

Markov Chains, Conductance and Mixing Time

Lecturer: Santosh Vempala

Scribe: Kevin Zatloukal

1 Markov Chains

A random walk is a Markov chain. As before, we have a state space Ω . We also have a set of subsets \mathcal{A} of Ω that form a σ -algebra, i.e., \mathcal{A} is closed under complements and countable unions. It is an immediate consequence that \mathcal{A} is closed under countable intersections and that $\emptyset, \Omega \in \mathcal{A}$. For any $u \in \Omega$ and $A \in \mathcal{A}$, we have the one-step probability, $P_u(A)$, which tells us the probability of being in A after taking one step from u . Lastly, we have a starting distribution Q_0 on Ω , which gives us a probability $Q_0(A)$ of starting in the set $A \in \mathcal{A}$.

With this setup, a **Markov chain** is a sequence of points w_0, w_1, w_2, \dots such that $P(w_0 \in A) = Q_0(A)$ and

$$P(w_{i+1} \in A \mid w_0 = u_0, \dots, w_i = u_i) = P(w_{i+1} \in A \mid w_i = u_i) = P_{u_i}(A),$$

for any $A \in \mathcal{A}$. A distribution Q is called **stationary** if, for all $A \in \mathcal{A}$,

$$Q(A) = \int_{u \in \Omega} P_u(A) dQ(u).$$

In other words, Q is stationary if the probability of being in A is the same after one step. We also have a generalized version of the symmetry we saw above in the transition probabilities ($p_{xy} = p_{yx}$). The Markov chain is called **time-reversible** if, for all $A, B \in \mathcal{A}$,

$$P(w_{i+1} \in B \mid w_i \in A) = P(w_{i+1} \in A \mid w_i \in B).$$

2 Ball Walk

The following algorithm, called Ball-Walk, is a continuous random walk. In this case, the set corresponding to the neighborhood of x is $B_\delta(x) = x + \delta\mathbb{B}_n$.

Algorithm Ball-Walk(δ):

1. Let x be a starting point in K .
2. Repeat sufficiently many times:
 - Choose a random $y \in B_\delta(x)$.
 - If $y \in K$, set $x = y$.

Here, the state space is all of the set, so $\Omega = K$. The σ -algebra is the measurable subsets of K , as is usual. We can define a density function for the probability of transitioning from $u \in K$ to v , provided that $u \neq v$:

$$p(u, v) = \begin{cases} \frac{1}{\text{Vol}(\delta\mathbb{B}_n)} & \text{if } v \in K \cap B_\delta(u), \\ 0 & \text{otherwise.} \end{cases}$$

The probability of staying at u is

$$P_u(u) = 1 - \frac{\text{Vol}(K \cap B_\delta(u))}{\text{Vol}(\delta\mathbb{B}_n)}.$$

Putting these together, the probability of transitioning from u to any measurable subset A is

$$P_u(A) = \begin{cases} \frac{\text{Vol}(A \cap K \cap B_\delta(u))}{\text{Vol}(\delta\mathbb{B}_n)} + P_u(u) & \text{if } u \in A \\ \frac{\text{Vol}(A \cap K \cap B_\delta(u))}{\text{Vol}(\delta\mathbb{B}_n)} & \text{if } u \notin A. \end{cases}$$

Since the density function is symmetric, it is easy to see that the uniform distribution is stationary. We can also verify this directly. We can compute the probability of being in A after one step by adding up the probability that we transition to u for each $u \in A$. Thus, after one step from the uniform distribution $dQ(u) = \frac{1}{\text{Vol}(K)} du$, the probability of being in A is

$$\begin{aligned} & \int_{u \in A} P_u(u) dQ(u) + \int_{u \in A} \int_{v \in K \cap B_\delta(u) \setminus \{u\}} \frac{1}{\text{Vol}(\delta\mathbb{B}_n)} dQ(v) du \\ &= \int_{u \in A} \left(1 - \frac{\text{Vol}(K \cap B_\delta(u))}{\text{Vol}(\delta\mathbb{B}_n)} \right) dQ(u) + \int_{u \in A} \int_{v \in K \cap B_\delta(u)} \frac{1}{\text{Vol}(\delta\mathbb{B}_n)} dQ(v) du \\ &= \int_{u \in A} dQ(u) + \frac{1}{\text{Vol}(\delta\mathbb{B}_n)} \int_{u \in A} \left(\int_{v \in K \cap B_\delta(u)} dv - \text{Vol}(K \cap B_\delta(u)) \right) dQ(u) \\ &= \int_{u \in A} dQ(u) \\ &= Q(A) \end{aligned}$$

Another way to look at the one-step probability distributions is in terms of flow. For any subset $A \in \mathcal{A}$, the ergodic flow $\Phi(A)$ is the probability of transitioning from A to $\Omega \setminus A$, i.e.,

$$\Phi(A) = \int_{u \in A} P_u(\Omega \setminus A) dQ.$$

Intuitively, in a stationary distribution, we should have $\Phi(A) = \Phi(\Omega \setminus A)$. In fact, this is a characterization of the stationary distribution.

Theorem 1. *A distribution Q is stationary iff $\Phi(A) = \Phi(\Omega \setminus A)$ for all $A \in \mathcal{A}$.*

Proof. Consider their difference

$$\begin{aligned}\Phi(A) - \Phi(\Omega \setminus A) &= \int_{u \in A} P_u(\Omega \setminus A) dQ(u) - \int_{u \notin A} P_u(A) dQ(u) \\ &= \int_{u \in A} (1 - P_u(A)) dQ(u) - \int_{u \notin A} P_u(A) dQ(u) \\ &= Q(A) - \int_{u \in \Omega} P_u(A) dQ(u).\end{aligned}$$

The latter quantity is the probability of staying in A after one step, so $\Phi(A) - \Phi(\Omega \setminus A) = 0$ iff Q is stationary. \square

Now, let's return to the question of how quickly (and whether) this random walk converges to the stationary distribution. As before, it is convenient to make the walk lazy by giving probability $\frac{1}{2}$ of staying in place instead of taking a step. This means that $P_u(\{u\}) \geq \frac{1}{2}$ and, more generally, for any $A \in \mathcal{A}$, $P_u(A) \geq \frac{1}{2}$ if $u \in A$ and $P_u(A) < \frac{1}{2}$ if $u \notin A$. Also as before, the notion of conductance will be useful. In this general context, we define conductance as

$$\phi = \min_{A \in \mathcal{A}, Q(A) \leq \frac{1}{2}} \frac{\Phi(A)}{Q(A)},$$

where Q is the stationary distribution. This is the probability of transitioning to $\Omega \setminus A$ given that we are starting in A .

We must also generalize the notion of the distance of a distribution from stationary. The straightforward generalization of our previous definition is

$$\begin{aligned}|Q_t - Q| &= \frac{1}{2} \int_{u \in \Omega} |dQ_t(u) - dQ(u)| \\ &= \frac{1}{2} \int_{u \in Q_t^+} dQ_t(u) - dQ(u) + \frac{1}{2} \int_{u \in Q_t^-} dQ(u) - dQ_t(u),\end{aligned}$$

where

$$\begin{aligned}Q_t^+ &= \{u \in \Omega \mid dQ_t(u) \geq dQ(u)\}, \text{ and} \\ Q_t^- &= \{u \in \Omega \mid dQ_t(u) < dQ(u)\}.\end{aligned}$$

Since we know that

$$1 = \int_{u \in Q_t^+} dQ_t(u) + \int_{u \in Q_t^-} dQ_t(u) = \int_{u \in Q_t^+} dQ(u) + \int_{u \in Q_t^-} dQ(u),$$

we can rearrange to get

$$\int_{u \in Q_t^+} dQ_t(u) - \int_{u \in Q_t^+} dQ(u) = \int_{u \in Q_t^-} dQ(u) - \int_{u \in Q_t^-} dQ_t(u),$$

which shows that the two terms from above are equal. Therefore, we can just as well define

$$|Q_t - Q| = \int_{u \in Q_t^+} dQ_t(u) - dQ(u) = \sup_{A \in \mathcal{A}} Q_t(A) - Q(A).$$

We will take this as our definition of the distance $d(Q_t, Q)$ between distributions Q_t and Q .

We would like to know how large t must be before $d(Q_t, Q) \leq \frac{1}{2}d(Q_0, Q)$. This is one definition of the mixing time. It would be nice to get a bound by showing that $Q_t(A) - Q(A)$ drops quickly for every A . This is not the case. Instead, we can look at $\sup_{Q(A)=x} Q_t(A) - Q(A)$ for each fixed $x \in [0, 1]$. A bound for every x would imply the bound we need. To prove a bound on this by induction, we will define it in a formally weaker way. Our upper bound is

$$h_t(x) = \sup_{g \in \mathcal{F}_x} \int_{u \in \Omega} g(u) (dQ_t(u) - dQ(u)) = \sup_{g \in \mathcal{F}_x} \int_{u \in \Omega} g(u) dQ_t(u) - x,$$

where \mathcal{F}_x is the set of functions

$$\mathcal{F}_x = \left\{ g : \Omega \rightarrow [0, 1] : \int_{u \in \Omega} g(u) dQ(u) = x \right\}.$$

It is clear that $h_t(x)$ is an upper bound on $\sup_{Q(A)=x} Q_t(A) - Q(A)$ since g could be the characteristic of A . The following lemma shows that these two quantities are in fact equal as long as Q is atom-free, i.e., there is no $x \in \Omega$ such that $Q(\{x\}) > 0$.

Lemma 2. *If Q is atom-free, then $h_t(x) = \sup_{Q(A)=x} Q_t(A) - Q(A)$.*

Proof (Sketch). Consider the set of points X that maximize dQ_t/dQ , their value density. (This part would not be possible if Q had an atom.) Put $g(x) = 1$ for all $x \in X$. These points give the maximum payoff per unit of weight from g , so it is optimal to put as much weight on them as possible. Now, find the set of maximizing points in $\Omega \setminus X$. Set $g(x) = 1$ at these points. Continue until the set of points with $g(x) = 1$ has measure x . \square

In fact, this argument shows that when Q is atom-free, we can find a set A that achieves the supremum. When Q has atoms, we can include the high value atoms and use this procedure on the non-atom subset of Ω ; however, we made to include a fraction of one atom to achieve the supremum. This shows that $h_t(x)$ can be achieved by a function g that is 0 – 1 valued everywhere except for at most one point.

Another important fact about h_t is the following.

Lemma 3. *The function h_t is concave.*

Proof. Let $g_1 \in \mathcal{F}_x$ and $g_2 \in \mathcal{F}_y$. Then, we can see that $\alpha g_1 + (1 - \alpha)g_2 \in \mathcal{F}_{\alpha x + (1 - \alpha)y}$, which implies that

$$\begin{aligned} & h_t(\alpha x + (1 - \alpha)y) \\ & \geq \int_{u \in \Omega} \alpha g_1(u) + (1 - \alpha)g_2(u) dQ_t(u) - (\alpha x + (1 - \alpha)y) \\ & = \alpha \left(\int_{u \in \Omega} g_1(u) dQ_t(u) - x \right) + (1 - \alpha) \left(\int_{u \in \Omega} g_2(u) dQ_t(u) - y \right). \end{aligned}$$

Since this holds for any such g_1 and g_2 , it must hold if we take the sup, which gives us $h_t(\alpha x + (1 - \alpha)y) \geq \alpha h_t(x) + (1 - \alpha)h_t(y)$. Thus, h_t is concave. \square

Now, we come to the main lemma relating h_t to h_{t-1} . This will allow us to put a bound on $d(Q_t, Q)$.

Lemma 4. *Let Q be atom-free, and $y = \min\{x, 1 - x\}$. Then*

$$h_t(x) \leq \frac{1}{2}h_{t-1}(x - 2\phi y) + \frac{1}{2}h_{t-1}(x + 2\phi y).$$

Proof. Assume that $y = x$, i.e., $x \leq \frac{1}{2}$. The other part is similar. We will construct two functions, g_1 and g_2 , and use these to bound $h_t(x)$. Let $A \in \mathcal{A}$ be a subset to be chosen later with $Q(A) = x$. Let

$$g_1(u) = \begin{cases} 2P_u(A) - 1 & \text{if } u \in A, \\ 0 & \text{if } u \notin A, \end{cases} \quad \text{and} \quad g_2(u) = \begin{cases} 1 & \text{if } u \in A, \\ 2P_u(A) & \text{if } u \notin A. \end{cases}$$

First, note that $(\frac{1}{2}g_1 + \frac{1}{2}g_2)(u) = P_u(A)$ for all $u \in \Omega$, which means that

$$\begin{aligned} & \frac{1}{2} \int_{u \in \Omega} g_1(u) dQ_{t-1}(u) + \frac{1}{2} \int_{u \in \Omega} g_2(u) dQ_{t-1}(u) \\ & = \int_{u \in \Omega} \left(\frac{1}{2}g_1(u) + \frac{1}{2}g_2(u) \right) dQ_{t-1}(u) \\ & = \int_{u \in \Omega} P_u(A) dQ_{t-1}(u) \\ & = Q_t(A) \end{aligned}$$

On the other hand,

$$\frac{1}{2} \int_{u \in \Omega} g_1(u) dQ(u) + \frac{1}{2} \int_{u \in \Omega} g_2(u) dQ(u) = \int_{u \in \Omega} P_u(A) dQ(u) = Q(A)$$

since Q is stationary. Hence $(\frac{1}{2}g_1 + \frac{1}{2}g_2) \in \mathcal{F}_x$. Putting these together, we have

$$\begin{aligned} & \frac{1}{2} \int_{u \in \Omega} g_1(u) (dQ_{t-1}(u) - dQ(u)) + \frac{1}{2} \int_{u \in \Omega} g_2(u) (dQ_{t-1}(u) - dQ(u)) \\ &= Q_t(A) - Q(A). \end{aligned}$$

Since Q is atom-free, there is a subset $A \subseteq \Omega$ such that

$$\begin{aligned} h_t(x) &= Q_t(A) - Q(A) \\ &= \frac{1}{2} \int_{u \in \Omega} g_1(u) (dQ_{t-1}(u) - dQ(u)) + \\ & \quad \frac{1}{2} \int_{u \in \Omega} g_2(u) (dQ_{t-1}(u) - dQ(u)) \\ &\leq \frac{1}{2} h_{t-1}(x_1) + \frac{1}{2} h_{t-1}(x_2), \end{aligned}$$

where $x_1 = \int_{u \in \Omega} g_1(u) dQ(u)$ and $x_2 = \int_{u \in \Omega} g_2(u) dQ(u)$. Now, we know that $\frac{1}{2}x_1 + \frac{1}{2}x_2 = x$. Specifically, we can see that

$$\begin{aligned} x_1 &= \int_{u \in \Omega} g_1(u) dQ(u) \\ &= 2 \int_{u \in A} P_u(A) dQ(u) - \int_{u \in A} dQ(u) \\ &= 2 \int_{u \in A} (1 - P_u(\Omega \setminus A)) dQ(u) - x \\ &= x - 2 \int_{u \in A} P_u(\Omega \setminus A) dQ(u) \\ &= x - 2\Phi(A) \\ &\leq x - 2\phi x \\ &= x(1 - 2\phi). \end{aligned}$$

This implies that $x_2 \geq x(1 + 2\phi)$. Since h_{t-1} is concave, the chord from x_1 to x_2 on h_{t-1} lies below the chord from $x(1 - 2\phi)$ to $x(1 + 2\phi)$. (See Figure 1.) Therefore, $h_t(x) \leq \frac{1}{2}h_{t-1}(x(1 - 2\phi)) + \frac{1}{2}h_{t-1}(x(1 + 2\phi))$. \square

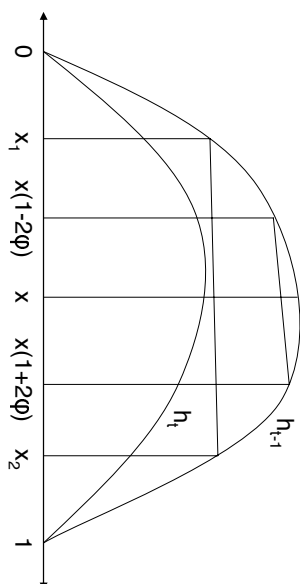


Figure 1: The value of $h_t(x)$ lies below the line between $h_{t-1}(x_1)$ and $h_{t-1}(x_2)$. Since $x_1 \leq x(1 - 2\phi) \leq x \leq x(1 + 2\phi) \leq x_2$ and h_{t-1} is concave, this implies that $h_t(x)$ lies below the line between $h_{t-1}(x(1 - 2\phi))$ and $h_{t-1}(x(1 + 2\phi))$.

Given some information about Q_0 , this lemma allows us to put bounds on the rate of convergence to the stationary distribution.

Theorem 5. *Let $0 \leq s \leq 1$ and C_0 and C_1 be such that*

$$h_0(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\}.$$

Then

$$h_t(x) \leq C_0 + C_1 \min\{\sqrt{x-s}, \sqrt{1-x-s}\} \left(1 - \frac{\phi^2}{2}\right)^t.$$

Proof. We argue by induction on t for $s = 0$. The inequality is true for $t = 0$ by our hypothesis. Now, suppose that $x \leq \frac{1}{2}$ and the inequality holds for all values less than t . From the lemma, we know that

$$\begin{aligned} h_t(x) &\leq \frac{1}{2}h_{t-1}(x(1 - 2\phi)) + \frac{1}{2}h_{t-1}(x(1 + 2\phi)) \\ &\leq C_0 + \frac{1}{2}C_1(\sqrt{x(1 - 2\phi)} + \sqrt{x(1 + 2\phi)})(1 - \frac{1}{2}\phi^2)^{t-1} \\ &= C_0 + \frac{1}{2}C_1\sqrt{x}(\sqrt{1 - 2\phi} + \sqrt{1 + 2\phi})(1 - \frac{1}{2}\phi^2)^{t-1} \\ &\leq C_0 + \frac{1}{2}C_1\sqrt{x}(1 - \frac{1}{2}\phi^2)^t, \end{aligned}$$

provided that $\sqrt{1 - 2\phi} + \sqrt{1 + 2\phi} \leq 2(1 - \frac{1}{2}\phi^2)$. The latter inequality can be verified by squaring both sides, rearranging, and squaring again. \square

Corollary 6. *If $M = \sup_{A \in \mathcal{A}} Q_0(A)/Q(A)$, then we have*

$$d(Q_t, Q) \leq \sqrt{M}(1 - \frac{1}{2}\phi^2)^t.$$

Proof. By the definition of M , we know that

$$h_0(x) \leq \min\{Mx, 1\} - x \leq \min\{Mx, 1\}.$$

Next, we will show that $\min\{Mx, 1\} \leq \sqrt{Mx}$. If $Mx = \min\{Mx, 1\}$, then $Mx \leq 1$, which implies that $Mx \leq \sqrt{Mx}$. If $1 = \min\{Mx, 1\}$, then $1 \leq Mx$, which implies that $1 \leq \sqrt{Mx} \leq Mx$. So we have shown that

$$h_0(x) \leq \min\{Mx, 1\} \leq \sqrt{Mx}.$$

Thus, by the last theorem, we know that

$$d(Q_t, Q) \leq \max_x h_t(x) \leq \sqrt{M}(1 - \frac{1}{2}\phi^2)^t.$$

□