

# Randomized Algorithms

## Week 7: Rapid Mixing-Coupling

Rao Kosaraju

### 7.1 Rapid Mixing

We have observed that if an ergodic Markov Chain starts in an initial distribution and performs many steps, then its distribution will approach its stationary distribution  $\pi$ . Here we want to upper bound the number of steps that need to be performed to approach  $\pi$  within a prescribed error.

There are two main methods for bounding the number of steps:

1. Coupling
  - Coupling
  - Path coupling
2. Eigenvalues
  - Eigenvalue Gap
  - Conductance
  - Canonical Paths

Here we study the coupling method.

**Definition 1.** Let  $X$  and  $Y$  be two random variables taking values  $1, 2, \dots, s$ . A coupling of  $X$  and  $Y$  is a specification of a joint distribution function of  $X$  and  $Y$ ,  $p_{XY}$ , such that the marginal distributions are  $p_X$  and  $p_Y$ , i.e.

$$p_X(i) = \sum_j p_{XY}(i, j);$$

$$p_Y(j) = \sum_i p_{XY}(i, j).$$

For example, consider two random variables,  $X$  and  $Y$ , given by their  $p_X$  and  $p_Y$  as in Table 1:

	1	2	3
$p_X$	1/4	1/3	5/12
$p_Y$	1/3	1/2	1/6

	Y = 1	Y = 2	Y = 3
X = 1	1/4	0	0
X = 2	0	1/3	0
X = 3	1/12	1/6	1/6

	Y = 1	Y = 2	Y = 3
X = 1	1/12	1/8	1/24
X = 2	1/9	1/6	1/18
X = 3	5/36	5/24	5/72

**Table 1:** Distributions of random variables X and Y

Two typical couplings are given in Table 2. Verify that the marginals are  $p_X$  and  $p_Y$ . For example, for the first coupling  $\sum_j p_{XY}(3, j) = \frac{1}{12} + \frac{1}{6} + \frac{1}{6} = \frac{5}{12} = p_X(3)$ . The first coupling maximizes  $P(X = Y)$  or equivalently minimizes  $P(X \neq Y)$ . Since for each  $i$ ,  $p_{XY}(i, i) = \min\{p_X(i), p_Y(i)\}$ , this coupling minimizes  $P(X \neq Y)$ . The second coupling makes  $X$  and  $Y$  independent.

**Table 2:** Two couplings  $p_{XY}$

**Lemma 1.** For any  $p_X$  and  $p_Y$  a coupling that minimizes  $P(X \neq Y)$  achieves  $P(X \neq Y) = \|p_X, p_Y\|_v$ .

Note that in the first coupling,  $P(X \neq Y) = \frac{1}{4}$  and  $\|p_X, p_Y\|_v = \frac{1}{2}\|p_X, p_Y\|_1 = \frac{1}{2}(\frac{1}{12} + \frac{1}{6} + \frac{3}{12}) = \frac{1}{4}$ .

For our purposes we only need  $\|p_X, p_Y\|_v \leq P(X \neq Y)$ . Our interest is in upper bounding the variation distance; but quite often it is more convenient to upper bound  $P(X \neq Y)$  and invoke this inequality.

Now, we extend the notion of the coupling of two random variables to the coupling of a Markov chain

## 7.2 Coupling of a Markov chain

Let the Markov chain be over a state space  $S$ , with the transition matrix  $P$ . Coupling is defined by running two copies of the same Markov chain such that each copy has the marginals as those of the given Markov chain i.e. if the sequence of the paired random variables of the given Markov chain is  $(X_0, Y_0), (X_1, Y_1), \dots, (X_t, Y_t)$ , then for every  $a, b, c$

$$P(X_{t+1} = b | X_t = a) = P(a, b), \text{ and}$$

$$P(Y_{t+1} = b | Y_t = a) = P(a, b).$$

In addition, the first copy is Markovian. Hence,

$$P(X_{t+1} = a_{t+1} | X_t = a_t, X_{t-1} = a_{t-1}, \dots, X_0 = a_0) = P(X_{t+1} = a_{t+1} | X_t = a_t).$$

The first copy runs independent of the second copy. That is,

$$P(X_{t+1} = a_{t+1} | X_t = a_t, Y_t = b_t, Y_{t-1} = b_{t-1}, \dots, Y_0 = b_0) = P(X_{t+1} = a_{t+1} | X_t = a_t).$$

The second copy need not be Markovian, and usually does depend upon the first copy. For example,  $Y_{t+1}$  can depend on  $Y_t, \dots, Y_0, X_t, \dots, X_0$ .

But in most applications, the coupling can be accomplished while keeping the second copy also Markovian. Then,

$$P(Y_{t+1} = a_{t+1} | Y_t = a_t, Y_{t-1} = a_{t-1}, \dots, Y_0 = a_0) = P(Y_{t+1} = a_{t+1} | Y_t = a_t).$$

**Coupling Lemma:** If there exists a coupling such that for some time  $t_0$  and for every two states  $a$  and  $b$ ,  $P(X_{t_0} \neq Y_{t_0} | X_0 = a, Y_0 = b) \leq \epsilon$ , then  $\tau(\epsilon) \leq t_0$ . That is, for any initial distribution  $q_0$  and any time  $t \geq t_0$ ,  $\|q_t, \pi\|_v \leq \epsilon$ .

**Example 1.** *Random Walk in n-dim Hypercube*

Let us go back to our n-dimensional hypercube problem. Figure 1 shows one node of a n-dimensional hypercube and its neighbors.

The random walk Markov chain chooses a coordinate  $i \in \{1, 2, \dots, n\}$  uniformly at random. Then with equal probability it sets the  $i^{th}$  bit to a 0 or a 1. Note that from any state it stays in that state with probability  $\frac{1}{2}(n \times \frac{1}{2} \times \frac{1}{n}) = \frac{1}{2}$  and complements each of the  $n$  positions with probability  $\frac{1}{2n}$ . That is,  $a_0, \dots, a_i, \dots, a_{n-1} \rightarrow a_0, \dots, \bar{a}_i, \dots, a_{n-1}$  has probability  $= \frac{1}{2n}$ . It is easy to verify that the Markov chain is ergodic and its stationary distribution is the uniform distribution, i.e.  $\pi = (\frac{1}{2^n}, \frac{1}{2^n}, \dots, \frac{1}{2^n})$ .

### 7.3 Coupling Procedure

We establish the mixing properties of the above random walk Markov chain by coupling two copies of the Markov chain.

Let  $(a_0 \dots a_{n-1})$  and  $(b_0 \dots b_{n-1})$  be the states in which the first and the second copies start. The first copy independently follows the above random walk. The second copy moves so that the corresponding bits of the two copies become equal. The second copy will mimic the first copy by changing exactly the same bit that was changed by the first copy. Typical transitions of the two copies are shown below. The underline indicates the coordinate that was selected by the first copy. For example, in the first step, the first copy chooses coordinate 2 and complements it to 0. The second copy also chooses coordinate 2 and sets it 0 irrespective of its current value. Note that for the second copy also, the probability of complementing the chosen coordinate is  $\frac{1}{2n}$ .

**First Copy:** 0100  $\rightarrow$  0000  $\rightarrow$  1000  $\rightarrow$  1000  $\rightarrow$

**Second Copy:** 1111  $\rightarrow$  1011  $\rightarrow$  1011  $\rightarrow$  1011  $\rightarrow$

Note that in this coupling both the copies are Markovian, and the second copy depends upon the first.

How long does it take for the two states to become equal? The two states will surely become equal when the first copy chooses each of the  $n$  coordinates at least once. (In fact, the first copy needs to choose only the coordinates in which the bits of the starting states differ.) The operation of the first copy is identical to the classic “coupon collector problem”.

**Example 2.** *Coupon Collector Problem*

There are  $n$  boxes, each containing a different coupon. We choose a box independently and uniformly at random and if the coupon in it was not collected before, we collect its coupon. For this problem, what is the probability that some coupon is not collected in  $t_0$  steps?

**Lemma 2.** *After  $t_0 = n \ln(\frac{n}{\epsilon})$  steps, the probability that a coupon is missed is  $\leq \epsilon$ .*

*Proof.* First fix a box  $i$ .

$$\text{Let } X_j = \begin{cases} 1 & \text{if } j^{\text{th}} \text{ choice chooses box } i \\ 0 & \text{otherwise} \end{cases}$$

Note that  $P(X_j = 1) = \frac{1}{n}$ .

$P(\text{each of the } t_0 \text{ choices misses box } i) = (1 - \frac{1}{n})^{t_0}$ . Hence, by Boole’s inequality,

$P(\text{some box is missed in each of the } t_0 \text{ choices}) \leq n(1 - \frac{1}{n})^{t_0}$ . We need  $n(1 - \frac{1}{n})^{t_0} \leq \epsilon$ . The inequality is satisfied if  $ne^{-\frac{t_0}{n}} \leq \epsilon$  ( $1 - x \leq e^{-x}$ ). That is,  $t_0 = n \ln(\frac{n}{\epsilon})$ .

□

**Example 3.** *Shuffling Cards*

In this problem, we are given a deck of  $n$  cards and we want to know, the number of “shuffling steps” needed to achieve a near random distribution. A shuffling step consists of choosing a card uniformly at random and bringing it to the top of the deck. In this example, there are  $n!$  states, one for each of the  $n!$  permutations of the  $n$  cards. Once again, the resulting Markov chain is ergodic, and its stationary distribution is the uniform distribution  $(\frac{1}{n!} \frac{1}{n!} \cdots \frac{1}{n!})$ .

We couple two copies of the Markov chain, as follows:

The first copy performs the above shuffling steps independent of the second copy. The second copy picks the same card (same element) from the second deck and brings it to the top. For example, if the king of clubs is chosen from the first deck, then the king of clubs is brought to the top of the first deck. Then the king of clubs, irrespective of its position in the deck, is brought to the top of the second deck of cards. By doing an analysis similar to the  $n$ -dim hypercube protein we can show that within  $n \ln \frac{n}{\epsilon}$  steps the two copies will be within a variation distance of  $\epsilon$ .

## 7.4 Path Coupling

We first motivate the path coupling technique by analysing the coupling of the  $n$ -dim hypercube Markov chain by a different method.

For any two addresses  $a$  and  $b$ , let  $d(a, b)$  be the number of positions in which  $a$  and  $b$  differ; e.g.  $d(011, 101) = 2$ . After the coupled Markov chains run for  $t$  steps, let the state be  $(X_t, Y_t)$ , with  $X_0 = a$  and  $Y_0 = b$ . Let the random variable  $d_t$  be  $d(X_t, Y_t)$ .

Now, we analyse  $E(d_t | d_{t-1} = k)$ . If the first copy chooses any one of the  $k$  coordinates in which the two copies  $X_{t-1}$  and  $Y_{t-1}$  differ,  $d_t$  becomes  $k - 1$ . The probability of choosing such a coordinate is  $\frac{k}{n}$ . For any other coordinate choice  $d_t$  remains at  $k$ . Hence,

$$\begin{aligned} E(d_t | d_{t-1} = k) &= \frac{k}{n}(k - 1) + \frac{n-k}{n}k \\ &= (1 - \frac{1}{n})k. \end{aligned}$$

Hence,  $E(d_t | d_{t-1}) = (1 - \frac{1}{n})d_{t-1}$ .

Taking expectations both sides,

$$E(d_t) = (1 - \frac{1}{n}) E(d_{t-1})$$

Repeatedly applying this  $t$  times results in

$$E(d_t) = (1 - \frac{1}{n})^t E(d_0).$$

Since  $d_0$  cannot be more than  $n$ ,  $E(d_t) \leq (1 - \frac{1}{n})^t n$ . Note that  $P(X_t \neq Y_t) = P(d_t \geq 1)$ .

Hence by Markov inequality,  $P(d_t \geq 1) \leq E(d_t)$ .

Hence,  $P(X_t \neq Y_t) \leq (1 - \frac{1}{n})^t$ .

This can be upper bounded by  $\epsilon$  if  $t \geq \frac{\ln(\frac{n}{\epsilon})}{\ln(1 - \frac{1}{n})}$ . Thus, it suffices if  $t \geq \frac{\ln(\frac{n}{\epsilon})}{\frac{1}{n}} = n \ln(\frac{n}{\epsilon})$ .

In this analysis, we first defined a metric  $d$  between every pair of states  $a, b \in S$  such that  $d(a, b) \geq 1$  if  $a \neq b$ . Then we specified a coupling or one step evolution of any two states. Next, we derived that  $E(d(X_t, Y_t)) < \beta * E(d(X_{t-1}, Y_{t-1}))$ , where  $\beta = (1 - \frac{1}{n}) < 1$ .

### Path Coupling:

Now, we discuss a refinement of this technique known as *Path Coupling*, which permits us to only couple a carefully chosen subset  $A$  of the pairs in  $S \times S$ . The transitive closure of  $A$  needs to be  $S \times S$ . We define a function  $d$  for every pair  $(a, b) \in A$ . Then we extend the

function  $d$  to every pair of vertices  $(x, y)$  from  $S$  as follows:

Find a shortest distance path between  $x, y$  (using  $d$  distances). Let this path be  $x = x_1 \rightarrow x_2 \rightarrow \dots \rightarrow x_{k+1} = y$ .

Then, let  $d(x, y) = \sum_{i=1}^k d(x_i, x_{i+1})$ .

**Note** that if the function  $d$  satisfies the property that for every  $(a, b) \in S \times S, d(a, b)$  is the shortest distance between  $a$  and  $b$ ; the function  $d$  over  $S \times S$  is a metric. Let the diameter,  $D$ , be  $\max_{x, y \in S} d(x, y)$ .

**Theorem 1.** *If we can couple two copies of Markov Chain such that for every  $(a, b) \in A, E[d(X_t, Y_t) | X_{t-1} = a, Y_{t-1} = b] \leq (1 - \beta) * d(a, b)$  for some  $\beta, 0 < \beta < 1$ , then the coupling can be extended to all pairs  $(c, d) \in S \times S$  such that  $E[d(X_t, Y_t) | X_{t-1} = c, Y_{t-1} = d] \leq (1 - \beta) * d(c, d)$ . Hence,  $\tau(\epsilon) = \frac{1}{\beta} * \ln \frac{D}{\epsilon}$ .*

**Application 1.** *n-dim hypercube mixing*

Define a set of pairs  $(a, b), a = (a_1, a_2, \dots, a_n), b = (b_1, b_2, \dots, b_n)$  such that  $a$  and  $b$  differ in only one coordinate i.e.  $d(a, b) = 1$ . [**Note** that if  $c$  and  $d$  differ in  $k$  coordinates then the shortest path between  $c$  &  $d$  has  $k$  edges and hence,  $d(c, d) = k$ .] Let us say that  $a$  and  $b$  differ in the coordinate  $p$ . At time  $t - 1: X_{t-1} = a, Y_{t-1} = b$ . For the first copy,  $X_{t-1}$ , our coupling procedure chooses a coordinate,  $i$ , uniformly at random and sets the  $i^{th}$  bit of  $X_{t-1}$  to 0 or 1 with equal probability. The second copy, chooses the same coordinate  $i$  and sets the  $i^{th}$  bit of  $Y_{t-1}$  to the same value.

Now, we will look at the  $(X_t, Y_t)$ .

$$d(X_t, Y_t) = \begin{cases} 0 & \text{with probability} = \frac{1}{n}, \text{ when the coordinate chosen is } p \\ 1 & \text{with probability} = (1 - \frac{1}{n}) \end{cases}$$

$$E[d(X_t, Y_t)] = \frac{1}{n} * 0 + (1 - \frac{1}{n}) * 1 = 1 - \frac{1}{n} < d(a, b). \text{ Hence, } \tau(\epsilon) = \frac{\ln \frac{D}{\epsilon}}{1 - (1 - \frac{1}{n})} = n \ln \frac{n}{\epsilon}.$$

**Application 2.** *Graph Coloring*

In this problem we want to color an undirected graph  $G$  so that no two adjacent vertices have the same color. We are given a graph  $G$  of maximum degree  $\Delta$ . Then  $G$  can be colored with  $\Delta + 1$  colors. We show that if we have  $3\Delta + 1$  colors then we can color  $G$  uniformly at random. In order to do this, first we define a Markov chain over all the possible colors of  $G$ . Let  $G$  have  $n$  vertices. Let a coloring (i.e. a state of the Markov chain) of  $G$  be  $(c_1, c_2, \dots, c_n)$ , where  $c_i$  is the color of the  $i^{th}$  vertex. Choose an  $i$  uniformly at random and choose a color,  $c$ , from the  $3\Delta + 1$  colors uniformly at random. If the  $i^{th}$  vertex can be colored with  $c$  then

make the change otherwise keep the current coloring. It can be proven that this Markov chain is ergodic and reversible, and its stationary distribution is the uniform distribution. The Markov chain is aperiodic since at each state there is a nonzero probability of staying in that state. The Markov chain is irreducible since you can reach from any coloring to any other coloring.

The coupling is done as follows:

Here  $S$  is the set of all possible colorings. Let  $A$  be  $\{(a, b) \mid a \text{ and } b \text{ differ in exactly one position, i.e. the colors of one vertex are different}\}$ . Then note that  $D \leq 2n$ .

At time  $t - 1$ , let  $X_{t-1} = (a_1, a_2, \dots, a_n)$  and  $Y_{t-1} = (b_1, b_2, \dots, b_n)$  represent two states such that the states differ in one coordinate (i.e. they differ in the color at one vertex). The evolution to  $X_t$  is as follows: Choose a coordinate  $i$  and a color  $c$ . Change the color of the  $i^{\text{th}}$  vertex to  $c$  if possible. The evolution of  $Y_{t-1}$  to  $Y_t$  is similar. The total number of possible choices for choosing a coordinate,  $i$  and a color  $c$  are  $n(3\Delta + 1)$ . Let  $(a_1, a_2, \dots, a_n)$  and  $(b_1, b_2, \dots, b_n)$  differ in the  $l^{\text{th}}$  coordinate.

1. If the chosen  $i$  is  $l$  and the color is not one of the  $\Delta$  colors adjacent to  $l$ , then the new colorings will be identical and  $d(X_t, Y_t) = 0$ . This happens in  $2\Delta + 1$  cases.
2. If the chosen  $i$  is adjacent to the coordinate  $l$  and the color chosen is  $a_l$  or  $b_l$  then  $X_t$  and  $Y_t$  can differ in the  $i^{\text{th}}$  coordinate in addition to the  $l^{\text{th}}$  coordinate. Therefore,  $d(X_t, Y_t) \leq 2$ . This happens in  $2\Delta$  cases.
3. In all other cases,  $X_t$  and  $Y_t$  will differ in one coordinate - the  $l^{\text{th}}$  coordinate i.e.  $d(X_t, Y_t) = 1$ .

Hence,  $E[d(X_t, Y_t) \mid X_{t-1} = (a_1, a_2, \dots, a_n), Y_{t-1} = (b_1, b_2, \dots, b_n)] \leq 1 - \frac{(2\Delta+1)-2\Delta}{n(3\Delta+1)} = 1 - \frac{1}{n(3\Delta+1)}$ .

Hence,  $\tau(\epsilon) \leq n(3\Delta + 1) \ln \frac{3n}{\epsilon}$ .