Rapid Mixing via Coupling

1 Recap

Recall the total variation distance between two distributions $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$:

$$d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \coloneqq \max_{A \subseteq \Omega} |\boldsymbol{\mu}_1(A) - \boldsymbol{\mu}_2(A)|.$$

That is, the largest absolute difference in probability mass between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ for any set A. It's easy to show that

$$\begin{aligned} d_{TV}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) &= \frac{1}{2}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_1 \\ &= \frac{1}{2}\sum_i |\boldsymbol{\mu}_1(i) - \boldsymbol{\mu}_2(i)|. \end{aligned}$$

Let $\boldsymbol{\pi}$ be the stationary distribution. Let $\boldsymbol{\mu}_i^t$ be the distribution of states after t steps, starting at i. That is, $\boldsymbol{\mu}_i^t \coloneqq \boldsymbol{e}_i \mathbf{P}^t$ where $\boldsymbol{e}_i$ is the vector with 1 in coordinate i and 0 everywhere else.

We will develop techniques for proving when

$$d_{TV}(\boldsymbol{\pi}, \boldsymbol{\mu}_i^t) \to 0 \text{ quickly.}$$

2 Coupling

At the heart of coupling arguments is the following:

Lemma 1. *If X, Y are random variables (non-independent) with probability distributions $\boldsymbol{\mu}_X(x) = \Pr[X = x]$ and $\boldsymbol{\mu}_Y(y) = \Pr[Y = y]$ then*

$$d_{TV}(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y) \le \Pr[X \ne Y].$$

Proof. Fix a set A. Then

$$\begin{aligned} \boldsymbol{\mu}_X(A) - \boldsymbol{\mu}_Y(A) &= \Pr[X \in A] - \Pr[Y \in A] \\ &= \sum_{a \in A, y} \Pr[X = a, Y = y] - \sum_{a \in A, x} \Pr[X = x, Y = a] \\ &= \sum_{a \in A, y \notin A} \Pr[X = a, Y = y] - \sum_{a \in A, x \notin A} \Pr[X = x, y = a] \\ &\le \sum_{a \in A, y \notin A} \Pr[X = a, Y = y] \\ &\le \Pr[X \ne Y]. \end{aligned}$$

This holds for all A, so

$$d_{TV}(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y) = \max_A |\boldsymbol{\mu}_X(A) - \boldsymbol{\mu}_Y(A)| \le \Pr[X \ne Y].$$

$\square$

We can use this for an analysis technique called coupling. We can generate two copies $(X_t), (Y_t)$ of the same Markov chain, usually correlating their random choices. The pair (X_t, Y_t) is called a coupling of the chain.

Importantly, each of X and Y viewed in isolation must be a valid copy of the original chain, *i.e.*, have the correct transition probabilities.

If $(X(t), Y(t))$ is itself a Markov chain (which is not necessary, but often the case), the coupling is called Markovian. Note that this would not be the case if we correlate using information from the past.

The coupling time (random variable) of $(X(t), Y(t))$ is defined as

$$T \coloneqq \min\{t \mid X_n = Y_n \text{ for all } n > t\}.$$

That is, the first time they align and continue to stay in the same state. We will always couple chains so that once they are in the same state, they stay in the same state forever.

Corollary 2. *If $\boldsymbol{\mu}_X^t, \boldsymbol{\mu}_Y^t$ are the distributions of $X(t), Y(t)$ then*

$$\begin{aligned} d_{TV}(\boldsymbol{\mu}_X^t, \boldsymbol{\mu}_Y^t) &\le \Pr[X(t) \neq Y(t)] \\ &= \Pr[T > t]. \end{aligned}$$

We will apply this lemma when Y is already in stationary distribution $\boldsymbol{\pi}$, while X started from a worst-case state i. Define

$$d^1(t) \coloneqq \max_i d_{TV}(\boldsymbol{\mu}_i^t, \boldsymbol{\pi}).$$

Lemma 3. *d^1 is non-increasing and $2d^1$ is submultiplicative, i.e., $2d^1(t_1 + t_2) \le 2d^1(t_1) \cdot 2d^1(t_2)$.*

Proof. Triangle inequality. □

The implication is that once $d^1(t) \le 1/4$, we get exponential improvement in the distance. So we focus primarily on how long it takes until $d^1(t) \le 1/4$.

Definition 4. The coupling time of $(X(t), Y(t))$ is the smallest t such that

$$\Pr[X(t) \neq Y(t)] \le \frac{1}{4}.$$

Because after this, we get exponential improvement.

Our analysis technique is to define a coupling by suitable correlations such that after "few" steps, *i.e.*, for small t,

$$\Pr[X(t) \neq Y(t)] \le 1/4.$$

A chain is called rapidly mixing if it converges in time polylogarithmic in the number of states. To prove rapid mixing using coupling, show that for $t = \operatorname{polylog}(n)$, you have $\Pr[X(t) \neq Y(t)] \le 1/4$. By Markov's inequality, up to a constant factor, it is sufficient to analyze $\mathbb{E}[T]$.

2.1 PageRank

Given a directed graph G (web graph), perform the following Markov chain: with probability $6/7$ follow a uniformly random outgoing link from the current node; with probability $1/7$ jump to a uniformly random node. Notice this is no longer a random walk. How quickly does this converge to its stationary distribution?

Define a coupling $(X(t), Y(t))$ of two copies of this chain, and prove that it couples "quickly". First, we correlate the decision of whether to walk or jump. If they jump, jump to the same node. If they walk, choose independently unless they are at the same node, then choose the same node.

The coupling time is exactly the time of the first jump, so it is distributed geometrically with success probability $1/7$. In expectation the chain couples in 7 steps, so by Markov, after at most $4 \cdot 7 = 28$ steps, the probability of being coupled is $\ge 3/4$. A more careful analysis gives 9 steps w.p. $\ge 3/4$.

2.2 Random Walk on a d-dimensional Hypercube

To avoid periodicity, we add a self-loop at each node with probability $1/2$ of taking it; otherwise, take one of the d edges uniformly at random. What is the stationary distribution? It is uniform, because by considering the "staying put" probability as d self-loops, the graph is $2d$-regular and undirected which implies a uniform stationary distribution.

Note 5. A general technique: if you want a uniform distribution, it is sufficient to build a regular undirected graph for your random walk.

How quickly does the distribution converge? We will find a coupling. To help find a coupling, we view a step as deciding (1) whether to flip a bit in the node's ID string, and (2) which bit to flip. This is the same as choosing a uniformly random bit i and setting it to 0 or 1 uniformly at random. Our coupling is for both chains to pick the same i and the same new value for bit i.

What does it take to couple X and Y? Every index must be chosen at least once. This is a coupon collector problem! So it takes $d \ln d$ in expectation and $d \ln d + o(d)$ w.h.p. Thus, our Markov chain is rapidly mixing because $d \ln d$ is polylogarithmic in n for $n = 2^d$.

2.3 Card shuffling with pairwise transpositions

We start with an arbitrary permutation of n cards. In each step, pick two positions i_1, i_2 i.i.d. uniformly (allow $i_1 = i_2$) and swap the cards at i_1 and i_2. Is this irreducible? Yes, by selection sort. Is it aperiodic? Yes, $i_1 = i_2$ gives self-loops. The stationary distribution is uniform because every node has degree n^2 and it is undirected.

How quickly does the distribution converge? Define a coupling. Let J_t be the number of positions for which $X_i(t) = Y_i(t)$, *i.e.*, card in position i in $X(t)$ equals the card in position i of $Y(t)$. An alternative view of the chain is that we pick an index i and a card c i.i.d. uniformly at random, then swaps the card at index i with card c.

Our coupling is to pick the same i and c in both chains. Does J ever increase with this coupling? No, at least one card is in the same position afterwards, and if both involved cards were in the same position nothing changes.