring the trivial group to be polycyclic of length 0, and defining a polycyclic group of length l to be an extension of a cyclic group by a polycyclic group of length $l - 1$. Thus polycyclic groups are polycyclic-by-cyclic, where the polycyclic factor has a shorter length than the full group.

(4) A *supersolvable group* is defined recursively by declaring the trivial group to be supersolvable of length 0, and defining a supersolvable group of length l to be a cyclic extension supersolvable group of length $l-1$. Thus supersolvable groups are cyclic-by-supersolvable, where the supersolvable factor has a shorter length that the full group. In other words, supersolvable groups are towers of cyclic extensions.

(5) A *solvable group* is defined recursively by declaring the trivial group to be solvable of length 0, and defining a solvable group of length l to be an extension of an abelian group by a solvable group of length $l-1$. Thus solvable groups are solvable-by-abelian, where the solvable factor has a shorter length. One can equivalently define solvable groups as abelian-by-solvable, where the solvable factor again has a shorter length (because the final term in the *derived series* is abelian and normal). In other words, a solvable group is a tower of abelian extensions.

(6) A *nilpotent group* is defined recursively by declaring the trivial group to be nilpotent of step 0, and defining a nilpotent group of step s to be a central extension of a nilpotent group of step $s-1$, thus nilpotent groups are central-by-nilpotent. In other words, a nilpotent group is a tower of central extensions.

Remark 2.4.3. The inclusions here are: cyclic implies abelian implies metabelian implies solvable, cyclic implies metacyclic implies supersolvable implies polycyclic implies solvable, metacyclic implies metabelian, and abelian implies nilpotent implies solvable.

The trivial group is the identity for the "-by-" operator: trivial-by-$\mathcal{P}$ or $\mathcal{P}$-by-trivial is the same thing as $\mathcal{P}$.

Now we comment on the associativity of the "-by-" operator. If N, H, K are groups, observe that an N-by-(H-by-K) group (i.e., an extension of an H-by-K group by N) is automatically an (N-by-H)-by-K group (i.e., an extension of K by an N-by-H group), since if we denote G by the N-by-(H-by-K) group, and π the quotient map from G to the H-by-K group, then $\pi^{-1}(H)$ is a N-by-H normal subgroup of G whose quotient is K. Thus, for instance, every cyclic-by-metacyclic group is metacyclic-by-cyclic, and more generally, every supersolvable group is polycyclic.

On the other hand, the converse is not true: not every (N-by-H)-by-K group is an N-by-(H-by-K) group. The problem is that N is normal in the N-by-H group, but need not be normal in the (N-by-H)-by-K group. For instance, the semi-direct product $\mathbf{Z}^2 \rtimes SL_2(\mathbf{Z})$ is ($\mathbf{Z}$-by-$\mathbf{Z}$)-by-$SL_2(\mathbf{Z})$ but not $\mathbf{Z}$-by-($\mathbf{Z}$-by-$SL_2(\mathbf{Z})$). So the "-by-" operation is not associative in general (for instance, there are polycyclic groups that are not supersolvable). However, if N is not just normal in the N-by-H group, but is *characteristic* in that group (i.e., invariant under all (outer) automorphisms of that group), then it is automatically normal in the larger (N-by-H)-by-K group, and then one can interpret the (N-by-H)-by-K group as an N-by-(H-by-K) group. So one recovers associativity when the first factor is characteristic. This explains why solvable groups can be recursively expressed both as abelian-by-solvable, and equivalently as solvable-by-abelian; this is ultimately because the *commutator subgroup* $[G, G]$ is a characteristic subgroup of G. An easy but useful related observation is that solvable-by-solvable groups are again solvable (with the length of the product being bounded by the sum of the length of the factors).

Given a group property $\mathcal{P}$, a group G is said to be *virtually* $\mathcal{P}$ if it has a finite index subgroup with the property $\mathcal{P}$; thus, for instance, a virtually abelian group is one with a finite index abelian subgroup, and so forth. As another example, "finite" is the same as "virtually trivial". The property of being virtually $\mathcal{P}$ is not directly expressible in terms of group extensions for arbitrary properties $\mathcal{P}$; however, if the group property $\mathcal{P}$ is *hereditary* in the sense that subgroups of a $\mathcal{P}$ group are also $\mathcal{P}$, then a virtually $\mathcal{P}$ group is the same concept as a $\mathcal{P}$-by-finite group. This is because every finite index subgroup H of a group G automatically contains[8] a finite index *normal* subgroup of G.

One also observes that if $\mathcal{P}$, $\mathcal{Q}$ are hereditary properties, then the property of $\mathcal{P}$-by-$\mathcal{Q}$ is hereditary also; if $0 \to P \to G \to Q \to 0$ is a $\mathcal{P}$-by-$\mathcal{Q}$ group, and G' is a subgroup of G, then the short exact sequence

$$0 \to (P \cap G') \to G' \to \pi(G') \to 0,$$

where $\pi : G \to Q$ is a projection map, demonstrates that G' is also a $\mathcal{P}$-by-$\mathcal{Q}$ group. Thus, for instance, the properties of being metabelian, metacyclic, polycyclic, supersolvable, solvable, or nilpotent, are hereditary. As a consequence, virtually nilpotent is the same as nilpotent-by-finite, etc.

We saw for hereditary properties $\mathcal{P}$ that "$\mathcal{P}$-by-finite" was the same concept as "virtually $\mathcal{P}$". It is natural to ask whether the same is true for "finite-by-$\mathcal{P}$". The answer is no; for instance, one can extend the an infinite vector space V over a finite field F by F (using some non-degenerate bilinear anti-symmetric form $\omega : V \times V \to F$, and defining $(v, f)(w, g) = (v+w, f+g+\omega(v, w))$ for $v, w \in V$ and $f, g \in F$) to create a nilpotent group which is finite-by-abelian, but not virtually abelian. Conversely, the semidirect product $\mathbf{Z} \rtimes \mathbf{Z}/2\mathbf{Z}$ (where $\mathbf{Z}/2\mathbf{Z}$ acts on $\mathbf{Z}$ by reflection) is virtually abelian, but not finite-by-abelian. On the other hand, for hereditary $\mathcal{P}$, a finite-by-$\mathcal{P}$ group is virtually (central finite)-by-$\mathcal{P}$. This is because if G is an extension of a $\mathcal{P}$ group P by a finite group F, then G acts by conjugation on the finite group F; the stabiliser G' of this action is then a finite index subgroup, whose intersection of F is then central in G'. The projection of G' onto P is also a $\mathcal{P}$ group by the hereditary nature of $\mathcal{P}$, and the claim follows.

[8] Proof: G acts on the finite quotient space G/H by left multiplication, hence the *stabiliser* of G/H has finite index in G. But this stabliser is also normal in G and contained in H.

Remark 2.4.4. There is a variant of the above result which is also useful. Suppose one has an H-by-K group G in which the action of K on H is virtually trivial (i.e., there are only a finite number of distinct automorphisms of H induced by K). Then G is virtually a central H'-by-K' group for some finite index subgroups H', K' of H, K.

One can phrase various results in group theory in a succinct form using this notation. For instance, one of the basic facts about (discrete) *amenable* groups is that amenable-by-amenable groups are amenable; see [**Ta2010**, §2.8]. As another example, the main result of a well-known paper of Larsen and Pink [**LaPi2011**] is a classification of finite linear groups over a field of characteristic p, namely that such groups are virtually (p-group by abelian) by (semisimple of Lie type), where one has bounds on the index of the "virtually" and on the type of the semisimple group.

2.5. A proof of Gromov's theorem

A celebrated theorem of Gromov [**Gr1981**] reads:

Theorem 2.5.1 (Gromov's theorem). *Every finitely generated group of polynomial growth is virtually nilpotent.*

The original proof of Gromov's theorem was quite non-elementary, using an infinitary limit and exploiting the work surrounding the solution to *Hilbert's fifth problem.* More recently, Kleiner [**Kl2010**] provided a proof which was more elementary (based in large part on an earlier paper of Colding and Minicozzi [**CoMi1997**]), though still not entirely so, relying in part on (a weak form of the) Tits alternative [**Ti1972**] and also on an ultrafilter argument of Korevaar and Schoen [**KoSc1997**] and Mok [**Mo1995**]. Kleiner's argument is discussed further in [**Ta2009**, §1.2].

Recently, Yehuda Shalom and I [**ShTa2010**] established a quantitative version of Gromov's theorem by making every component of Kleiner's argument finitary. Technically, this provides a fully elementary proof of Gromov's theorem (we do use one infinitary limit to simplify the argument a little bit, but this is not truly necessary); however, because we were trying to quantify as much of the result as possible, the argument became quite lengthy.

In this note I want to record a short version of the argument of Yehuda and myself which is not quantitative, but gives a self-contained and largely elementary proof of Gromov's theorem. The argument is not too far from the Kleiner argument, but incorporates a number of simplifications. In a number of places, there was a choice to take between a short argument that was "inefficient" in the sense that it did not lead to a good quantitative bound, and a lengthier argument which led to better quantitative bounds. I have opted for the former in all such cases.

2.5.1. Overview of argument. The argument requires four separate ingredients. The first is the existence of non-trivial Lipschitz harmonic functions $f : G \to \mathbf{R}$:

Theorem 2.5.2 (Existence of non-trivial Lipschitz harmonic functions). *Let G be an infinite group generated by a finite symmetric set S. Then there exists a non-constant function $f : G \to \mathbf{R}$ which is harmonic in the sense that*

$$f(x) = \frac{1}{|S|} \sum_{s \in S} f(xs)$$

for all $x \in G$, and Lipschitz in the sense that

$$|f(x) - f(sx)| \leq C$$

for all $x \in G$ and $s \in S$, and some $C < \infty$.

The second is that there are not *too* many such harmonic functions:

Theorem 2.5.3 (Kleiner's theorem). *Let G be a group of polynomial growth generated by a finite symmetric set S of generators. Then the vector space V of Lipschitz harmonic functions is finite-dimensional.*

The third ingredient is that Gromov's theorem is true in the compact linear group case:

Theorem 2.5.4 (Gromov's theorem in the compact linear case). *Let G be a finitely generated subgroup of a compact linear group $H \subset GL_n(\mathbf{C})$ of polynomial growth. Then G is virtually abelian.*

The final ingredient is that Gromov's theorem is inductively true once one can locate an infinite cyclic quotient:

Theorem 2.5.5 (Gromov's theorem with an cyclic quotient). *Let G be a finitely generated group which has polynomial growth of exponent at most d (i.e., the volume of a ball $B_S(r)$ grows like $O(r^d)$ for any fixed set of generators S). Suppose inductively that Gromov's theorem is already known for groups of polynomial growth of exponent at most $d-1$, and suppose that G contains a finite index subgroup G' which can be mapped homomorphically onto an infinite cyclic group. Then G is virtually nilpotent.*

We prove these four facts in later sections. For now, let us see how they combine to establish Gromov's theorem in full generality.

We assume that G has polynomial growth of order d, and assume inductively that Gromov's theorem has already been established for growth of order $d-1$ or less. We fix a symmetric set S of generators.

We may assume that G is infinite otherwise we are already done. So by Theorem 2.5.2, the space V of (complex) Lipschitz harmonic functions consists of more than just the constants $\mathbf{R}$. In particular, setting $W := V/\mathbf{C}$, we have a non-trivial short exact sequence

$$0 \to \mathbf{C} \to V \to W \to 0.$$

The left translation action of G preserves the space of Lipschitz harmonic functions, and is thus an action of G on V. Since G preserves constants, it is also an action of G on W. Now, on W, the homogeneous Lipschitz norm is a genuine norm, and is preserved by the action of G. Since all norms are equivalent on a finite-dimensional space, we can place an arbitrary Euclidean structure on W and conclude that this structure is preserved up to constants by G. So, the image of the action of G on W is precompact, and thus its closure is a compact linear group. By Theorem 2.5.4, this image is virtually abelian. If it is infinite, then we thus see that a finite index subgroup of G has an infinite abelian image, and thus has a surjective homomorphism onto the integers, and we are done by Theorem 2.5.5. So we may assume that this image is finite; thus there is a finite index subgroup G' of G that is trivial on W. The action of G' on V then collapses to the form $gf = f + \lambda_g(f)$ for some linear functional $\lambda_g \in V^*$ (in fact λ_g annihilates 1 and so comes from W^*). Note that λ is then an additive representation of G. If the image of this representation is infinite, then we are again done by Theorem 2.5.5, so we may assume that it is finite; thus there is a finite index subgroup G'' of G' that is trivial on V. In other words, all Lipschitz harmonic functions are G''-invariant, and thus take only finitely many values. But looking at the maximum such value and using harmonicity (i.e., using the *maximum principle*) we conclude that all Lipschitz harmonic functions are constant, a contradiction.

2.5.2. Building a Lipschitz harmonic function. Now we prove Theorem 2.5.2. We introduce the function

$$\mu := \frac{1}{|S|} \sum_{s \in S} \delta_s$$

where δ_s is the Kronecker delta function. The property of a function $f : G \to \mathbf{C}$ being harmonic is then simply that $f * \mu = f$, using the discrete convolution structure on the group.

To build such a function, we consider the functions

$$f_n := \frac{1}{n} \sum_{m=1}^{n} \mu^{(m)}$$

where $\mu^{(m)} := \mu * \ldots * \mu$ is the convolution of m copies of μ. This sequence of functions is "asymptotically harmonic" in the sense that

$$\|f_n\|_{\ell^1(G)} = 1$$

but

$$\|f_n - f_n * \mu\|_{\ell^1(G)} = O(1/n)$$

(we allow implied constants to depend on S).

There are now two cases. The first case is the **non-amenable case**, when we have

$$\|f_n - f_n * \delta_s\|_{\ell^1(G)} > \varepsilon > 0$$

for some $s \in S$, some $\varepsilon > 0$, and infinitely many n; informally, this means that the averaged iterated convolutions f_n are not getting smoother as $n \to \infty$. By the duality of $\ell^1(G)$ and $\ell^\infty(G)$, we see that for each such n we can find H_n with $\|H_n\|_{\ell^\infty(G)} = 1$ such that

$$|H_n * f_n(\mathrm{id}) - H_n * f_n(s)| > \varepsilon.$$

But Young's inequality, $H_n * f_n$ has $\ell^\infty(G)$ norm of at most 1, and

$$\|H_n * f_n - H_n * f_n * \mu\|_{L^\infty(G)} = O(1/n).$$

Using the sequential Banach-Alaoglu theorem we may take a subsequence limit and obtain a non-trivial bounded harmonic function. Since bounded functions are automatically Lipschitz, and the claim follows.

The second case is the **amenable case**, when we have

$$\|f_n - f_n * \delta_s\|_{\ell^1(G)} \to 0$$

as $n \to \infty$ for each $s \in S$. Setting $F_n := f_n^{1/2}$, one soon verifies that

$$\|F_n\|_{\ell^2(G)} = 1$$

and

$$\|F_n - F_n * \delta_s\|_{\ell^2(G)} = o(1).$$

In particular,

$$\|F_n - F_n * \mu\|_{\ell^2(G)} = o(1).$$

From this and the spectral theorem, we see that the positive-definite Laplacian operator $\Delta : \ell^2(G) \to \ell^2(G)$ defined by the formula

$$\Delta F := F - F * \mu$$

has non-trivial spectrum at the origin. On the other hand, as G is infinite, there are no non-trivial harmonic functions in $\ell^2(G)$ (as can be seen from the maximum principle), and so the spectrum at the origin is not coming from a zero eigenfunction. From this and the spectral theorem (taking spectral projections to $[0, \varepsilon]$ for small ε), one can find a sequence $G_n \in \ell^2(G)$ of functions such that

$$\sum_{g \in G} G_n(g) \Delta G_n(g) = 1$$

but

$$\|\Delta G_n\|_{\ell^2(G)} \to 0$$

as $n \to \infty$.

A summation by parts gives the Dirichlet energy identity

$$\sum_{g\in G} G_n(g)\Delta G_n(g) = \frac{1}{2|S|}\sum_{s\in S} \|G_n - G_n * \delta_s\|_{\ell^2(G)}^2$$

and thus

$$\|G_n - G_n * \delta_s\|_{\ell^2(G)} = O(1),$$

and also there exists $s_0 \in S$ such that

$$\|G_n - G_n * \delta_{s_0}\|_{\ell^2(G)} \gg 1$$

for infinitely many n. By the self-duality of $\ell^2(G)$, we may thus find a sequence $H_n \in \ell^2(G)$ with $\|H_n\|_{\ell^2(G)} = 1$ such that

$$|H_n * G_n(\mathrm{id}) - H_n * G_n(s_0)| \gg 1$$

for infinitely many n. From Young's inequality we also see that

$$\|H_n * G_n - H_n * G_n * \delta_s\|_{\ell^\infty(G)} = O(1)$$

(so $H_n * G_n$ is uniformly Lipschitz) and

$$\|\Delta(H_n * G_n)\|_{\ell^\infty(G)} \to 0$$

as $n \to \infty$, thus $H_n * G_n$ is asymptotically harmonic. Using the Arzelá-Ascoli theorem to take another subsequence limit (after first subtracting a constant to normalise $H_n * G_n$ to be zero at the identity, so that $H_n * G_n$ becomes locally bounded by the uniform Lipschitz property) we obtain the required non-trivial Lipschitz harmonic function.

Remark 2.5.6. In the case of groups of polynomial growth, one can verify that one is always in the "amenable" case. In the non-amenable case, the theory of Poisson boundaries gives a plentiful supply of *bounded* Lipschitz harmonic functions (in fact, there is an infinite-dimensional space of such).

2.5.3. Kleiner's theorem. We now prove Theorem 2.5.3. Our proof will basically repeat those in Kleiner's original paper [**Kl2010**]. For simplicity, let us assume a stronger condition than polynomial growth, namely *bounded doubling*

$$|B_S(2R)| \leq C|B_S(R)|$$

for some fixed constant C and all $R > 0$. In general, polynomial growth does not obviously imply bounded doubling at all scales, but there is a simple pigeonhole argument that gives bounded doubling on *most* scales, and this turns out to be enough to run the argument below. But in order not to deal with the (minor) technicalities arising from exceptional scales in which bounded doubling fails, I will assume bounded doubling at all scales. The full proof in the general case can, of course, be found in Kleiner's paper (which in turn was based upon an earlier argument of Colding and Minicozzi [**CoMi1997**]).

Let $\varepsilon > 0$ be a small parameter. The key lemma is

Lemma 2.5.7 (Elliptic regularity). *Cover $B_S(4R)$ by balls B of radius εR. Suppose that a harmonic function $f : G \to \mathbf{R}$ has mean zero on every such ball. Then one has*

$$\|f\|_{\ell^2(B_S(R))} \ll \varepsilon \|f\|_{\ell^2(B_S(4R))}.$$

Let's see how this lemma establishes the theorem. Consider some Lipschitz harmonic functions $u_1, \ldots, u_D$, which we normalise to all vanish at the identity. Let V be the space spanned by $u_1, \ldots, u_D$. For each R, the $L^2(B_S(R))$ inner product gives a quadratic form Q_R on V. Using this quadratic form, we can build a Gram matrix determinant

$$\det(Q_R(u_i, u_j))_{1 \leq i,j \leq D}.$$

From the Lipschitz nature of the harmonic functions, we have a bound of the form

$$\det(Q_R(u_i, u_j))_{1 \leq i,j \leq D} \ll R^D \tag{2.1}$$

as $R \to \infty$. On the other hand, we also have the monotonicity property

$$\det(Q_R(u_i, u_j))_{1 \leq i,j \leq D} \leq \det(Q_{4R}(u_i, u_j))_{1 \leq i,j \leq D}.$$

Now by bounded doubling, we can cover $B_S(4R)$ by $O_\varepsilon(1)$ balls of radius εR. By Lemma 2.5.7, the space of functions in V which have mean zero on each such ball is such that Q_R is bounded (as a quadratic form) by $O(\varepsilon)$ times Q_{4R} on this space. Furthermore, by linear algebra, this space has codimension $O_\varepsilon(1)$ in V. Using this, we obtain the improved bound

$$\det(Q_R(u_i, u_j))_{1 \leq i,j \leq D} \leq O(\varepsilon)^{D - O_\varepsilon(1)} \det(Q_{4R}(u_i, u_j))_{1 \leq i,j \leq D}.$$

For ε small enough and D large enough, the rate of growth $O(\varepsilon)^{D-O_\varepsilon(1)}$ is strictly less than 4^{-D}. Iterating this estimate by doubling R off to infinity, and comparing against (2.1), we conclude in the limit that

$$\det(Q_R(u_i, u_j))_{1 \leq i,j \leq D} = 0$$

for all R, and so $u_1, \ldots, u_D$ cannot be linearly independent. This implies that the space of Lipschitz harmonic functions has dimension at most $D+1$, and the claim follows.

It remains to prove the lemma. Fix the harmonic function f.

There are two basic ingredients here. The first is the reverse Poincaré inequality[9]

$$\sum_{x \in B_S(2R)} |\nabla f(x)|^2 \ll R^{-2} \sum_{x \in B(x_0, 4R)} |f(x)|^2$$

[9] This inequality is in the general spirit of the philosophy that functions that are harmonic on a ball, should be smooth on that ball.

where

$$|\nabla f(x)|^2 := \sum_{s\in S} |f(x) - f(xs)|^2.$$

This claim (which heavily exploits the harmonicity of f) is proven by writing $|f|^2$ as $f(f*\mu)$, multiplying by a suitable cutoff function adapted to $B(x_0, 2r)$ and equalling one on $B(x_0, r)$, and summing by parts; we omit the standard details.

The second claim is the Poincaré inequality

$$\sum_{x,y\in B(x_0,r)} |f(x) - f(y)|^2 \ll r^2 |B_S(r)| \sum_{x\in B(x_0,3r)} |\nabla f(x)|^2,$$

which does not require harmonicity. To prove this claim, observe that the left-hand side can be bounded by

$$\sum_{g\in B_S(2r)} \sum_{x\in B(x_0,r)} |f(x) - f(xg)|^2.$$

But by expanding each $g \in B_S(2r)$ as a word of length most $2r$ and using the triangle inequality in ℓ^2 and Cauchy-Schwarz, we have

$$\sum_{x\in B(x_0,r)} |f(x) - f(xg)|^2 \ll r^2 \sum_{x\in B(x_0,3r)} |\nabla f(x)|^2$$

and the claim follows.

If f has mean zero on $B(x_0, r)$, the Poincaré inequality implies that

$$\sum_{x\in B(x_0,r)} |f(x)|^2 \ll r^2 \sum_{x\in B(x_0,3r)} |\nabla f(x)|^2. \tag{2.2}$$

To prove the lemma, we first use bounded doubling to refine the family of balls $B = B(x_i, \varepsilon R)$ so that the triples $3B = B(x_i, 3\varepsilon R)$ have bounded overlap. Applying (2.2) for each such ball and summing we obtain the claim.

2.5.4. The compact linear case. Now we prove Theorem 2.5.4. It is a classical fact that all compact linear groups H are isomorphic to a subgroup[10] of a unitary group $U(n)$; indeed, if one takes the standard inner product on $\mathbf{C}^n$ and averages it by the Haar measure of H, one obtains an inner product which is H-invariant, and so H can be embedded inside the unitary group associated to this group. Thus it suffices to prove the claim when $H = U(n)$.

A key observation is that if two unitary elements g, h are close to the identity, then their commutator $[g, h] = ghg^{-1}h^{-1}$ is even closer to the

[10] Indeed, thanks to a theorem of Cartan, H is isomorphic to a *Lie* subgroup of $U(n)$, i.e., an analytic submanifold of $U(n)$ that is also a subgroup; but we will not need this fact here.

identity. Indeed, since multiplication on the left or right by unitary elements does not affect the operator norm, we have

$$\begin{aligned}\|[g,h]-1\|_{op} &= \|gh-hg\|_{op}\\ &= \|(g-1)(h-1)-(h-1)(g-1)\|_{op}\end{aligned}$$

and so by the triangle inequality

$$\|[g,h]-1\|_{op} \leq 2\|g-1\|_{op}\|h-1\|_{op}. \tag{2.3}$$

We now need to exploit (2.3) to prove Theorem 2.5.4. As a warm-up, we first prove the following slightly easier classical result:

Theorem 2.5.8 (Jordan's theorem)**.** *Let G be a finite subgroup of $U(n)$. Then G contains an abelian subgroup of index $O_n(1)$ (i.e., at most C_n, where C_n depends only on n).*

And indeed, the proof of the two results are very similar. Let us first prove Jordan's theorem. We do this by induction on n, the case $n=1$ being trivial. Suppose first that G contains a *central element* g (i.e., an element that commutes with all elements of G) which is not a multiple of the identity. Then, by definition, G is contained in the *centraliser* $Z(g) := \{a \in U(n) : ag = ga\}$ of g, which by the spectral theorem is isomorphic to a product $U(n_1) \times \ldots \times U(n_k)$ of smaller unitary groups. Projecting G to each of these factor groups and applying the induction hypothesis, we obtain the claim.

Thus we may assume that G contains no central elements other than multiples of the identity. Now pick a small $\varepsilon > 0$ (one could take $\varepsilon = 1/10$ in fact) and consider the subgroup G' of G generated by those elements of G that are within ε of the identity (in the operator norm). By considering a maximal ε-net of G we see that G' has index at most $O_{n,\varepsilon}(1)$ in G. By arguing as before, we may assume that G' has no central elements other than multiples of the identity.

If G' consists only of multiples of the identity, then we are done. If not, take an element g of G' that is not a multiple of the identity, and which is as close as possible to the identity (here is where we use that G is finite). By (2.3), we see that if ε is sufficiently small depending on n, and if h is one of the generators of G', then $[g,h]$ lies in G' and is closer to the identity than g, and is thus a multiple of the identity. On the other hand, $[g,h]$ has determinant 1. Given that it is so close to the identity, it must therefore be the identity (if ε is small enough). In other words, g is central in G', and is thus a multiple of the identity. But this contradicts the hypothesis that there are no central elements other than multiples of the identity, and we are done.

The proof of Theorem 2.5.4 is analogous. Again, we pick a small $\varepsilon > 0$, and define G' as before. If G' has a central element that is not a multiple of the identity, then we can again argue via induction, so suppose that there are no such elements.

Being finitely generated, it is not difficult to show that G' can be generated by a finite set S of generators within distance ε of the identity. Now pick an element $h_1 \in S$ which is not a multiple of the identity, and is at a distance δ_1 from the identity for some $0 < \delta_1 \leq \varepsilon$. We look at all the commutators $[g, h_1]$ where $g \in S$. By (2.3), they are all at distance $O(\varepsilon\delta_1)$ from the identity, and have determinant 1. If they are all constant multiples of the identity, then by arguing as before we see that h_1 is central in G', a contradiction, so we can find an element $h_2 := [g_1, h_1]$ for some $g_1 \in S$ which is a distance $\delta_2 = O_n(\varepsilon\delta_1)$ from the origin and is not a multiple of the identity. Continuing this, we can construct $h_3 := [g_2, h_2]$, etc., where each h_n is a distance $0 < \delta_n = O(\varepsilon\delta_{n-1})$ from the identity, and is a commutator of h_{n-1} with a generator.

Because of the lacunary nature of the distances of $h_1, h_2, h_3, \ldots$, we easily see that the words $h_1^{i_1} \ldots h_m^{i_m}$ with $0 \leq i_1, \ldots, i_m \leq c\varepsilon^{-1}$ are distinct for some small $c > 0$. On the other hand, all of these words lie in the ball of radius $O(m\varepsilon^{-1}2^m)$ generated by S. This contradicts the polynomial growth hypothesis for ε taken small enough and m large enough.

Remark 2.5.9. Theorem 2.5.4 can be deduced as a corollary of Gromov's theorem, though we do not do so here as this would be circular. Indeed, it is not hard to see that the image of a torsion-free nilpotent group in a unitary group must be abelian.

2.5.5. The case of an infinite abelian quotient. Now we prove Theorem 2.5.5 (which was already observed in Gromov's original paper [**Gr1981**], and also closely related to earlier work of Milnor [**Mi1968**] and of Wolf [**Wo1968**]).

Since G is finitely generated and has polynomial growth of order d, the finite index subgroup G' is also finitely generated of growth d. By hypothesis, there is a non-trivial homomorphism $\phi : G' \to \mathbf{Z}$. Using the Euclidean algorithm, one can move the generators $e_1, \ldots, e_m$ of G' around so that all but one of them, say $e_1, \ldots, e_{m-1}$, lie in the kernel $\ker(\phi)$ of ϕ; we thus see that this kernel must then be generated by $e_1, \ldots, e_{m-1}$ and their conjugates $e_m^k e_i e_m^{-k}$ by powers of e_m.

Let S_k be the set of $e_m^{k'} e_i e_m^{-k'}$ for $1 \leq i \leq m-1$ and $|k'| \leq k$, and let B_k be the words of length at most k generated by elements of S_k. Observe that if at least one of the elements in S_{k+1} is not contained in $B_k \cdot B_k^{-1}$, then B_{k+1} is at least twice as big as B_k. Because of polynomial growth, this implies that $S_{k+1} \subset B_k \cdot B_k^{-1}$ for some $k \geq 1$, which implies that $\ker(\phi)$ is generated by S_k.

Observe that the ball of radius R generated by S_k, e_m is at least $R/2$ times as large as the ball of radius $R/2$ generated by S_k. Since G' has growth d, we conclude that $\ker(\phi)$ has growth at most $d-1$, and is thus virtually nilpotent by hypothesis.

We have just seen that the kernel $\ker(\phi)$ contains a nilpotent subgroup N of some finite index M; it is thus finitely generated. We may take N to be a normal subgroup of $\ker(\phi)$. From Lagrange's theorem, we see that the group N' generated by the powers g^M with $g \in \ker(\phi)$ is then contained in N and is therefore nilpotent. N' is clearly a *characteristic subgroup* of $\ker(\phi)$ (i.e., preserved under all outer automorphisms), and is thus normal in N. The group N/N' is nilpotent and finitely generated with every element being of order M, and is thus finite; thus N' is finite index in $\ker(\phi)$. Since it is characteristic, it is, in particular, invariant under conjugation by e_m. If one lets $G'' = \mathbf{Z} \ltimes_{e_m} N'$ be the group generated by N' and e_m, we see that G'' is a finite index subgroup[11] of G. In particular, it has polynomial growth.

To conclude, we need to show that G'' is virtually nilpotent. It will suffice to show that the conjugation action of e_m^a on N' acts unipotently on N' for some finite $a > 0$. We can induct on the step of the nilpotent group N', assuming that the claim has already been proven for the quotient group $N'/Z(N')$ (where $Z(N')$ is the centre of N'), which has one lower step on N'. Thus it suffices to prove unipotence on just the centre $Z(N')$, which is a finitely generated abelian group and thus isomorphic to some $\mathbf{Z}^d \times H$ for some finite group H. The torsion group H must be preserved by this action. By *Lagrange's theorem*, the action on H becomes trivial after raising e_m to a suitable power, so we only need to consider the action on $\mathbf{Z}^d$. In this case the conjugation action can be viewed as a matrix A in $SL_d(\mathbf{Z})$. Because G'' has polynomial growth, the powers A^n of A for $n \in \mathbf{Z}$ cannot grow exponentially; in other words, all the eigenvalues of A have unit magnitude. On the other hand, these eigenvalues consist of Galois conjugacy classes of algebraic integers. But it is a classical result of Kronecker that the only algebraic integers α whose Galois conjugacy classes all have unit magnitude are the roots of unity[12]. We conclude that all the eigenvalues of A are roots of unity, i.e., some power of A is unipotent, and the claim follows.

[11] Note that as e_m is not annihilated by ϕ, it will have infinite torsion even after quotienting out by N'.

[12] Proof: The algebraic integers α^n for natural number n have bounded degree and all Galois conjugates bounded, so the minimal polynomials have bounded integer coefficients and must thus repeat themselves after finitely many n. Note also that the group $\mathbf{Z} \ltimes \mathbf{Z}[\alpha]$ is of polynomial growth, so one can in fact view Kronecker's theorem as a very special case of Gromov's theorem, making it unsurprising that it is needed at some point in the proof of that theorem.