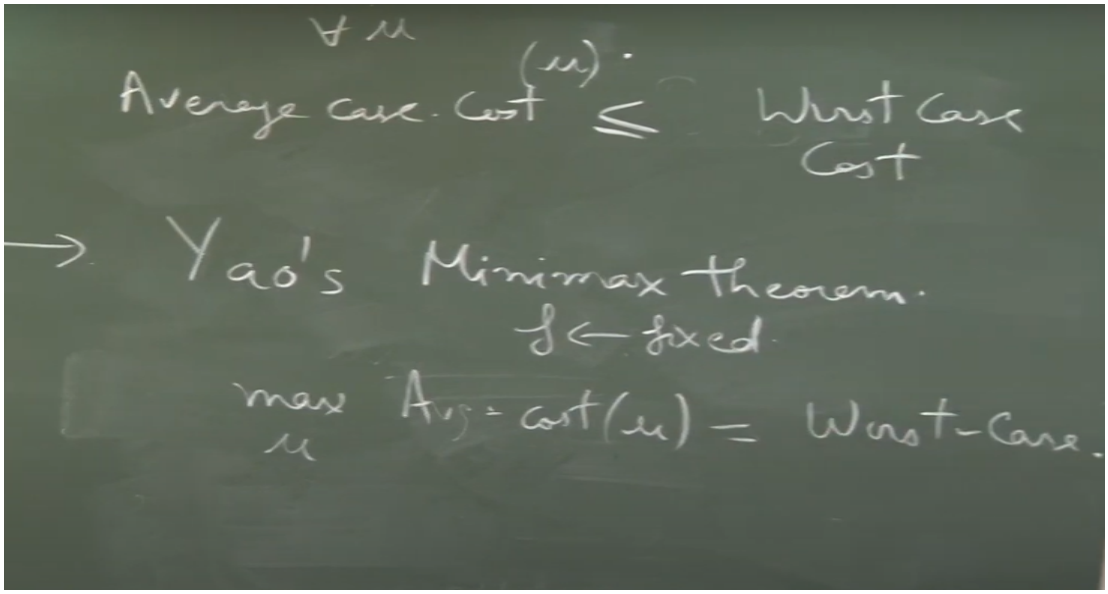
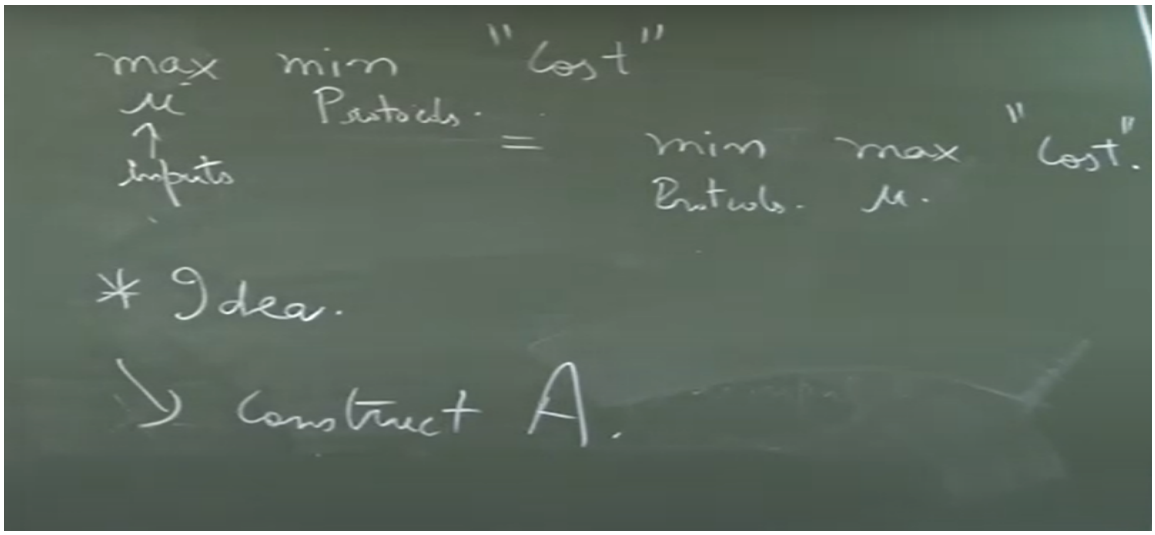
Linear Programming and its Applications to Computer Science

Prof. Rajat Mittal
Department of Computer Science and Engineering
Indian Institute Of Technology, Kanpur

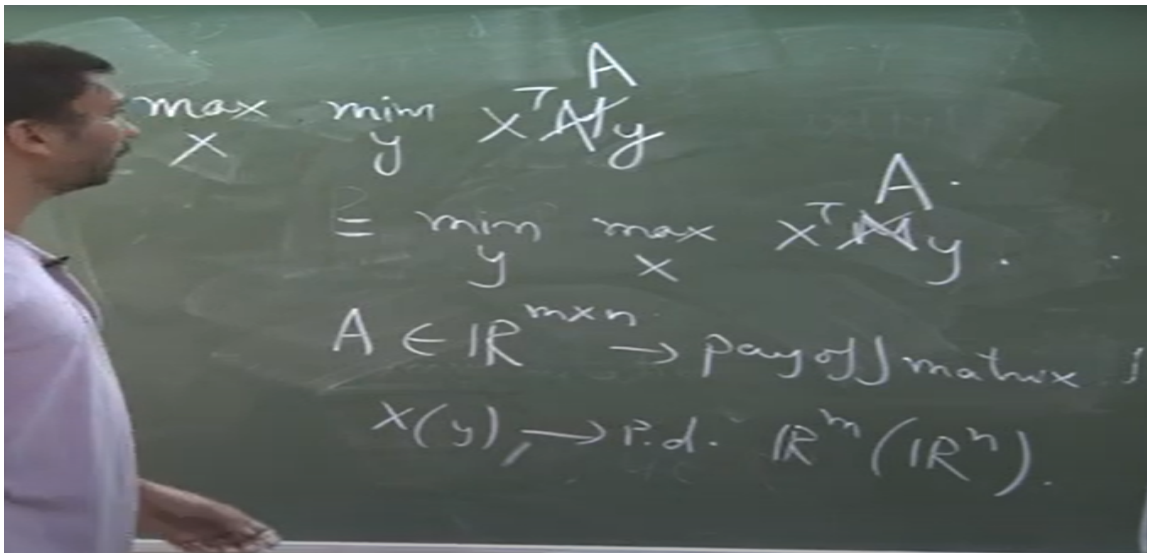
Lecture – 36
Yao's Minimax Theorem



X transpose A Y equal to Y transpose A X min X max Y max X min Y right. So, if I want to show something using Minimax theorem my task is to construct exactly right. So, to formalize this what I need to do is construct a matrix A right. And I kind of know by this idea what should be the rows and what should be the columns right. Let me write it out in any case I think it will be a good idea. Do not sleep, do not try to construct the A and see if it works.



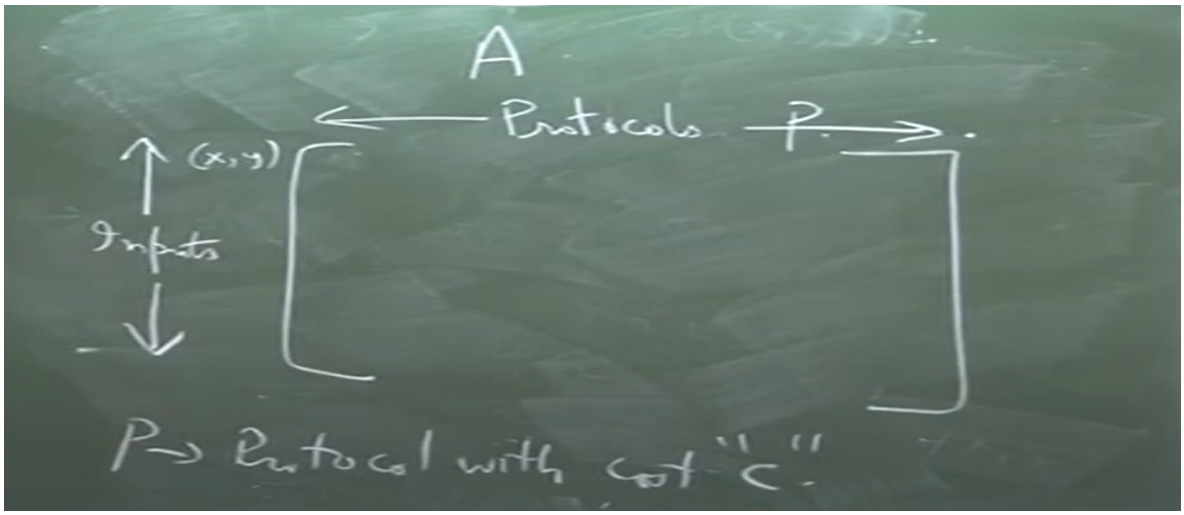
And what do I, I have changed A to M let me not do it A because I have been talking about A. Let us keep it A, A is the payoff matrix. These are probability distributions in R to the M or R to the M correct. So, I want to use this I want to make sense of A, I want to make sense of X transpose A right.



Now looking at this I should have all the inputs correct agreed. No no mu this is going to be a probability distribution over rows. There is no intersection here now this is a general communication complexity problem. So, every entry here is some pair X comma Y right.

And this is some protocol P. Ok Let us fix some cost C. So, here P when I say P it is a protocol with cost C. Now you might ask what C is? Remember average case is always less than worst case. What we are interested in is find saying that there exist a particular average case which is equal to worst case.

So, let us say C is the maximum over all possible average case protocols. But the probability distribution will have only finite support. Actually you cannot have infinite protocols because your cost is C. So, every protocol is given by the transcript. The transcript length is C every alphabet in the transcript is at most this much.



So, and you can even bound the randomness you can give a fixed amount of randomness this more than that amount of randomness. For example, more than R amount of C amount of randomness not needed you would not even be able to use it or something right. So, all that kind of things. No, C is not the worst case scenario. Again in all this discussion we have a F in mind.

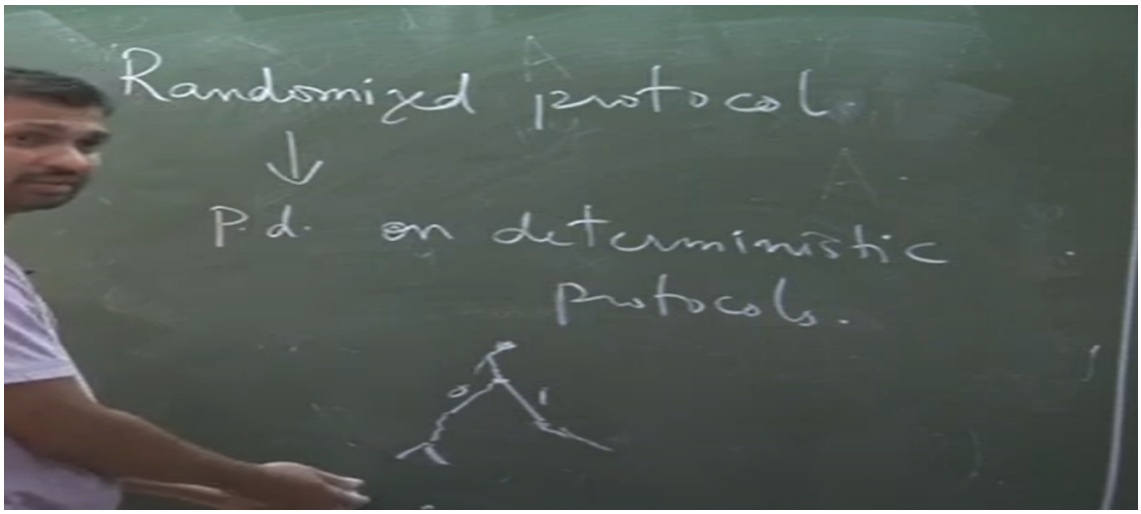
F is fixed. If you are lazy F is set to shyness right. Let us keep that F in mind, but it is true for any general F you want to prove that average case the max over average case is same as worst case. That will make our task easy like this will have a well defined. So, what do you want to do? You want to show that again we have to define a matrix.

So, that these two things like I can make sense of these two things this becomes the average cost this becomes the worst case. So, one way to do it is to say that oh let us say this thing is C this thing is C which means for any distribution mu there is a communication protocol with cost C which correctly answers with probability more than 2 by 3. And now I want to come up with a minimize protocol of cost C which works for every input that is the task I am doing. So, then I have restricted my P my these protocols to be only the ones with cost C clear. And now the next question is what is this entry? Like cost for the protocol is now C for every input right.

We have fixed that C is the upper limit even if they use less I do not care. So, I guess one important piece is missing here which I want to emphasize which is actually needed for all of this thing to work is what do I mean by a randomized protocol right. And you know this correct what is a randomized protocol? You have some randomness R you toss a coin and then you decide what to do? Do you want to send this bit or that bit? Do you want to compute addition? Do you want to multiply? So, on and so forth right this is the randomized protocol. But notice that in this case a randomized protocol is nothing, but a

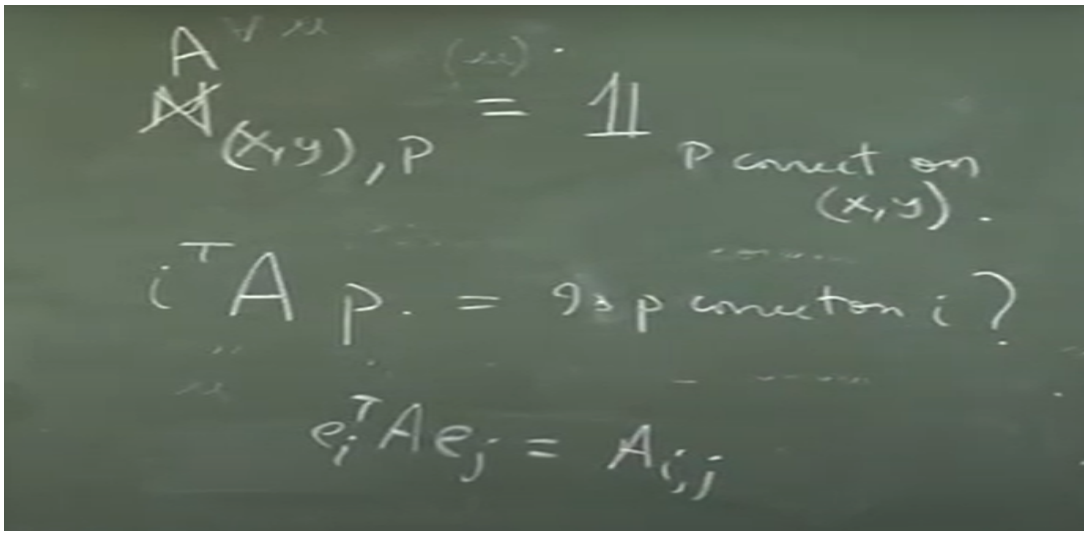
probability distribution on deterministic protocols. What I am saying is when the decision comes you do not need to toss a coin when the decision comes.

You can toss a coins beforehand. So, a way to do this would be you start you do some computation you realize oh I need to toss a coin. Now, this is 0 or 1. Now, your computation path differs again you toss a coin. What I am saying is when I look at the leaf there is some probability of doing this computation, but this leaf is basically a deterministic protocol.

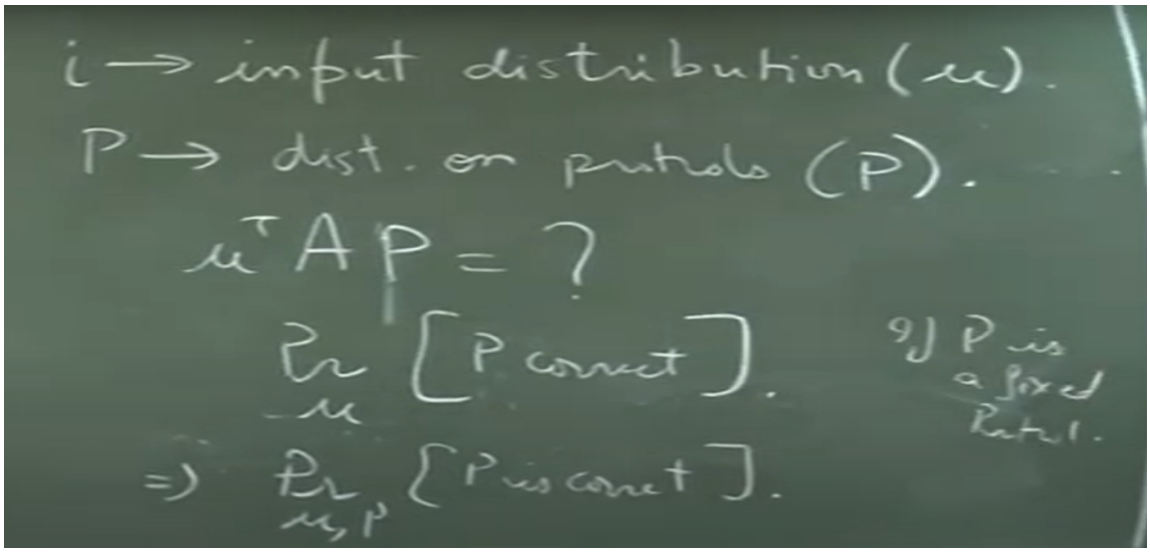


This leaf is basically a deterministic protocol. So, I can think of my this randomized protocol as a probability distribution on deterministic protocols. Other way to say it is when I talk about these protocols these are all deterministic and the randomized protocol is a probability distribution over it. Yes, so what is happening is now what the strategy of a column player is? It is going to be a probability distribution over protocols which is basically a randomized protocol. So, now the strategy of a column player is a randomized protocol and I can maximize minimize over that.

Now, with this strategy I can maximize minimize over that this knowledge can you tell me what would you like m x comma y. So, this is m i j. why I do I like m so much right the input is x comma y this is 1 of the rows x comma y is 1 of the rows p is 1 of the column. What should this be? What should we start with? p is a deterministic protocol sorry take a guess again important thing is there probability here no p is a deterministic protocol it is correct or not right that is the simplest thing. Let us start with that notice things have changed now right previously we said a protocol is correct if it works correctly on all x comma y. But now this is a set of all protocols which have cost c they might work on some x comma y they might not work on some x comma y right clear .



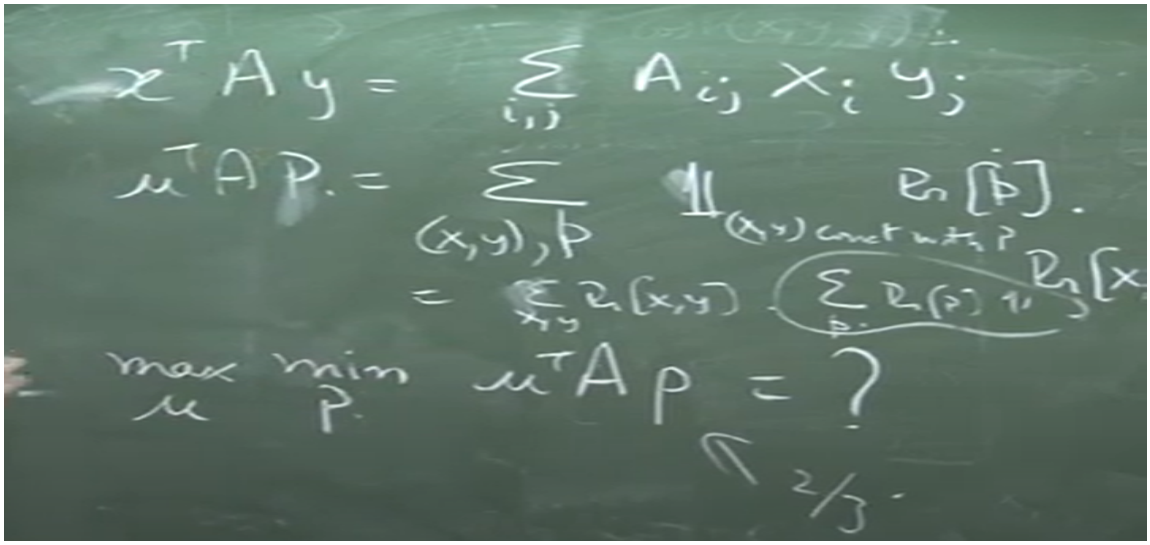
So, then if I have an input in mind I have the matrix a and I have a particular protocol in mind what is this same thing right you realize that this is a particular protocol you realize that this is the standard vectors what is this a i j or a j i correct. So, this is this is the standard vector a i j or a j i correct. So, this is this is the standard vector a i j or a j i correct. But now things are going to become interesting i is going to be a input distribution right let us call it mu and we want to change p to be distribution on protocols let us call it p right. And now I want to ask if I have a particular protocol in mind that is going to work correctly do not guess if you have a answer in mind are you convinced that is correct.



F

Let us start with p being a point protocol it is a particular deterministic protocol then what is this p is a fixed protocol mu is a distribution over inputs what does this thing give me you are in the right direction. But I think if you write it down on your pen and paper and think about it what is probability over sorry p is not a distribution p I am saying is a particular protocol. Yes, our intuition is correct, but you want to write the

complete correct thing because you want to apply mix maximum we do not want to give the incorrect proof right. So, this is exactly mu transpose AP. P is a fixed protocol, but how does it matter even if p is a probability distribution this is linear in that probability distribution correct. So, this and randomness r or p, but how does it matter even if p is a probability distribution do you agree with this do you want to spend some time with it.



If you are not convinced look at what this thing is remember x transpose A y is summation i comma j right. So, that means take out probability of p first take the summation that will give the correctness of probability of p like of a particular. So, this is small p this is big p big p is a probability distribution over small p's sounds good. See you should say ok slowly because once you say ok I am going to ask another question and if you understood the previous thing I should be able to answer this question now. That is why I said take some time and internalize this.

So, this is summation over x y probability x comma y summation over p probability p whether p is correct on x y. So, this is the probability that my randomized protocol works on x comma y. Now with linearity this becomes what is the probability that capital P works on input distribution. When I say probability over mu and P my randomized protocol P is correct p is correct what does this mean? This means summation over all my inputs my input is coming from mu what is the probability that my randomized protocol works on x comma y. Now this thing is exactly this thing, right because what is my randomized protocol I play p 1 with probability p of p 1 I play p 2 with probability p of p 2.

So, some of the probabilities of the correct protocols I get that good. So, this is the mu transpose AP is giving me the probability that p is correct what is the error probability. So, then what do I know about this very good how many bits did you use to describe your answer lot. You can describe this answer with like 2 words or 1 fraction very long answer the answer is a number. No we will that is the next step we will apply a new x

theorem after this that is why I said that you want to apply min max, but you want to set up quantities.

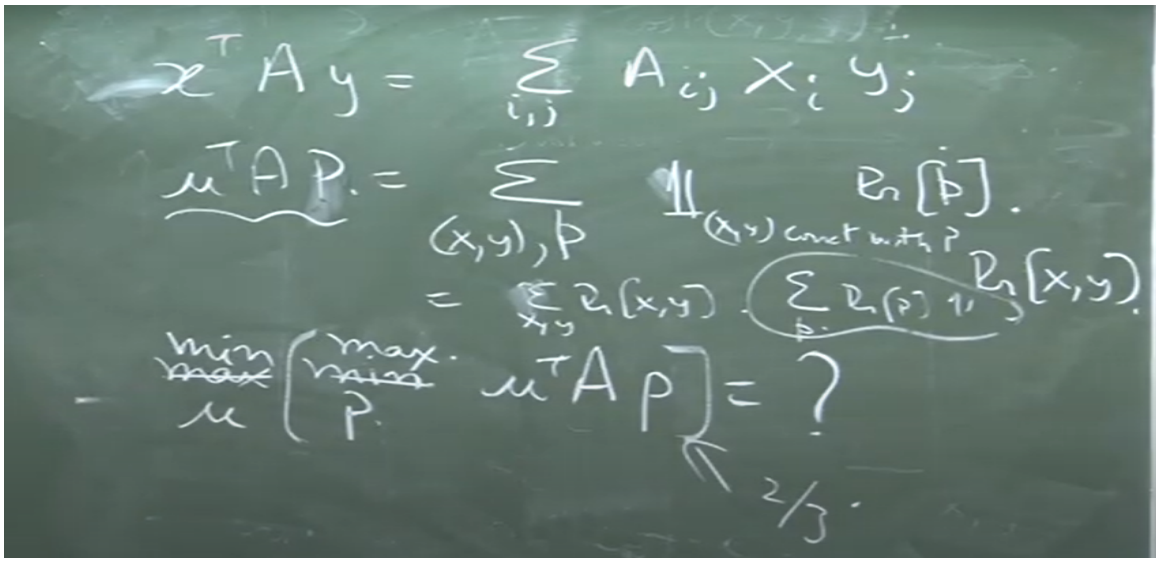
So, that everything has a nice meaning to it. Remember our A is not talking about only the protocols which have cost at most C what was C under C for any mu my error probability was at most right for every mu there exist at least 1 protocol for which this is a mu at least 2 by 3. So, if I take the max this is going to be at least 2 by 3 it will come out to be equal to 2 by 3 because we constructed this C right C was the minimum C such that everything worked out right. So, this is exactly 2 by 3 this is the x transpose M y this x transpose A y is going to be the success probability error probability, but success probability. Because I know for every mu this quantity is at least 2 by 3 for any mu there exist a protocol which succeeds with probability more than 2 by 3.

So, that means, when I take the max this quantity has to be at least 2 by 3. It depends on I how do I define C right, but even greater than equal to 2 by 3 things will work out. So, now, what do I know now I will apply Minimax theorem. So, now, what do I know right. This is the application of Minimax yes.

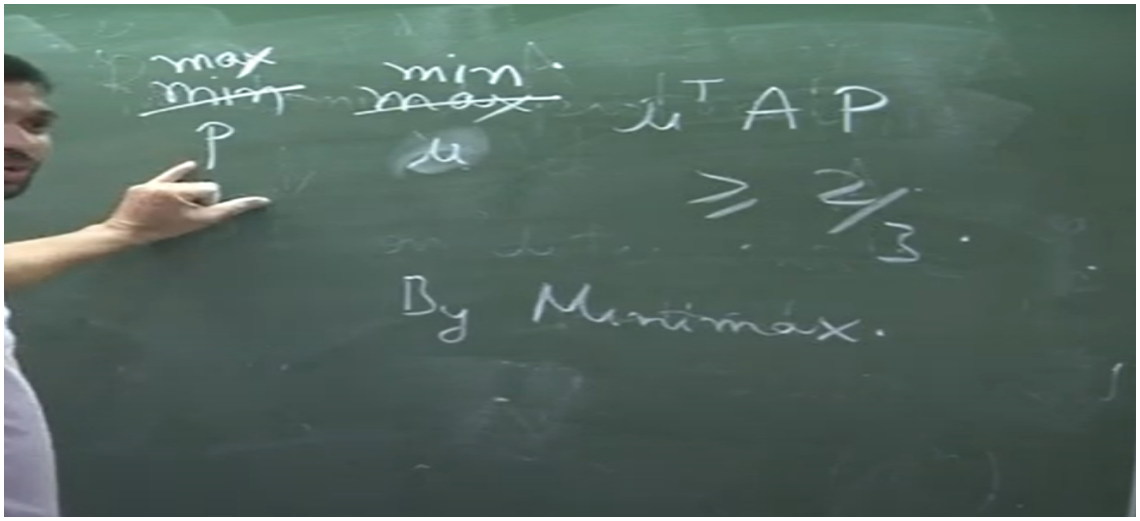
Like in the matrix. The columns are all deterministic protocols which have cost C a column corresponds to deterministic protocol which has communication cost C that is all. The correct protocols are the ones which have probability more than 2 by 3 that depends on whether I am going on average case or worst case or the worst case. In average case my correctness probability is over inputs also right that is why I wanted to write mu transpose A P. So, that I take the probability over everything right. Confusion we can write it is a lot to absorb, but these are all small ideas which make sense right.

So, once again remember the intuitive meaning of mu transpose A P right. If mu is the distribution P is now a randomized protocol right. A distribution on deterministic protocol is a randomized protocol. So, on this randomized protocol what is my success probability with respect to this distribution.

This is what I am capturing correct. Now, by the definition of average case complexity I know that for any mu. There exist a protocol with cost less than C which works correctly, which works correctly means exactly means that there exist a randomized protocol such that this quantity is at least 2 by 3 success probability is 2 by 3. This is the definition of being correct in average case. Yeah Oh Did I exchange error probability with let me just.



Yeah. You are complete right I think I have just exchange the success probability with failure probability or max with yeah sorry yeah you are right. So, what should I do is yeah good you pointed out very nice right. So, what happens is again I was using weaker I was saying everything is bigger than 2 by 3 then max is bigger than 2 by 3 that is a weaker thing right. What I know is that if everything is bigger than 2 by 3 then even minimum is bigger than 2 by 3. My mistake look at the correct thing now right min over mu max over p mu transpose AP right this is the success probability.

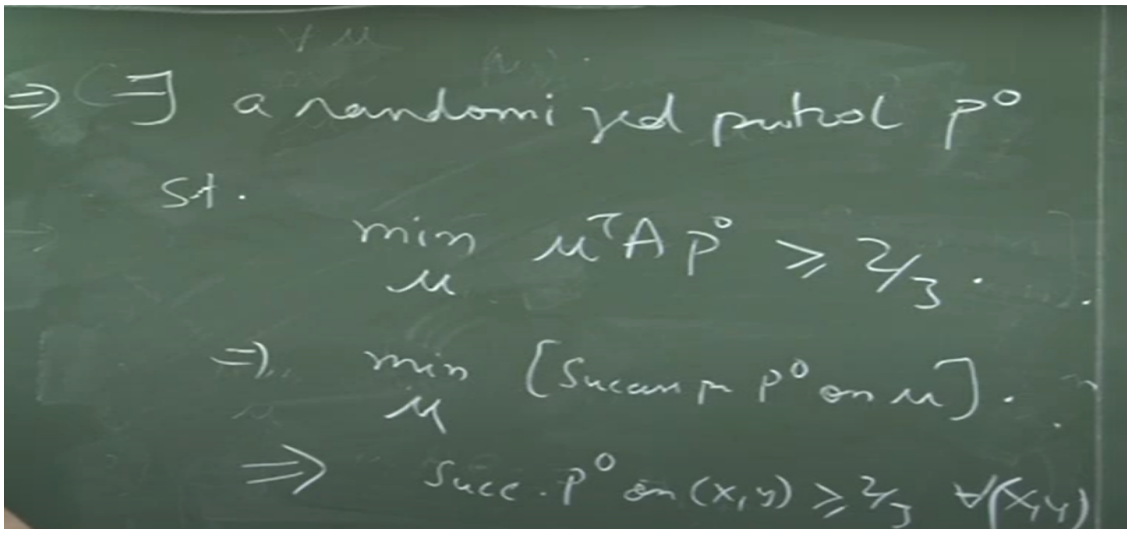


So, fix a distribution mu I know that there is a randomized protocol whose success probability is 2 by 3. So, that means the max over the protocols will also be at least 2 by 3 now it is. No, no, no, no, no the remember how did I define C. So, C was max over mu average case cost for mu. So, what does it mean for any mu there exist a protocol with cost C which works correctly on mu right this we will remember other way to say it is you fix a mu there is a randomized protocol such that the success probability is at least 2 by 3 is more than 2 by 3 right there is a protocol which is correct.

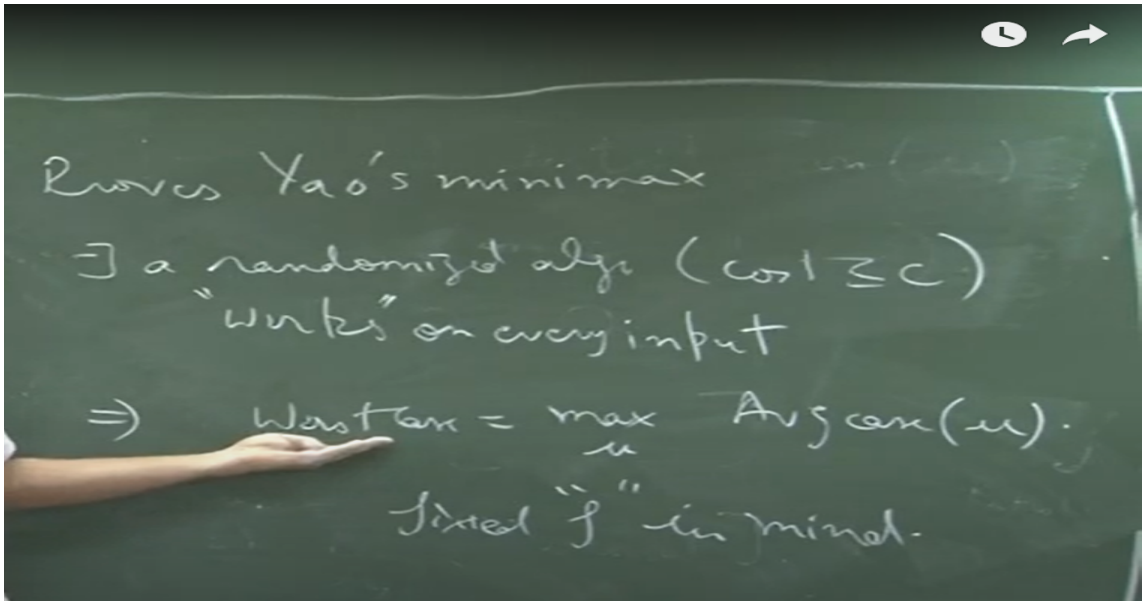
So, since this cost is maximum fixing a mu there is always a communication protocol which is correct on mu correct on mu means success probability is more than 2 by 3 right. With this definition in mind now if I maximize over p these are all randomized protocols with cost C and clearly there is a protocol here which works correctly on mu correctly on mu means this quantity is more than 2 by 3 right. So, since this quantity is more than 2 by 3 for every mu even the minimum is more than 2 by 3 good I think. So, I was using the weaker version here obviously then I would not get things right I was saying that oh everything here is bigger than 2 by 3. So, the maximum is bigger than 2 by 3 that is a very weak statement to make what I want to say is that everything is bigger than 2 by 3.

So, even minimum is bigger than 2 by 3 that makes much more sense. So, then I get that this is bigger than 2 by 3, but this also means that max this is more than 2 by 3 good I would have got unstuck here if I had taken the opposite definition I would have gotten stuck here right by min max is that we are good right. So, I was wrong before when I was writing max min with min max you get the exact correct interpretation right. And you can do min max I think you just have to change this with failure probability or something and then things will work out I think not sure, but I think that is. So, this gives me this expression is bigger than 2 by 3 right and again what am I doing I am maximizing over all randomized protocols correct the success probability right.

So, this will give me the best success probability within cost c right this implies there exist a randomized protocol call it P 0 such that correct I should say P 0. Now, what does it mean this is the success probability of P 0 on in this case on mu right this is the definition this is how I understand this quantity correct. And now I can take mu to be any probability distribution in effect I can take it to be the point distributions. So, mu could be probability 1 I choose a input pair x comma y then this is saying that my P 0 this these are same statements if you fix a protocol success probability on all x comma y is greater than equal to 2 by 3 is same as saying success probability on any distribution over inputs is greater than equal to 2 by 3 those 2 are the same we have done this thing 1000 times in linear programming term the optimal is achieve at the vertices that is all. Vertices are the point distributions. When you take the all possible convex combinations you get the convex amount of it this is the same phenomena which is appearing many times.



And I know many of you are looking at your watch, but be patient few more minutes right. So, this tells us there exist a worst case algorithm which works with probability more than 2 by 3 which has cost at most c right. So, this proves Yau's Minimax there exist a randomized algorithm cost less than c works on every input again we know what works means works means probability is correct with more than 2 by 3. This implies worst case is equal to so over mu case complexity over mu again this is the Minimax theorem, but again when I write this I have a fixed F n minus 1 right. Now, this might seem like a nice Mathematical theorem which says average case surprisingly average case complexity is same as worst case correct, but the main important thing is using this you can give a lower bound.



How do I give a lower bound? I am generally interested in let us say worst case complexity of set destroyness and this is actually the way in which the first lower bound for worst case complexity came about right. If I want to give a worst case bound I can show that if I show that there exist a mu such that I have to make sure that I do it

correctly right for communication cost. So, suppose I find these are called hard distributions. So, I have an input in my input distribution in mind if I show that for that hard distribution this is less than 2 by 3. That means the worst case complexity is more than C right because this is same that this is not going to be 2 by 3 this is going to be smaller than 2 by 3.

If I restrict it to cost C there is going to be less than 2 by 3. That means this is less than 2 by 3 there is no good randomize algorithm which works for all the inputs. So, what is this saying? This is saying that the average case complexity is not C right. If the worst if the average case complexity max over mu is not C then the worst case complexity has to be higher than C. This is saying that there is no protocol of cost C which works for input distribution mu correct.

This statement is telling you any communication protocol which has cost less than C its success probability is less than 2 by 3 does not work for distribution mu. That means average case complexity is higher than C that means worst case complexity is also higher than C. And now even proving this is very easy once again because now I do not have to worry about P being a probability distribution. I can reduce it to max over all deterministic protocols. Success probability is less than 2 if I can find a hard distribution such that all deterministic protocols fail on it I am done all deterministic protocols of cost C fail on it I am done.
I know that worst case also will be higher than 2 by 3. So, and this constructing this hard distribution is actually a very nice very creative endeavor. It is called a hard distribution because you kind of come up with X comma Y pairs which are hard to distinguish. So, I will leave you just by one last question what do you think what kind of input pairs are hard to distinguish for the set disjointness problem one element in common. So, you take X comma Y such that X comma Y are completely disjoint or they have just one element in common these are in some sense. So, you look at these kind of things put a probability decision over it and that is what actually did and prove the set disjointness problem.

This was one of the papers in your list of projects, but nobody picked it nobody picked it right. Let us see if I have time I will present that, but these are lot of proofs work like this I have just given you Minimax and randomized communication complexity. There are randomized algorithm there are many many randomized settings all of which benefit from Yao's Minimax.