# LECTURE 3: THE SIMPLE WALK IN HIGH DIMENSIONS

Let us continue our discussion on the simple walks to higher dimensions. To do so, it helps to introduce a more abstract walk first (and briefly).

## §1. THE SIMPLE WALK ON A GRAPH

**(1.1) Graphs.** A *graph* is a collection of points (or *vertices*) and a set of neighboring relations (*edges*) between these vertices. An example of a graph is $\mathbf{Z}^1$—the one-dimensional integer lattice—which can be thought of as a graph: The vertices are $0, \pm 1, \pm 2, \ldots$ and there is an edge between two vertices $a$ and $b$ if and only if $|a-b| = 1$. In particular, every vertex has two neighbors.

An obvious generalization to this is $\mathbf{Z}^d$, which is the $d$-dimensional integer lattice. This can be thought of as a graph with vertices of type $(z_1, \ldots, z_d)$ where the $z_i$'s are integers, and there is an edge between $z = (z_1, \ldots, z_d)$ and $w = (w_1, \ldots, w_d)$ if and only if $\sum_{i=1}^d |w_i - z_i| = 1$ (check this!) So every vertex has $(2d)$ neigbors on this graph. (Can you compute this from the formal definition that I have written?)

A third example of an interesting graph is a *binary tree*. Here, you start with one vertex; it then branches into two; each of these branches into two, and so on. Check that at the $n$th level of this construction, there are $2^n$ vertices. The edges are the natural ones: Two vertices are neighbors (i.e., have an edge in common) if and only if one of them branched off into the other. You should check that evey vertex except for the first one (the *root*) has three neighbors, whereas the root has two neighbors.

As a fourth and final example, consider the *complete graph on $n$ vertices*. Here, the graph is made up of a finite number $(n)$ of vertices, and everyone is the neighbor of everyone else.

**(1.2) The Simple Walk.** The simple walk on a graph is the random process that starts someplace in the graph (call it the origin if you want), and then moves to one of the nearest neighboring vertices with equal probability. (**Warning:** This makes sense only if the graph has no vertices with infinitely many neighbors, of course.) And the walk proceeds this way, everytime going to a nearestneighbor independently of all his/her other moves, and always, all neighbors are equally likely.

## §2. THE SIMPLE WALK ON $\mathbf{Z}^2$

Returning to $S_1, S_2, \ldots$ being the simple random walk on the planar integer lattice $\mathbf{Z}^2$, we ask, "*how many times is the walk expected to return to its origin?*" We have already seen in (2.1, Lecture 2) that the one-dimensional walk returns to the origin about $\sqrt{n}$-times in the first $n$ steps, as $n \to \infty$. One should expect fewer returns for the planar walk, since there is "more space." Here is the precise result.

**(2.1) Expected Number of Returns.** *If $N_n$ denotes the number of times the simple walk returns to the original before time $n$, then for $n$ even,*

$$E\{N_n\} = \sum_{j=1}^{n/2} 4^{-2j} \binom{2j}{j}^2.$$

*In particular, for some constant $c$, $E\{N_n\} \sim c\log(n)$.*

*A Semi-Proof:* I gave a geometric proof of this in the lecture; the idea was that if you rotate the $xy$-plane, you rotate the simple walk $S_n$ on to the simple walk $\widetilde{S}_n$ which is a simple walk on the lattice in which the neighbors of the origin $(0,0)$ are the 4 points,

$$\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right),\left(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right),\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right),\left(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right).$$

Since we have only turned the plane, $S_n = 0$ if and only if $\widetilde{S}_n = 0$, so these two events have the same probability, but $P\{\widetilde{S}_n = 0\} \sim c/\sqrt{n}$ (cf. the Stirling-formula approximation in (2.4, Lecture 2)). So, $P\{\widetilde{S}_n = 0\} \sim C^2/n$. On the other hand, just as in the one-dimensional case, $E\{N_n\} = \sum_{j=1}^n P\{S_j = 0\}$, so that $E\{N_n\} \sim \sum_{j=1}^n C^2/j$. Let us see how this sum behaves:

$$\begin{aligned} E\{N_n\} &\sim C^2 \sum_{j=1}^n \frac{1}{j} = C^2 \frac{1}{n} \sum_{j=1}^n \frac{1}{(j/n)} \\ &\sim C^2 \int_{1/n}^1 \ln(x)\,dx, \end{aligned}$$

by a Riemann-Sum approximation. (How did the lower limit of the integral become $(1/n)$?) As $n \to \infty$, this behaves like $c\log(n)$—check!

When done carefully, as we did in the lecture, the exact calculation follows also. ♣

## §3. THE SIMPLE WALK ON $\mathbf{Z}^d$, $\mathbf{d} \geq 3$

In higher dimensions the rotation trick fails, but our intuition that the coordinates of $S_n$ are almost independent simple walks is in a sense correct and can be made precise. This leads to $P\{S_n = 0\} \sim (C/\sqrt{n})^d = cn^{-d/2}$. On the other hand, since $d \geq 3$, this sums and we have

**(3.1) The Simple Walk in $\mathbf{d} \geq 3$ is transient.** *We have $E\{N_\infty\} < +\infty$. Therefore, the expected number of times to hit any point is finite. Therefore, after a finite (but random) number of steps, $S_n$ will leave any finite neighborhood of the origin, and this is the property that the word "transient" is referring to.*

**§4. THE SELF-AVOIDING WALK** Certain models of polymer chemistry lead to the *self–avoiding walk*, which is defined as follows: First consider all paths of length $n$ in your favorite infinite lattice, say $\mathbf{Z}^d$. On the latter, there are $(2d)^n$ such paths, but many of them self-intersect, i.e., there are distinct $i, j \leq n$ such that $\pi_i = \pi_j$. Let $\chi_n$ denote the total number of self-avoiding paths of length $n$, and from these $\chi_n$ self-avoiding paths, choose one at random. This is the *self-avoiding walk* of length $n$.

**(4.1) Bounds on $\chi_n$.** *I claim that for every $n$, $d^n \le \chi_n \le (2d)^n$.*

Actually much better bounds are possible (say when $d = 2$), but this is good enough.

*Proof:* To get the upper bound of $(2d)^n$ note that every self-avoiding path is a path, and so $\chi_n \le$ the number of *all* paths of length $n$, which is $(2d)^n$. The lower bound is not much more difficult. When $d = 2$, note that every path that only goes "up" or to the "right" is self-avoiding. There are clearly $2^n$ such paths. Note that paths of this type (i.e., the "up-right" paths) are those that move in the direction of either vector $(1, 0)$ or $(0, 1)$.

When $d = 3$, the analogue of "up-right" paths are those that move in the direction of $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. There are $3^n$ such paths. In general, only choose the directions that keep you going "up" in the positive quadrant, and note that these paths are (i) self-avoiding; and (ii) there are $d^n$ many of them. ♣

**(4.2) The Connectivity Constant $\mathbf{C(d)}$.** *There exists a constant $d \le C(d) \le 2d$, such that*

$$\lim_{n\to\infty} \frac{\chi_n}{n} = C(d).$$

*This $C(d)$ is called the connectivity constant.*

**(4.3) Remarks.**

**a.** Such a result holds on many infinite graphs that are "self-similar."

**b.** In rough terms, the above states that $\chi_n$ behaves (roughly again!) like $(C(d))^n$ for large value of $n$.

*Proof:* Note that on every self-avoiding path of length $n+m$, certainly the first $n$ steps are self-avoiding, and the next $m$ steps are also self-avoiding. Therefore,

$$\chi_{n+m} \le \chi_n \cdot \chi_m.$$

In words, the sequence $\chi_1, \chi_2, \ldots$ is *submultiplicative.* This is equivalent to the *subadditivity* of $\log(\chi_n)$'s, i.e.,

$$\log(\chi_{n+m}) \le \log(\chi_n) + \log(\chi_m).$$

Therefore, by the subbadditivity lemma below, $\log(\chi_n)/n$ has a limit. Note that this limit is between $d$ and $(2d)$ by (4.1). ♣

**(4.4) The Subadditivity Lemma.** *Any sequence $a_1, a_2, \ldots$ that is subadditive (i.e., $a_{n+m} \le a_n + a_m$) satisfies*

$$\lim_{k\to\infty} \frac{a_k}{k} = \min_{n\ge 1}\left(\frac{a_n}{n}\right).$$

In particular, the above limit is always $\le a_1$ which is finite. However, this limit could be $-\infty$!

**(4.5) Limits.** I will prove this shortly. However, we need to be careful when dealing with limits, especially since the entire point of this exercise is to show that the limit exists. So let us start with some preliminaries: For any sequence $x_1, x_2, \ldots$

$$\limsup_{k\to\infty} x_k := \min_{n\ge 1} \max_{j\ge n} x_j, \qquad \text{and} \qquad \liminf_{k\to\infty} x_k := \max_{n\ge 1} \min_{j\ge n} x_j.$$

In other words, the lim sup is the largest possible accumulation point of the $x_j$'s and the lim inf is the smallest. It should be obvious that for any sequence $x_1, x_2 \ldots$, we always have $\liminf_j x_j \leq \limsup_j x_j$. When the two are equal, this value is the limit $\lim_j x_j$, and this is the only case in which the limit exists.

> *Exercise 1.* For our first example, consider the sequence $x_j := 1/j$ $(j = 1, 2 \cdots)$. Then you should check that $\liminf_{j\to\infty} x_j = \limsup_{j\to\infty} = 0$. More generally, check that for any sequence $x_1, x_2, \ldots$, $\lim_{j\to\infty} x_j$ exists if and only if $\liminf_{j\to\infty} x_j = \limsup_{j\to\infty} x_j$.
>
> *Exercise 2.* Show that the sequence $x_j := (-1)^j/j$ $(j = 1, 2, \ldots)$ has no limit. Do this by explicitly computing $\liminf_j x_j$ and $\limsup_j x_j$.
>
> *Exercise 3.* A point $a$ is defined to be an `accumulation point` for the sequence $x_1, x_2, \ldots$ if there exists a subsequence $n(k) \to \infty$, such that $x_{n(k)} \to a$. Show that $\limsup_j x_j$ and $\liminf_j x_j$ are always accumulation points of $(x_j)$.
>
> *Exercise 4.* Show that the sequence of Exercise 2 only has 2 accumulation points. Construct a sequence $x_1, x_2, \ldots$ that has $k$ accumulation points for any predescribed integer $k$. Can you construct a sequence $x_1, x_2, \ldots$ that has infinitely many accumulation points?

Now we are ready for

**(4.6) Proof of (4.4).** Since $a_k/k \geq \min_n(a_n/n)$ for any $k$, it follows that

$$\liminf_{k\to\infty} \frac{a_k}{k} \geq \min_n \left(\frac{a_n}{n}\right).$$

It suffices to show that $\limsup_{k\to\infty}(a_k/k) \leq \min_n(a_n/n)$. (For then, the lim sup and the lim inf agree.) We do this in a few easy stages: Thanks to subbadditivity, $a_k \leq a_{k-1} + a_1$. But the same inequality shows that $a_{k-1} \leq a_{k-2} + a_1$, so that by iterating this we get

$$\begin{aligned}
a_k &\leq a_{k-1} + a_1 \\
&\leq a_{k-2} + a_1 + a_1 = a_{k-2} + 2a_1 \\
&\leq a_{k-3} + 3a_1 \\
&\ \ \vdots \\
&\leq ka_1.
\end{aligned}$$

Therefore, $\limsup_k(a_k/k) \leq a_1$. Next, we show that this lim sup is also $\leq (a_2/2)$. "By induction," this argument boosts itself up to show that for any $n$, $\limsup_k(a_k/k) \leq (a_n/n)$, which is what we want to show but in disguise.

To finish, I will show that

$$\limsup_{k\to\infty} \frac{a_k}{k} \leq \frac{a_2}{2}. \tag{4.7}$$

I will then leave the "induction" part up to you as a nice exercise.

By subbaditivity, for all $k > 2$, $a_k \leq a_{k-2} + a_2$. Applying it again, subadditivity yields $a_k \leq a_{k-4} + 2a_2$ for all $k > 4$ and so on. In general, we see that for all $k > 2j$,

$$a_k \leq a_{k-2j} + ja_2. \tag{4.8}$$

Now, if $k$ is even, choose $j = (k/2) - 1$ to see that (a) $k > 2j$; and so (b) $a_k \leq (k/2)a_2$. If $k$ is odd, choose $j = (k-1)/2$ to see that (c) $k > 2j$; and so (d) $a_k \leq a_1 + \frac{k-1}{2}a_2$. So regardless of whether or not $k$ is even, we always have

$$a_k \leq \left(\frac{k}{2}\right) a_2 + |a_1| + |a_2|.$$

(why?) Divide by $k$ and let $k \to \infty$ to deduce (4.7). ♣

**(4.9) Exercise on the Connectivity Constant.** Improve (4.1) by showing that in all dimensions, $\chi_n \leq (2d)\cdot(2d-1)^{n-1}$. Conclude from this and from (4.2) the following slightly better bound on the connectivity constant: $d \leq C(d) \leq (2d-1)$, e.g., $2 \leq C(2) \leq 3$.
(`Hint.` For step 1, you have $(2d)$ choices, but then you cannot go back to where you were.)