CSE 525: Randomized Algorithms **Spring 2023**

Lecture 16: Coupling

Lecturer: Shayan Oveis Gharan *May 18th*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Definition 16.1 (Coupling)**.** *Let μ, ν be probability distributions over Ω. A coupling between μ, ν is a probability distribution π on $\Omega \times \Omega$ that preserves the marginals of μ, ν respectively. In particular, for all $x \in \Omega$,*

$$\sum_y \pi(x, y) = \mu(x) \text{ and } \sum_y \pi(y, x) = \nu(x).$$

Let us give an example: Consider the following two distributions over $\{1, 2, 3\}$. Consider the following

	1	2	3
$\mu(.)$	0.2	0.5	0.3
$\nu(.)$	0.3	0.4	0.3

coupling $\pi(1, 1) = 0.2, \pi(2, 2) = 0.4, \pi(3, 3) = 0.3, \pi(2, 1) = 0.1$. Observe that all marginal probabilities are satisfied; for example $\pi(2, 2) + \pi(2, 1) = 0.5 = \mu(2)$. Furthermore, if (X, Y) is a sample of π, we have that $\mathbb{P}_\pi [X = Y] = 0.9$. You can compare this coupling with an independent coupling in which $\tilde{pi}(i, j) = \mu(i)\nu(j)$. In that case we would have $\mathbb{P}_{\tilde{\pi}} [X = y] = \pi(1)\nu(1) + \pi(2)\nu(2) + \pi(3)\nu(3) = 0.35$.

Lemma 16.2 (Coupling Lemma)**.** *Let μ and ν be probability distributions on Ω, and let X and Y be random variables with distributions μ and ν, respectively. Then*

1. $\mathbb{P}[X \neq Y] \geq \|\mu - \nu\|_{TV}$.
2. *There exists a coupling between μ and ν such that $\mathbb{P}[X \neq Y] = \|\mu - \nu\|_{TV}$.*

Proof. We prove the 2nd part, that is we construct the optimal coupling between μ and ν. The coupling will be very similar to the above example. For any i in the support of these distributions, we let $\pi(i, i) = \min\{\mu(i), \nu(i)\}$. For all other pairs $i \neq j$ we sequentially choose $\pi(i, j)$. Obviously, there is a way to match the remaining mass in the distributions such that all marginals are preserved. For example, when we process $i \neq j$ we define $\pi(i, j) = \min\{\mu(i) - \pi(i, .), \nu(j) - \pi(., j)\}$.

Note that obviously, this coupling has the highest probability that $X = Y$ because for any i, we must have

$$\mathbb{P}_{(X,Y)\sim\pi} [X = i, Y = i] \leq \min\{\mu(i), \nu(i)\}$$

in order to satisfy the marginal of i. Therefore,

$$\mathbb{P}[X \neq Y] = 1 - \sum_i \pi(i, i) = \sum_i \mu(i) - \min\{\mu(i), \nu(i)\} = \|\mu - \nu\|_{TV}.$$

□

16.1 Coupling for Bounding Mixing Time

In the last section we saw the coupling as a technique to prove upper bound on the mixing time. Here, we make this more formal, and we will see many examples on how to use this idea to bound the mixing time of a chain.

Definition 16.3 (Markovian Coupling)**.** *A coupling of a Markov chain is a pair process X_t, Y_t such that each process in isolation looks like an honest simulation of the chain, i.e., for all states x, y,*

$$\sum_{x',y'} \mathbb{P}\left[X_{t+1} = y, Y_{t+1} = y' | X_t = x, Y_t = x'\right] = K(x,y)$$

and similarly,

$$\sum_{x',y'} \mathbb{P}\left[Y_{t+1} = y, X_{t+1} = y' | Y_t = x, X_t = x'\right] = K(x,y),$$

and for all $t \geq 0$, if $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

We need another definition before we prove a bound on mixing time using a Markov chain coupling.

Definition 16.4 (Stopping Time)**.** *A stopping time with respect to a sequence of random variables $X_1, X_2, \ldots$ is an integer random variable T with the property that for each $t \in \{1, 2, \ldots\}$, the occurrence or non-occurrence of the event $T = t$ depends only on the values of $X_1, X_2, \ldots, X_t$.*

For example, consider a random walker on a line who starts from the origin. Consider the first time that it is at distance n away from the origin. This is a stopping time.

Lemma 16.5. *Let X_t, Y_t be a coupling of a Markov chain where $X_0 = x$ and $Y_0 \sim \pi$. Let*

$$T_{X,Y} = \min\{t : X_t = Y_t\},$$

be the stopping time until the two process meet. Then,

$$\|K^t(x,.) - \pi\|_{TV} \leq \mathbb{P}\left[T_{X,Y} > t\right].$$

Proof. The main observation is that the Markovian coupling between X, Y gives a coupling between the two random variables X_t, Y_t at any time $t \geq 1$. Therefore by the coupling lemma.

$$\|K^t(x,.) - \pi\|_{TV} \leq \mathbb{P}\left[X_t \neq Y_t\right] = \mathbb{P}\left[T_{X,Y} > t\right].$$

The last equality is by the definition of the coupling. □

In the next section we prove a bound on the mixing time of a hypercube using this idea.

16.2 Simple Random Walk on a Hypercube $\{0,1\}^n$

Consider the n dimensional hypercube with 2^n vertices, where every vertex is labelled with an n bit string. Two vertices are adjacent if their n bit strings differ in exactly one bit. Consider the following Markov chain.

i) At any vertex x, with probability $1/2$ do nothing (self-loop)

ii) Otherwise, pick a uniformly random coordinate and flip it.

Equivalently, we can consider the following Markov chain: At any vertex x choose a uniformly random coordinate i and substitute it with a uniformly random bit $b \in \{0, 1\}$.

Having the second description, considering the following coupling between X_t, Y_t: X_t and Y_t choose the same coordinate i and the same bit b.

Observe that this is a valid coupling, because each X_t, Y_t is following the same random walk. Furthermore, observe that $T_{X,Y}$ is at most the time by which each coordinate is chosen at least once. This is because once we choose a coordinate i from that moment X_t, Y_t will agree on that coordinate.

So, it is enough to find the expected time that it takes to choose each coordinate at least once, and then we can use the Markov's inequality.

This problem is known as the *coupon collector* problem: At each time step, the collector gets one out of n coupons uniformly at random. His aim is to continue till he has seen every coupon at least once. Let T_k be the time it takes to see k coupons assuming he has already seen $k-1$ coupons. Obviously $T_1 = 1$, and $\sum_{i=1}^{n} T_i$ is the time to take all coupons. It turns out that for each k, $\mathbb{E}[T_{k+1}] = \frac{1}{1-k/n}$. This is because he has already seen k coupons; so the next coupon will be new only with probability $1 - k/n$. This is a geometric random variable, so its mean is $\frac{1}{1-k/n}$. By linearity of expectation we get

$$\mathbb{E}[T_1 + \cdots + T_n] = \sum_{k=1}^{n} \frac{1}{1-(k-1)/n} = \sum_{k=1}^{n} \frac{n}{n-k+1} = nH_n.$$

In general, it is not hard to see that the coupon collector time is highly concentrated around its expectation, in the sense that

$$\mathbb{P}[T_1 + \cdots + T_n > n \ln n + cn] \leq e^{-c}.$$

We prove a weaker bound: For any i, the probability that coupon i is not collected by time $t = cn \ln n$ is at most $(1-1/n)^t \leq n^{-c}$. So, by union bound, the probability that all coupons are not collected by time $2n \ln n$ is at most $1/n$.

$$\mathbb{P}[T_1 + \cdots + T_n > cn \ln n] \leq n^{-c+1}.$$

16.3 Top-to-Random Shuffling

Recall that in the Top-to-Random shuffling: Each time we pick up the top card in a deck and place it at uniformly random location in the deck. We want to design a coupling strategy to study the mixing time of this chain.

It turns out it is much easier to design a coupling for the following *inverse* chain: Each time choose a uniformly random card and place it at the top of the deck.

We claim that both of these walks have the same mixing time. This fact is in fact true for any Markov chain on groups. In this case note that we are considering a Markov chain on the symmetric group S_n. In general consider any group G and suppose we have a set of generators $\{g_1, \ldots, g_k\}$. In each time step we choose a generator from this set according to a probability distribution μ and we apply it to the current state. The inverse chain is defined as follows: Consider the set of generators $\{g_1^{-1}, \ldots, g_k^{-1}\}$. At any state x, choose a generator g_i^{-1} from the same distribution μ and apply it to x. Observe that both of these walks are doubly stochastic, so have a uniform stationary distribution.

Now for any path in the original walk $x \circ \sigma_1 \circ \cdots \circ \sigma_t$ we can construct a path in the inverse walk

$$f(x \circ \sigma_1 \circ \cdots \circ \sigma_t) = x \circ \sigma_1^{-1} \circ \cdots \circ \sigma_t^{-1}.$$

Observe that these two paths occur with exactly the same probability. Furthermore, if two paths $x \circ \sigma$ and $x \circ \tau$ reach the same state, so does the inverse path, $f(x \circ \sigma) = f(x \circ \tau)$. This means that f defines a bijection between the elements of the group, i.e., the states, in the original walk and the states of the inverse walk. Therefore, the distribution of the original walk at time t, $K^t(x, .)$, and the distribution of the inverse walk at time t, $\tilde{K}^t(x, .)$, are the same up to relabelling the states. But since the stationary distribution of the chain is uniform,

$$\|K^t(x, .) - \pi\|_{TV} = \|\tilde{K}^t(x, .) - \pi\|_{TV}.$$

Having this tool all we need to do to study the Top-to-Random shuffling is to study the mixing time of the Random-to-Top shuffling. Consider the following coupling: At any time both chains If X_t chooses the card labelled i to move to the top, Y_t also chooses the card labelled i and moves it to the top.

Note that this is a valid coupling. In particular, the chain X_t is exactly running the original Markov chain. Y_t is also running the original chain because it chooses each of the n card uniformly at random.

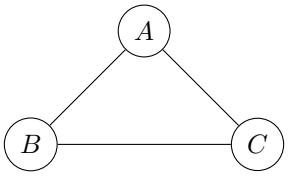
Now, let us study the time T_{XY} at which we get $X_t = Y_t$. Observe that whenever we choose the card labelled i, from now on this card will be in exactly the same location in both chains. So, we have $X_t = Y_t$ the first time by which we have chosen each card at least once. This is again the coupon collector problem. So, this chain mixes in time $n \ln n$.

16.4 Graph Coloring

Let G be a graph with maximum degree Δ. Obviously, any such graph has a coloring with $q = \Delta + 1$ colors: Greedily color every vertex with a color with is not used by any of its neighbors. There is a well-known Brooks theorem which says that any graph has a coloring with Δ colors unless it has a $\Delta + 1$ clique or $\Delta = 2$ and it has an odd cycle. For $q < \Delta$ even the decision problem is NP-hard.

So, let us focus on the case the case where $q \geq \Delta + 1$. We want to see if the Metropolis chain mixes rapidly. Note that the chain is very simple to describe: At each time we choose a uniformly random vertex and a uniformly random color and we color that vertex with that color if possible.

It turns out that if $q = \Delta + 1$, then the above chain is not irreducible as shown below: Even though the



graph has 6 proper coloring we cannot move out of the above state. In other words, the Metropolis chain is not irreducible in this case.

First, we argue that the Metropolis chain is irreducible if $q \geq \Delta + 2$. It is enough to show that we can go from any proper coloring to any fixed proper coloring, because the chain is reversible. We show that if $q \geq \Delta + 2$ Consider a vertex v and suppose it has a color c in the target coloring. If we can re-color v wi

The following is a major open problem in the field of counting/sampling:

Conjecture 16.6. *The Metropolis chain mixes in time $O(n \log n)$ if $q \geq \Delta + 2$.*

It turns out that even though there have been a huge amount of efforts we are still very far from prove the above conjecture. In this lecture and the next we will discuss several of the techniques to bound mixing time for different ranges of q.

Theorem 16.7. *For any $q \geq 4\Delta + 1$ the mixing time is $O(n \log n)$.*

Let us construct a coupling between two copies of the chain. Let X_t, Y_t be two colorings of graph G. Consider the natural metric where $d(X_t, Y_t)$ is equal to the number of vertices with different colors. We construct a coupling of X_{t+1}, Y_{t+1} such that

$$\mathbb{E}\left[d(X_{t+1}, Y_{t+1})\right] \leq \mathbb{E}\left[d(X_t, Y_t)\right]\left(1 - \frac{q - 4\Delta}{qn}\right). \tag{16.1}$$

Note that once we have the above inequality by a repeated application we can show that

$$\mathbb{E}\left[d(X_t, Y_t)\right] \leq d(X_0, Y_0)\left(1 - \frac{q - 4\Delta}{qn}\right)^t \leq e^{-\frac{(q-4\Delta)t}{qn}}.$$

Now, note that $d(X_0, Y_0) \leq n$ because in the worst case all vertices have different colors. Therefore, for $t = 8\Delta n \log n$, we have $\mathbb{E}\left[d(X_t, Y_t)\right] \leq 1/n$ which implies that $\mathbb{P}\left[X_t \neq Y_t\right] \leq 1/n$.

So, it remains to prove (16.1). We use the natural coloring, i.e., we use the same coin for both chains. Both X_t, Y_t will choose the same vertex v and try to color it with c if possible. Now, let us consider under what circumstances the distance of X_t, Y_t will decrease or increase.

- **Good Moves:** This is a move in which the distance between X_t, Y_t decreases. Suppose we choose a vertex of disagreement, say v, and we choose a color that belongs to none of the neighbors of v in X_t, Y_t. In such a case the distance decreases by 1. There are $d(X_t, Y_t)$ such vertices. Note that v has at most Δ neighbors so at most 2Δ colors are used on the neighbors of v. So, if the color c is different from these 2Δ, v will have the same color in X_{t+1}, Y_{t+1}, so $d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) - 1$. The probability of this event is $\frac{d(X_t, Y_t)(q-2\Delta)}{qn}$.
- **Bad Moves:** This is a move in which the distance between X_t, Y_t increases. Now, suppose we choose a vertex of agreement, say v and a color c. A bad event is if c is an invalid color for v in exactly one of the chains. In such a case the distance increases by 1. Observe that such an event happens if v is a neighbor of a disagreement vertex, say u, and c is the color of u in one of the two chains. The disagreement vertices have at most $\Delta d_t(X_t, Y_t)$ neighbors, and for any such neighbors there are at 2 bad colors. Therefore,

$$\mathbb{P}\left[d(X_{t+1}, Y_{t+1}) = d(X_t, Y_t) + 1\right] \leq \frac{2\Delta d_t(X_t, Y_t)}{qn}$$

- **Neutral Moves:** In any other move the distance remains invariant.

Putting these together we have,

$$\mathbb{E}\left[d(X_{t+1}, Y_{t+1}) | X_t, Y_t\right] \leq d(X_t, Y_t) - \frac{d(X_t, Y_t)(q - 2\Delta)}{qn} + \frac{d(X_t, Y_t)2\Delta}{qn}$$

Taking expectation from both sides proves (16.1).

References