PIMS 2025 – Random walks on polynomial growth groups

Lecturer: Professor Tianyi Zheng

Notes by: Andrew Lin

June 2025

1 June 17, 2025

Today, we'll discuss some classical facts about volume growth on groups and how that relates to random walks. We'll then understand how to get estimates for simple random walk and then processes with jumps, and down the line we'll see connections to heat kernel estimates.

Example 1

Our model is as follows: let G be a countable group (often finitely generate) and let μ be a probability measure on G which will be our random walk step distribution. We will sample a sequence of iid random variables $X_1, X_2, X_3, \cdots$ from μ, and we multiply them together (so that at time 0 we're at the identity element, and at W_n we're at $X_1 X_2 \cdots X_n$). This is called the **μ-random walk on G**.

The simplest model would be μ to be iid on $\{\pm 1\}$ and $G = \mathbb{Z}$, recovering the original simple random walk. Such walks were first studied by Kesten around 1959, and there are connections to what is called "analytic group theory" by Gelander. For example, Kesten proved that G is **amenable** if and only if every nondegenerate symmetric μ-random walk has $\mu^{(2n)}(\text{id})$ decaying sub-exponentially (that is, slower than $e^{-\lambda n}$ for any λ). Here "nondegenerate" means that the support of μ generates G and "symmetric" means that $\mu(g) = \mu(G^{-1})$. The point is that this was the first example where we **use random walks to characterize properties of a group**, a similar thing has been done for the Liouville property.

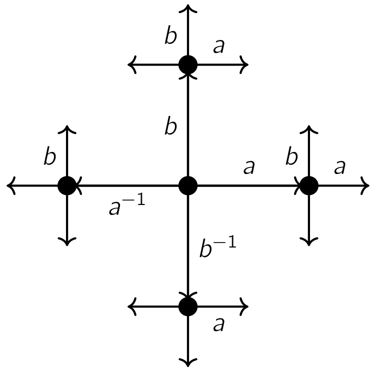
There are also connections of random walks to the geometry of groups (asking questions like volume growth, the shape of the expansion of the set via isoperimetry, and so on). And turning to things that are less well-understood, we may also study these random walks via their relations to dynamics of group actions. For example, we've seen the fractal structure of the Sierpinski gasket; it turns out these kinds of sets come from certain iterated function systems or zero sets of rational functions, and for these associated "self-similar groups" we can ask about their analytic properties.

Example 2

We'll explain further now what we mean by polynomial growth of groups. Suppose that G is equipped with some finite generating set – for example we can have the free group on two generators $F_2 = \langle a, b \rangle$ (with no relations or cancellations) or $\mathbb{Z}^2 = \langle a, b \,|\, [a, b] = 1 \rangle$.

For any such finite generating set we can form the **Cayley graph** (G, S), which is where we connect two vertices if we can go from one to another by a generator. Specifically, the vertices are given by the elements of G, and we have directed edges $\{(g, gs) : g \in G, s \in S\}$. Such a graph is always left-invariant (since if we multiply on the left by any group element, (hg, hgs) is an edge and so was (g, gs)).

The Cayley graph for $\mathbb{Z}^2$ just looks like the usual integer lattice (with edges pointing up and right), while the free group has a more complicated graph because there are no further relations – it's a regular tree (with a pointing to the right, its inverse pointing to the left, b pointing to up, and its inverse pointing down):



In the free group, we can understand the polynomial growth function quite easily: letting $d = d_S$ be the graph distance in the Cayley graph (G, S), we can calculate the **volume growth function**

$$V_{G,S}(r) = \#\{g : d(\text{id}, g) \leq r\}.$$

For the free group, we have $V_{G,S}(r) = 4 \cdot 3^{r-1}$, and on the integer lattice we have $V_{G,S}(r) \sim r^2$ (so exponential versus polynomial behavior in the two cases).

Example 3

The **Heisenberg group** is generated by the following two 3×3 matrices:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Notice that the commutator $[a, b] = aba^{-1}b^{-1}$ satisfies

$$[X^m, Y^n] = \begin{bmatrix} 1 & m & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & n \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -m & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -n \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & mn \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This has some implications for volume growth, since if we take the group generated by X and Y and also define $Z = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, all elements can always be written in the form $X^m Y^n Z^\ell$ for some integers m, n, ℓ. The commutator calculation then tells us that

$$Z^{n^2} = X^n Y^n X^{-n} Y^{-n},$$

so the scaling we can get is up to exponents of size r, r, r^2 in m, n, ℓ, leading to the following:

Proposition 4

For the Heisenberg group, the volume growth satisfies $V_{G,S}(r) \sim r^4$.

We thus started with two elements off the diagonal, and then taking commutators to get things further off the diagonal:

Definition 5

For a group G, we can take the lower central series

$$G_1 = G, \quad G_2 = [G, G], \quad G_3 = [G_2, G], \quad \cdots$$

and we say that G is **nilpotent** of class c if $G_{c+1} = \{\text{id}\}$.

For the Heisenberg group, we see that $[X, Z] = [Y, Z] = \text{id}$, and therefore the Heisenberg group is nilpotent of class 2. (The nilpotent groups are exactly those that can be embedded in some sufficiently large upper triangular matrix group.) This nilpotent structure is important, and the classical theorem in this area is the following:

Theorem 6 (Gromov polynomial growth theorem)

Let $G = \langle S \rangle$ be a finitely generated group. Suppose there are some constants $C > 0, d > 0$ such that $V_{G,S}(r) \leq Cr^d$. Then G has a finite index subgroup H such that H is nilpotent.

So these upper triangular matrices are essentially the only way to get polynomial growth, and we'll see how the structure of these nilpotent groups comes up.

Definition 7

Let $\mathcal{M}_k$ be the space of **marked groups** $(G, (S_1, \cdots, S_k))$, where S_i are an ordered list of generators. This space is equipped with the Cayley topology of local convergence, meaning that (G, S) and (H, T) are close if there is a large radius such that the metric balls around the identity coincide

This next result is basically telling us that polynomial growth is a local property, which might be surprising:

Theorem 8 (Breuillard–Green–Tao)

For any $d > 0$, there exists some radius $n_0 \in \mathbb{N}$ depending only on d, such that if $V_{G,S}(r) \leq r^d|S|$ for some $r \geq n_0$, then G is virtually nilpotent (so we can pass to a finite index nilpotent group).

So we can consider only the marked groups of polynomial growth, and what we're saying is that this set is open in $\mathcal{M}_k$. In contrast, it's not true at all that we can see what our group does with a finite ball, in the following sense:

Definition 9

A group G has **intermediate growth** if $V_{G,S}(r)$ is subexponential but not polynomial in r (for example with growth function $e^{\sqrt{r}}$).

Theorem 10 (Grigorchuk)

There exists a continuous map from $\{0, 1, 2\}^{\mathbb{N}}$ (with the product topology) to 4-generated groups $\mathcal{M}_4$ (meaning that if we have two strings that coincide with a long prefix, then our resulting groups will have large balls where they coincide) with the following properties (say that a string ω maps to G_ω):

1. If ω is eventually periodic with all three letters $0, 1, 2$ appearing infinitely often, then $V_{G_\omega}(r) \lesssim e^{r^{\alpha_\omega}}$ for some constant $\alpha_\omega < 1$.
2. If ω is eventually constant, then V_{G_ω} grows exponentially.

This is in stark contrast to the previous result – we may have large balls which look very similar in the Cayley graphs, but then the growth is very different asymptotically. But assuming polynomial growth, we in fact have much more highly constrained conditions.

Fact 11

We next want to turn to limiting behavior of random walks. One popular topic in this which we won't go into much is **boundary theory** – this has some ergodic theory flavor, where once we have a μ-random walk (W_n), we can ask about the **tail σ-field** $\mathcal{T}$ associated with the process or the **invariant σ-field**. The idea is that we will end up in some boundary set after doing random walk, and so we get some connections to bounded harmonic functions and martingale solutions.

Thinking about the simple random walk as a simple example, there isn't very much we can say that's interesting about the boundary theory. But we do have theorems telling us things like $\frac{W_n}{n} \to 0$ almost surely (law of large numbers), that $\frac{W_n}{\sqrt{n}} \to N(0, 1)$ (central limit theroem), or $\left(\frac{W_{\lfloor nt \rfloor}}{\sqrt{n}}\right) \to B_t$ (invariance principle, or equivalently a functional central limit theorem). The idea is that we're considering $\mathbb{Z}$ as a subset of $\mathbb{R}$ and we have a valuation operator $\delta_t : \mathbb{R} \to \mathbb{R}$ so that we can make sense of $\frac{W_n}{\sqrt{n}}$ and "scale down space" accordingly. In particular, we might ask how we would construct scaling limits when we don't live in something like $\mathbb{R}$.

There are a few ways to think about such situations and "scaling space." One thing we can do is take a metric point of view, where we start with some metric space (G, d_S) and then rescale the metric to $(G, \frac{1}{n} d_S)$; it may seem a little sketchy to take $n \to \infty$, but there is in fact a way to do so using **asymptotic cones** in a well-posed way. Some difficulties will automatically arise – for example the limit may not necessarily exist, and that limiting space may not be locally compact in general. Furthermore, we still need a way of thinking about the random walk on this limiting space, and there aren't quite ways of doing this outside the polynomial growth regime.

But we can also take a more algebraic approach. First, we ask what spaces can actually admit such scalings: specifically, we want to know what locally compact groups admit a contractive automorphism, meaning that we can make sense of "scaling down by a factor" like in our simple random walk. Thus we want to first construct some continuous $\tau : G \to G$ with the property that for any $g \in G$, repeated iterations of τ will have us converge to the identity. It turns out that this is already a lot to ask for – there's already a very restrictive classification of what is possible:

Theorem 12 (Siebert)
If G admits a continuous contractive automorphism, then $G = G_e \times D$ for D totally disconnected and G_e connected, where G_e is a simply connected nilpotent Lie group with a positive gradation, meaning that the Lie algebra decomposes as $\mathfrak{g} = \bigoplus \mathfrak{m}_s$ with $[\mathfrak{m}_s, \mathfrak{m}_t] \subseteq \mathfrak{m}_{s+t}$ for all s, t. Defining τ_r via $\tau_r(x) = \sum_s r^s \pi_s(x)$, where π_s is the projection from $\mathfrak{g}$ to $\mathfrak{m}_s$, we then have a contraction if $r < 1$.

The Heisenberg group we've seen is an example of such a group. And a result of Glöcker-Willis characterizes the totally disconnected part in a complicated way, but it's quite well-understood.

The point is that we end up looking at scaling limits on polynomial growth groups, because somehow this is the only reasonable place we can end up stating precise limit theorems. Otherwise, we need to invent something new to even make sense of the process!

2 June 19, 2025

Today, we'll do some basic estimates of random walk on polynomial growth groups. We'll start with on-diagonal estimates:

Example 13
Let Γ be a finitely generated group generated by some set $\langle S \rangle$, and let μ be a finitely supported, nondegenerate, symmetric probability measure on Γ (for example uniform on S and its inverses). We can associate a Dirichlet form to this process, since we're symmetric with respect to the counting measure π on Γ. This yields a Markov operator $P_\mu : \ell^2(\Gamma) \to \ell^2(\Gamma)$, defined by

$$P_\mu f(x) = \sum_{s \in \Gamma} f(xs)\mu(s).$$

The associated Dirichlet form will be (we won't have to worry about the domain of the form and so on because everything is discrete and the semigroup is just convolutions of μ)

$$\begin{aligned}\mathcal{E}(f, f) &= \langle (I - P_\mu)f, f \rangle \\ &= \frac{1}{2} \sum_{s, x \in \Gamma} \left(f(x) - f(xs)\right)^2 \mu(s)\end{aligned}$$

where we've "symmetrized" the expression after expanding out. Our first task is to understand how the return probability $\mu^{(2n)}(\mathrm{id})$ decays, and since we're looking at symmetric random walks this is exactly

$$\mu^{(2n)}(\mathrm{id}) = \sum_{g \in \Gamma} \mu^{(n)}(g)\mu^{(n)}(g^{-1})$$

by doing casework on the nth step, and then this is exactly (by symmetry) $||\mu^{(n)}||_2^2$. So in particular the norm $||P_\mu^{2n}||_{1 \to \infty}$ is governed by Nash inequalities:

Fact 14
We wish to find an inequality of the form

$$||f||_2^2 \leq \mathcal{N}\left(\frac{||f||_1^2}{||f||_2^2}\right)\mathcal{E}(f,f),$$

where $\mathcal{N} : (0,\infty) \to (0,\infty)$ is the **Nash profile**. We should think of the argument of this function as the volume (indeed, we get the size of A if $f = 1_A$),

So we need to find such a function which makes the inequality hold for all f; it serves a similar role as a spectral gap but with constant depending on the function. And once we do this, we can turn things into ultracontractivity bounds.

In general it's quite difficult to find such a function $\mathcal{N}$, but we can get the right answer in some cases coming from volume-growth lower bounds. We get the following fact from translation-invariance:

Proposition 15 (Pseudo-Poincaré inequality)
Define the shift operator $R_g f(x) = f(xg)$. Then

$$||f - R_g f||_2^2 \leq C|g|_S^2 \mathcal{E}(f,f),$$

where $|g|_S$ is the word length and we can take $C = \frac{1}{\min_{s \in S \cup S^{-1}} \mu(s)}$.

Proof. We have an element $g \in G$ and we can let $\ell = |g|_S$ be its word length (meaning that $g = x_1 x_2 \cdots x_\ell$). Then by the triangle inequality

$$||f - R_g f||_2 \leq \sum_{j=1}^{\ell} \left|\left|R_{x_1 \cdots x_j} f - R_{x_1 \cdots x_{j-1} f}\right|\right|_2,$$

but by translation-invariance we can simplify this to

$$\sum_{j=1}^{\ell} ||R_{x_j} f - f||_2.$$

Squaring everything and using Cauchy-Schwarz (so that we're inside the Dirichlet form, we thus get the bound

$$||f - R_g f||_2 \leq \ell \sum_{j=1}^{\ell} ||R_{x_j} f - f||_2^2.$$

Now all of these terms lie in $\mathcal{E}(f,f) = \frac{1}{2}\sum_{s \in \Gamma} ||R_s f - f||_2^2 \mu(s)$; the price we have to pay is just the factor $\mu(s)$, which has some positive minimum because we are finitely generated. This yields $\frac{2\ell}{\min \mu(s)} \mathcal{E}(f,f)$, as desired. □

The passage to obtaining a Nash inequality is now as follows. Write the average

$$f_r(x) = \frac{1}{|B_r(x)|} \sum_{y \in B_r(x)} f(y);$$

our goal is to get a Nash inequality by using the triangle inequality

$$||f_2||_2^2 \leq 2||f - f_r||_2^2 + 2||f_r||^2.$$

We will apply pseudo-Poincaré to the second term:

$$\left\| f - \frac{1}{B(\mathrm{id}, r)} \sum_{g \in B(\mathrm{id}, r)} R_g f \right\|_2 \le \frac{1}{|B(\mathrm{id}, r)|} \sum_{g \in B(\mathrm{id}, r)} ||f - R_g f||_2$$
$$\le r\mathcal{E}(f, f)^{1/2}$$

(since within the ball $B(\mathrm{id}, r)$ the word length is always at most r). So plugging this in and also use $|f_r(x)| \le \frac{||f||_1}{B(\mathrm{id}, r)}$ on the second term (details left as an exercise) to get

$$||f||_2^2 \lesssim 2r^2\mathcal{E}(f, f) + 2 \cdot \frac{||f||_1^2}{|B(\mathrm{id}, r)|}.$$

So we have two terms corresponding to the Dirichlet form and the L^1 norm; we can now optimize r large enough so that the latter term is less than half of $||f||_2^2$. We thus get the following result:

Theorem 16 (Varopoulos bound)
We have (for some constant C)

$$||f||_2^2 \le C \left(V^{-1} \left(\frac{4||f||_1^2}{||f||_2^2} \right) \right)^2 \mathcal{E}(f, f),$$

where $V(r) = |B(\mathrm{id}, r)|$ is the volume growth function.

We should really think of this V^{-1} as having a function which takes in a volume of our group and yields a radius.

Example 17
Suppose we know that we have a polynomial-growth lower bound $V(r) \ge cr^D$. Then this inequality yields a Nash inequality with Nash profile $\mathcal{N}(v) \le v^{2/D}$, yielding bounds on the return probability $\mu^{(2n)}(\mathrm{id}) \lesssim n^{-D/2}$. On the other hand if we have exponential growth (so $V(r) \ge e^{\lambda r}$ for some $\lambda > 0$), then $\mathcal{N}(v) \le (\log v)^2$, yielding $\mu^{(2n)}(\mathrm{id}) \lesssim e^{-n^{1/3}}$.

This "slowest possible" $n^{1/3}$ in the exponent is not necessarily a good bound: if we think about return probability behaviors, we can have things like the non-amenable free group where the return probability is exponentially small. And even if we focus only on subexponential situations, we have for any function $t^{1/3} \le \phi(t) \le t$ with $\frac{\phi(t)}{t} \to 0$ and satisfying some regularity, there is some finitely generated group of exponential growth with $\mu^{(2n)}(\mathrm{id}) \sim e^{-\phi(t)}$. So volume is not sufficient information to tell the behavior of the Nash profile or on-diagonal decay; we have to look more into the isoperimetry of the group for that. But the strength is that we don't need anything else besides a volume lower bound!

On the other hand, the polynomial growth bound is indeed sharp, and there are various ways to do this.

Theorem 18 (Hebsich, Saloff-Coste)
Assume as before that μ is symmetric and finitely supported, and assume that Γ has volume growth $V(r) \sim r^D$. Then there are some constants C_1, C_2, C_3, C_4 such that

$$C_1 n^{-D/2} \exp\left(-\frac{|g|_S^2}{C_2 n}\right) \le \mu^{(2n)}(g) \le C_3 n^{-D/2} \exp\left(-\frac{|g|_S^2}{C_4 n}\right).$$

The idea is that we can always upgrade an on-diagonal bound into off-diagonal ones, and then we can use near-diagonal regularity to get lower bounds. So heat kernel bounds will do the job here!

What we want to do next is understand the central limit theorem type results; we discussed last time that in order to have a valid process, we need a certain structure which we can scale, and our object is quite constrained. The earliest example is the following (coming from a paper called "Compositions of measures on the simplest nilpotent group"):

Theorem 19 (Tutubalin, 1964)

Recall that the Heisenberg group $H_3(\mathbb{Z})$ is generated by two particular matrices X, Y. We can similarly define the Heisenberg group $H_3(\mathbb{R})$ to be the set of matrices of the form $\begin{bmatrix} 1 & * & * \\ 0 & 1 & * \\ 0 & 0 & 1 \end{bmatrix}$. Let $\pi : H_3(\mathbb{R}) \to \mathbb{R}^2$ project the matrix to its $(1, 2)$ and $(2, 3)$ entries. Suppose μ is compactly supported and centered (meaning that $\mathbb{E}[\pi(X_1)] = 0$ for $X_1 \sim \mu$). Define the group automorphism

$$\delta_t \left(\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & tx & t^2 z \\ 0 & 1 & ty \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $\delta_{\frac{1}{\sqrt{n}}}(W_n)$ admits a density with respect to Haar measure on $H_3(\mathbb{R})$.

Furthermore, we can also get a functional central limit theorem, and the limiting process (which will take the place of Brownian motion) will have generator called the "sub-Laplacian" $\mathcal{L}_\mu$, which is like the Laplace operator on the $(1, 2)$ (x) and $(2, 3)$ (y) coordinates.

Fact 20

We can also consider the non-centered case, meaning that we see a drift when projected onto the x and y coordinates. Then if $v_\mu = \mathbb{E}[\pi(X_1)]$ is nonzero, we must subtract off the drift and look at fluctuations around the mean. If W_n is our μ-random walk, we can then define the recentered walk $W_n \exp(-nv_\mu)$, and this time the contributions to the "z" coordinate cannot be ignored. Indeed, we now define the dilation

$$D_t \left(\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & tx & t^3 z \\ 0 & 1 & ty \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $D_{\frac{1}{\sqrt{n}}}(W_n \exp(-nv_\mu))$ converges to a Gaussian distribution on $\mathbb{R}^3$.

The general procedure for nilpotent groups goes like the following: first we find a filtration adapted to the random walk which captures how fast we spread out in each coordinate, and we also need an approximate dilation which is consistent with the filtration (this is the scaling function), and a certain "stratification" of the group structure so that in the limit, we get some limit group on which the dilations become automorphisms.

We'll next mention an estimate – this filtration on the nilpotent group is somehow the most fundamental structure that control how fast random walks go. We'll talk about that in more detail next time, though.

Theorem 21 (Brieussel, Tessera, Zheng)
Let G be a locally compact, compactly generated group of polynomial growth (now polynomial growth means looking at $m(S^n)$ under the Haar measure); let $|\cdot|_S$ be the corresponding word length for the generating set. Let μ be compactly supported and centered. Then we get the tail growth (for some constant $c > 0$)

$$\mathbb{P}\left(\max_{1 \le t \le n} |W_t|_S \ge x\right) \le C \exp\left(-c\frac{x^2}{n}\right).$$

The reason why we see this once we have a compactly supported set is a calculation using the lower central series – we don't need detailed heat kernel estimates or anything. But there is the following useful application:

Corollary 22
Let Γ be a finitely generated group sitting in the group of $n \times n$ upper-triangular matrices with rational coefficients (these are called **torsion-free solvable groups of finite Prüfer rank**). Then simple random walk on Γ also satisfies the same Gaussian-type tail estimate.

There is some structure theory behind this, showing that such groups can be embedded into solvable Lie groups with exponentially-distorted normal subgroup. The point is that the exponentially distorted kernel is quite big, but to get control over displacement we can reduce to the polynomial growth version, and because of the filtration we can get estimates using very generic methods. And if we want to understand behavior of return probabilities, it turns out that we get

$$\mu^{(2n)}(\text{id}) \sim e^{-n^{1/3}},$$

so we're still quite small in the sense that we're controlled by volume and we're still kind of traveling a distance $\sqrt{n}$ away. We'll see some more examples of filtrations next time and understand some more scaling limits; the point is that for limiting behavior on nilpotent groups, it's just about getting the right filtration and stratification!

3 June 20, 2025

We'll look at some central limit theorems for random walks on nilpotent groups. We'll focus on the symmetric case – for the non-centered case, we can see some papers by Bénard and Breuillard

Our setup today will be a finitely generated nilpotent group $\Gamma = \langle S \rangle$, where we fix a symmetric probability measure μ on Γ. We want a space on which we can scale and take limits, so we can embed Γ in a nilpotent group as long as torsion-free (meaning there are no finite-order elements other than the identity).

Fact 23
In such a setting, we have the **Malcev embedding**, which says that Γ embeds as a cocompact lattice in a simply connected nilpotent Lie group N.

We'll let $\mathfrak{n}$ be its Lie algebra, which will give us coordinates to work with for the random walk on the group. In general we can ask whether we can find this "envelope" N, but that is a subject of its own.

Example 24

Some examples to keep in mind for this are $\mathbb{Z}^d$ embedded in $\mathbb{R}^d$, as well as the Heisenberg group $H_3(\mathbb{Z})$ embedded in $H_3(\mathbb{R})$.

We can now introduce the filtration – we'll include two classes of examples. First, suppose μ has finite generating support (the simple random walk case); last time, we had some discussion of scaling and displacement of the random walk. More formally, we have the exponential map $\exp : \mathfrak{n} \to N$ which we can think of as a coordinate map, and we can multiply using the **Baker-Campbell-Hausdorff formula**

$$\exp(X) * \exp(Y) = \exp\left(X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] + \frac{1}{12}[Y, [X, Y]] + \cdots\right);$$

since N is nilpotent, all brackets of a certain length are going to vanish and thus the formula will end, meaning all coordinate operations are polynomials. The adapted filtration is then the lower central series

$$\mathfrak{n}_1 = \mathfrak{n}, \quad \mathfrak{n}_2 = [\mathfrak{n}, \mathfrak{n}], \quad \mathfrak{n}_3 = [\mathfrak{n}, \mathfrak{n}_2], \cdots, \mathfrak{n}_{c+1} = \{0\},$$

where c is the "nilpotency class." We write $\mathcal{F}$ for the filtration $\mathfrak{n}_1 \supseteq \mathfrak{n}_2 \supseteq \cdots \supseteq \mathfrak{n}_{c+1}$, and now we'll choose $\mathfrak{m}_i$ so that $\mathfrak{m}_i \oplus \mathfrak{n}_{i+1} = \mathfrak{n}_i$ as vector space complements so that we can write $\mathfrak{n} = \bigoplus_{i=1}^{c} \mathfrak{m}_i$ in some non-canonical way.

If we now project onto $\mathfrak{m}_1$, then we're just doing a simple random walk on the corresponding **abelian** group because all brackets disappear, so we know exactly what happens (diffusion behavior, central limit theorem, and so on). But if we move on to $\mathfrak{m}_2$, we only need to worry about terms like $\exp(X) * \exp(Y) = \exp\left(X + Y + \frac{1}{2}[X, Y]\right)$, and thus we change the scale at which we see contributions when convolving after n steps. What we're saying is that the projection $\pi_i : \mathfrak{n} \to \mathfrak{m}_i$ should send our random walk W_n as

$$||\pi_i(W_n)||_2 \approx n^{i/2}.$$

This can in fact be rigorously justified by comparing these norms with word distances.

Definition 25

Given a filtration and corresponding vector space decomposition, we compute the **homogeneous quasi-norm** via

$$|x|_{\mathcal{F}} = \sum_{i=1}^{c} ||\pi_i(x)||^{1/i}.$$

The idea is that every term is on the order of $\sqrt{n}$, and now the Guivarc'h lemma tells us that if Ω is a compact generating set of N, then we have some constant C so that

$$\frac{1}{C}|x|_{\mathcal{F}} \le |\exp(x)|_{\Omega} \le C(|x|_{\mathcal{F}} + 1),$$

and thus up to constants it makes sense to work with the quasi-norm. We can then translate the Hebsich and Saloff-Coste result to see that this is the right estimate. So we know how the coordinates behave and know the way in which to scale the Lie algebra – this is called the **approximate dilation** $\delta_t : \mathfrak{n} \to \mathfrak{n}$, defined by

$$\delta_t(x) = \sum_{i=1}^{c} t^i \pi_i(x).$$

The scaling is correct if we take simple random walk, and we would like to evaluate the limit of $\delta_{\frac{1}{\sqrt{n}}}(W_n)$. Note though that δ_t is not necessarily an automorphism of $\mathfrak{n}$, so we actually have to change the bracket structure (and this is what is called **stratification**): for a nilpotent group N, we define a modification of the usual element $*$ as follows. Letting x and y stand in for their Lie group equivalents for shorthand (so exponentiate and then go backward), we define

$$x *_t y = \delta_{1/t}\left(\delta_t(x) * \delta_t(y)\right),$$

and then define $x *_\infty y = \lim_{t\to\infty} x *_y y$. The space $(N, *_\infty)$ then admits each δ_t as an automorphism.

Example 26

The Heisenberg group itself is already stratified, so to give a non-stratified example we need to go up in dimension. The smallest example is in dimension 7, so we have some basis vectors $X_1, \cdots, X_7$ and the Lie bracket is given by

$$[X_1, X_i] = X_{i+1} \text{ for } i \in \{2,3,4,5,6\}, \quad [X_2, X_3] = X_6, \quad [X_2, X_4] = -[X_2, X_5] = [X_3, X_4] = X_7,$$

and $[X_s, X_t] = 0$ for $s + t > 7$. This is a seven-dimensional nilpotent LIe algebra

This is not stratified – the first slice means we look for basis vectors that cannot be written as brackets, which would be

$$\mathfrak{m}_1 = \langle X_1, X_2 \rangle.$$

Next, we see the brackets that we cannot be written as brackets of length 3 and so on, and we see

$$\mathfrak{m}_2 = \langle X_3 \rangle, \quad \mathfrak{m}_3 = \langle X_4 \rangle, \quad \cdots, \quad \mathfrak{m}_6 = \langle X_7 \rangle.$$

So if we define our dilations with respect to this, we won't get an automorphism, since

$$[\delta_t(X_2), \delta_t(X_3)] = [tX_2, t^2 X_3] = t^3 X_6,$$

which is different from $\delta_t([X_2, X_3]) = t^5 X_6$ (since we can write X_6 as a bracket of length 5). So instead we want to try to stratify, and we do so by dilating to be large and then dilating back. That's exactly what gets rid of the **additional relations that mismatch levels** which give us "shortcuts;" the only relations left are $[X_1, X_i] = X_{i+1}$. So the central limit theorem we get can be stated as follows:

Theorem 27 (Crepel-Raugi central limit theorem)

We have in the space of continuous paths that

$$\left(\delta_{1/\sqrt{n}}(W_{\lfloor ns \rfloor})\right)_{s\geq 0} \implies (B(s))_{s\geq 0},$$

where $B(s)$ is the diffusion on the stratified space $(\mathcal{N}, *_\infty)$ with sub-Laplacian generator "only having directions in the first layer $\mathfrak{m}_1$:"

$$\mathcal{L} = \frac{1}{2} \sum_{i=1}^{\dim \mathfrak{m}_1} E_i^2,$$

where E_i is a basis of the first slice of the Lie algebra such that $\pi_1(\mu)$ has covariance matrix Id.

So in the end, we see the diffusion driven by the sub-Laplacian, moving on the group $\mathcal{N}$ whose infinitesimal directions are only in the first slice. In our example above, the idea is that when we scale, X_6 will get scaled by $n^{5/2}$, which is stronger than what the brackets can contribute so the relation goes away. Notably, it's important that X_1, X_2 generate the whole group here so that the limiting diffusion has a density. (And the noncentered case is more complicated, so we won't describe it here.)

We're now ready to go to non-local operators – by background on the limiting processes for graded nilpotent connected Lie groups G (studied by Hunt), we must have a symmetric Lévy process, and such processes have a generator of the form (this is the **Lévy-Khintchine formula**)

$$\mathcal{L}f(g) = \sum_{i,j=1}^{d} a_{ij} X_i X_j f(g) + \int_{G \setminus \{e\}} f(gh) - f(g) d\mathcal{J}(h)$$

which is a decomposition into a diffusion term and a jump term. Here $\mathcal{J}$ is some Radon measure on $G \setminus \{e\}$ satisfying the property that

$$\int \min\left(1, ||x||^2\right) d\mathcal{J}(x) < \infty.$$

Example 28

We'll care about the case where we actually have **pure** jump processes – consider the following "α-stable-like walk" μ which is easy to think about. Let Γ have generating set $(S_1, \cdots, S_k)$, and let $\alpha_i \in (0, 2)$ for all $1 \le i \le k$. Let μ_i be a measure on $\mathbb{Z}$ with $\mu_i(x) = \frac{C_i}{(1+|x|)^{\alpha_i+1}}$ (this mimicks the usual α-stable random walk in one dimension); we push it forward to $\langle S_i \rangle$ and write μ_i for the pushforward as well. We then define the random walk

$$\mu = \frac{1}{k} \sum_{l=1}^{k} \mu_i.$$

So basically we choose some directions and then do a spread out α-stable walk along those directions. For scaling limits of this, we need to understand how to scale in terms of the α_is as well: on $\mathbb{Z}$ the correct scaling is

$$\frac{1}{n^{1/\alpha_i}} \mu_i^{(n)} \to \alpha_i\text{-stable law},$$

so we may think that we should scale appropriately like that, but unfortunately the brackets complicate things a bit. So what we need to do is associate a filtration adapted to this random walk μ; it turns out we can do things in the same way as the lower central series for simple random walk, but we need to introduce weights to our brackets. Specifically, we do the following: each S_i corresponds to some direction in the Lie algebra

$$S_i = \exp(\sigma_i), \quad \sigma_i \in \mathfrak{n},$$

and now we assign a weight (how much to scale to get a central limit theorem) $w(\sigma_i) = \frac{1}{\alpha_i}$ to σ_i. We then further also assign weights to the **formal commutators**

$$w([\sigma_i, \sigma_j]) = \frac{1}{\alpha_i} + \frac{1}{\alpha_j}$$

and so on, inductively making it so that $w([\sigma_i, X]) = w(X) + w(\sigma_i)$. (Again, this is completely formal without looking at the structure of the Lie algebra.) What we do after that is to evaluate the commutators (actually use the bracket relations) and consider

$$\mathfrak{n}_s = \text{ideal generated by commutators of weight at least } s.$$

This yields a filtration of the form $\mathfrak{n} = \mathfrak{n}_{w_1} \supseteq \mathfrak{n}_{w_2} \supseteq \cdots \supseteq \mathfrak{n}_{w_{j_*}} \supseteq \{0\}$, where we set the different cutoffs based on all possible combinations of these weights. We can thus estimate (this result is due to Professor Zheng and Saloff-Coste)

$$\mu^{(2n)}(e) \sim n^{-\gamma(S,\alpha)}$$

when we have a generating set $S = (s_1, \cdots, s_k)$, some specified numbers $\alpha = (\alpha_1, \cdots, \alpha_k)$ between 0 and 2, and where the numbers

$$\gamma(S,\alpha) = \sum_{i=1}^{j_*} w_i \dim(\eta_{w_i}/\mathfrak{n}_{w_{i+1}});$$

we can think of this as the determinant of the scaling matrix and giving us the "on-diagonal terms." To get scaling limits, we have to do a bit more: this is based on the idea of heat kernel regularity (Hölder continuity estimates) and exit time estimates, and we can see the paper by Bass, Kumagai, and Uemura for details (which also allow for diffusion parts). The result is the following:

Theorem 29 (Chen, Kumagai, Wang, Saloff-Coste, Zheng)
For the random walk $\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_{\alpha_i S_i}$, define the dilation adapted to the filtration

$$\delta_t(x) = \sum_{i=1}^{j_*} t^{w_i}\pi_i(x),$$

where our projections π_i to $\mathfrak{m}_i$s are chosen as above. Assume that if we rescale our original measure down, we have $t\delta_{1/t}\mu$ converging vaguely to a Radon measure μ_∞ on $\mathbb{R}^d$. Then there is a stratified nilpotent group adapted to this dilation (same procedure as before according to the dilation); we think of μ_∞ as a jumping measure on that group. Then

$$\left(\delta_{1/n}W_{ns}\right)_{s\geq 0} \implies (Z_s)_{s\geq 0}$$

converges to a pure jump Lévy process on the stratified $(\mathcal{N}, *_\infty)$ with jump measure $\mathcal{J} = \mu_\infty$.

In some sense this is more elementary than the diffusion process – we're scaling α-stable kernels and we're going to see how we end up jumping in the limit by using the correct group structure using heat kernel estimates.