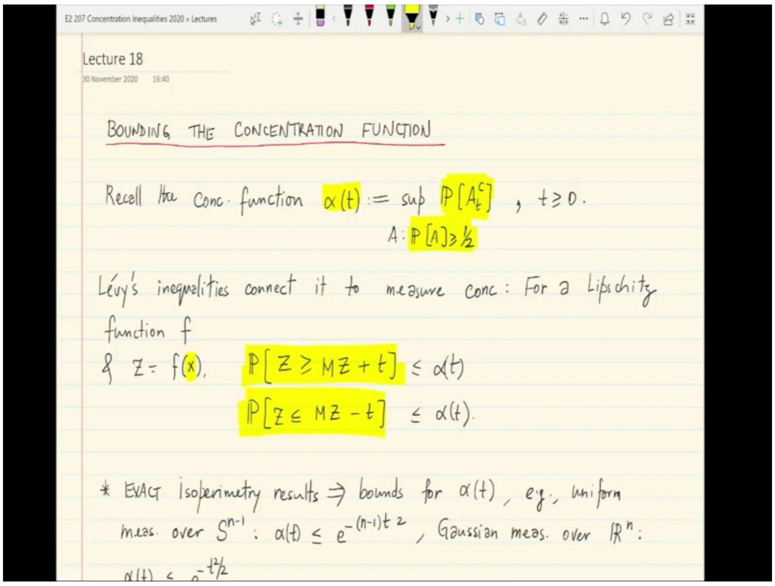
Concentration Inequalities

Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 19
Isoperimetry and bounded difference

(Refer Slide Time: 00:20)



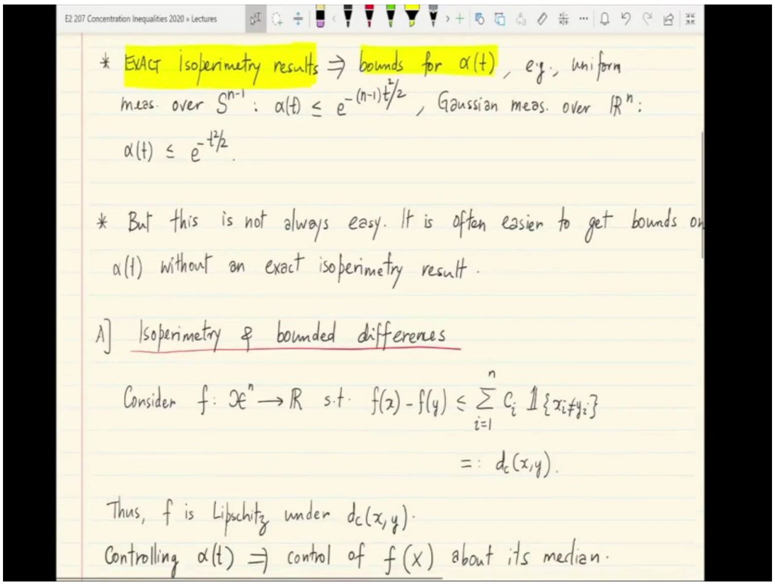
Hi all. Today we will continue understanding the connection between isoperimetry and concentration of measure. So, last time we showed that there was a certain concentration function that we defined which was given by in any probability space endowed with a metric.

We try to maximize the probability of the complement of a t blow up of any set A, which has significant probability measure namely probability measure half.

So, we basically saw that if we can control the concentration function alpha of t, then it actually serves as a very convenient upper bound to the probability that Lipschitz functions f of a bunch of random variables on that probability space concentrate around their medians ok.

So, it gives the alpha of t concentration function can give bounds on the deviation probabilities of Lipschitz functions about their medians by an amount of t.

(Refer Slide Time: 01:31)



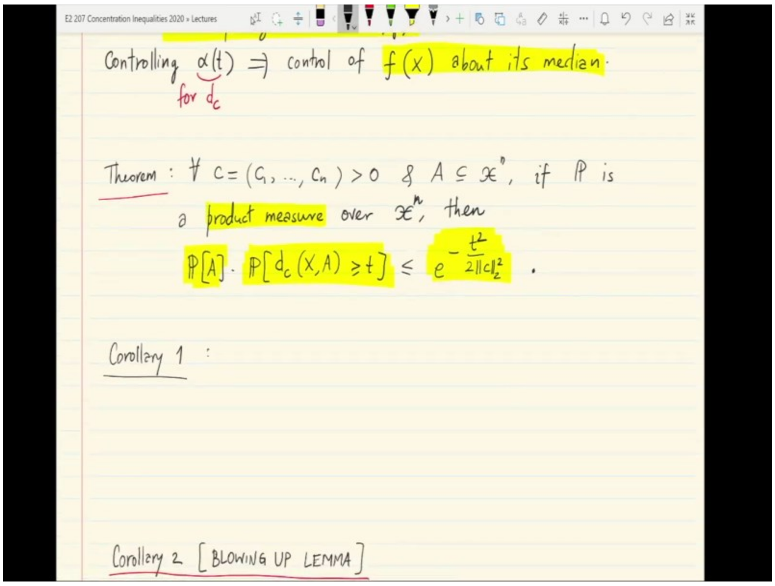
We also saw that bounds on alpha of t can be obtained by leveraging whenever possible exact isoperimetry results which basically tell you what are the optimal shapes with respect to achieving best possible surface area for a given volume or vice versa.

So, exact isoperimetry results imply bounds which are often explicitly computable for the concentration function alpha of t. So, for example, for the uniform measure or probability distribution over the surface of the unit sphere S n - 1. We saw that you could get a bound of alpha of t less than equal to e to the - n - 1 t square by 2. And likewise, for Gaussian measure in n dimensions, we saw that alpha of t was upper bounded by e to the - t square by 2.

But, the unfortunate fact is that bounding alpha of t using exact isoperimetry results is not always easy. It is; however, easier to get bounds on alpha of t without exact isoperimetry results and we will exhibit some of these techniques here.

So, the next section that we will start with is called isoperimetry and bounded differences. And this is about understanding how to prove or control the concentration function by exploiting boundary differences properties.

(Refer Slide Time: 03:01)



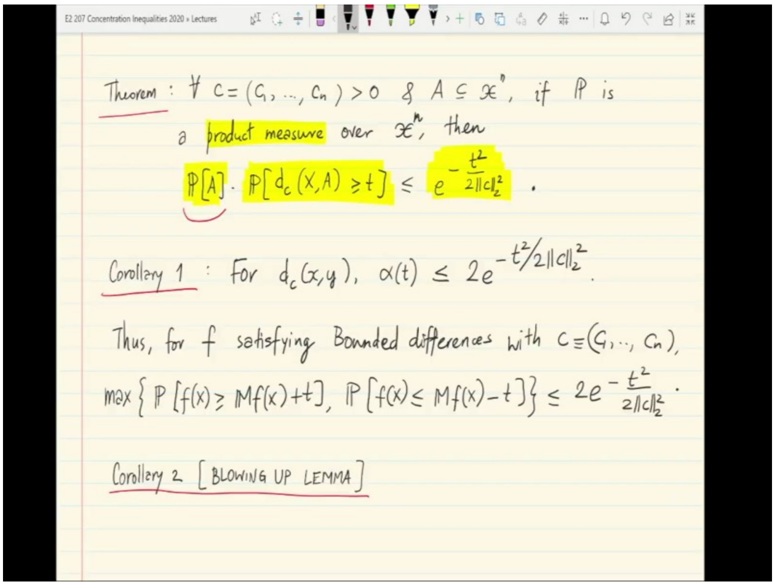
So, for the purposes of this section, it is convenient to consider a function f which has a bounded differences property defined on a space X n in that f x - f(y) is at most the sum of C i * indicator x i not equal to y i.

We also called this the C weighted hamming distance and in fact, this bounded differences property implies that f is Lipschitz under d c x y. So, this means that if we can control alpha of t for the distance measure d c then, this will imply control of the deviations of f(x) around its median by Levy's inequalities.

So, here is a representative result in this direction. It says that given a vector of bounded difference coefficients C 1 through C n and a set A any set A subset X A of this metric probability space endowed with a product measure.

So, we assume that x 1 through x n are independent random variables not necessarily identically distributed. Then we have this inequality that says that the probability of the set * the probability of deviating in C weighted hamming distance from that set by a quantity at least t is at most e to the - t square by 2 l 2 norm of c the whole square ok.

(Refer Slide Time: 04:41)



And if we assume this theorem, it has some interesting corollaries. So, the first corollary here, is what happens when you have d c x y as the distance measure. It basically means that alpha of t which was defined to be the supremum of the blow-up of supremum of the probability of the complement of the blow up of alpha by t of A by t when probability of A is at least half.

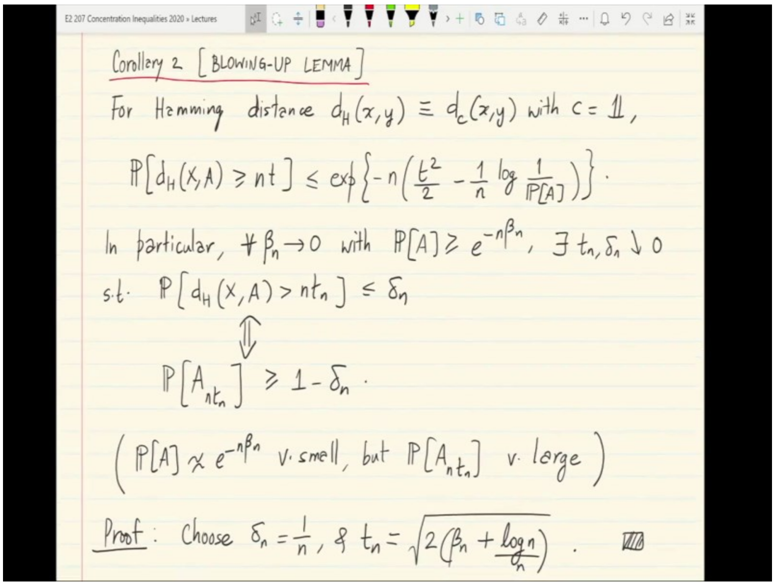
So, here, this probability of A is at least half. So, you can use the theorem to get the further lower bound on the left-hand side of half * this probability here. And then, move the half over to the other side to immediately get alpha of t as being upper bounded by twice e raised to - t square by 2 norm c the whole square.

So, for the C weighted hamming distance metric on a probability measure space with independent measure, we already have using the theorem that we have an upper bound on alpha of t of this order. And this immediately implies further that for f satisfying for a function f which satisfies bounded differences with coefficients C.

We have median concentration we have that the max of both these quantities Z exceeding. So, let me just say f(x) exceeding M f(x) which is any median + t and as well as f(x) being at most M f(x) - t; these probabilities are both upper bounded by twice e to the - t square by 2 c 1 twice the sum of the squares of the c i's.

So, this is the first corollary. The second corollary is something familiar to people in information theory which applies this theorem to a special weighted hamming distance which is the classical hamming distance.

(Refer Slide Time: 07:24)



So, for hamming distance d H x y which is just d c x y with C equal to all 1s. We have that the probability of d H X A exceeding n t. So, this.

So, we move the probability of A term all the way to the other end and then take exponent and log to get the following, e raised to - n t square by 2 - 1 over n log 1 by the probability of A ok. So, this is the blowing up lemma is basically a corollary of this theorem when you specialize to the specific d c function, where all the coefficients are the same. And a further consequence of this is the more useful part of this lemma in information theory.

Which says that information and coding theory in fact, it says that if you are given a sequence of functions given a sequence of numbers beta n that decay to 0 then one can construct sequences t n and delta n that go to 0 as well; such that the following holds sorry.

So, I should probably be careful. So, supposing that beta n is a sequence going to 0 with probability. So, if there is a set A for which the probability of A is at least e raised to - n * beta n with beta n going to 0.

Then there exists you can also find sequences t n and delta n going to 0 such that; the probability that d H; the hamming distance of X from A exceeds n t n is at most delta n or equivalently the probability. You can state this equivalently as the probability of the n t n blow up of A in our notation A n t n has exceedingly large measure.

Recall that delta n goes to 0. So, this is basically saying that. So, in words what does this lemma mean? In words this blowing up lemma says that P of A. So, imagine that there is a sequence beta n going to 0 that makes probability of A have a probability that can actually go to 0, but at rate which is slower than exponential.

So, P of A is about e to the - n beta n which is very small. A can have a very small probability, but the probability of A enlarged by n t n is suddenly extremely large. It is very close to 1. So, by blowing up a very small set, you can get to by a very small distance amount which is n t n where t n also goes to 0.

So, you can by blowing up A by expanding A to cover a distance extra distance of n * t n which is vanishingly small compared to n the maximum hamming distance in the set, you could actually inflate the probability or blow-up probability from very low almost 0 to almost 1 ok.

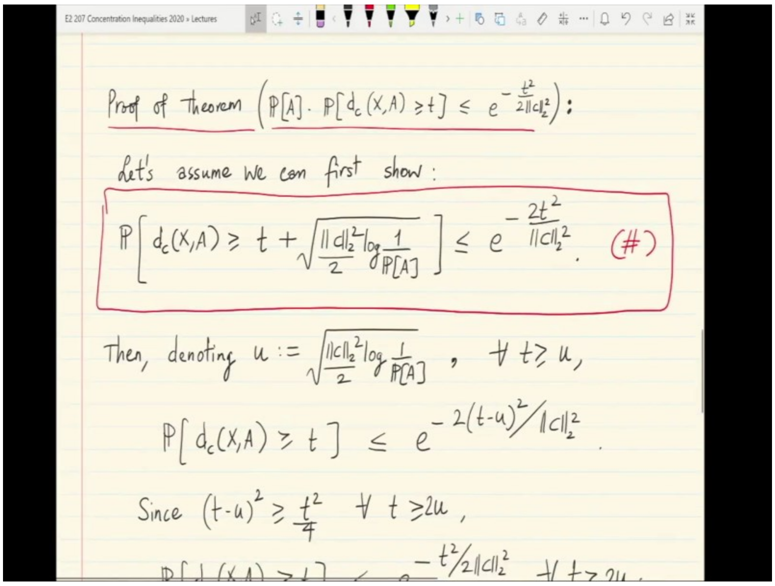
So, what is the proof of this second part? Well, the second part is just obtained by taking. So, you can do the calculations you could basically take choose t n. So, choose delta n as 1 by n and t n the required t n that gives you this is square root twice beta n + log n by n.

You can check that this choice will basically give you what you want. Notice that n t n is essentially of order square root n roughly order square root n. So, it says that if you blow up a very small set, in distance by order about square root n then, you basically get extremely large probability coverage.

And this property is heavily exploited in information theory. Particularly, for showing what are called strong converses and other coding theorems ok. So, these are two immediate corollaries of this result of this theorem that basically bounds the product of the probability of A into the probability of the complement of a t blow-up of A by e raised to - t square by 2 norm c the whole square.

So, let us in the remainder of this lecture, you will go ahead and prove this theorem. Recall that we want to show the following inequality, where the measure P is a product measure in the metric in the overall metric probability space.

(Refer Slide Time: 13:34)

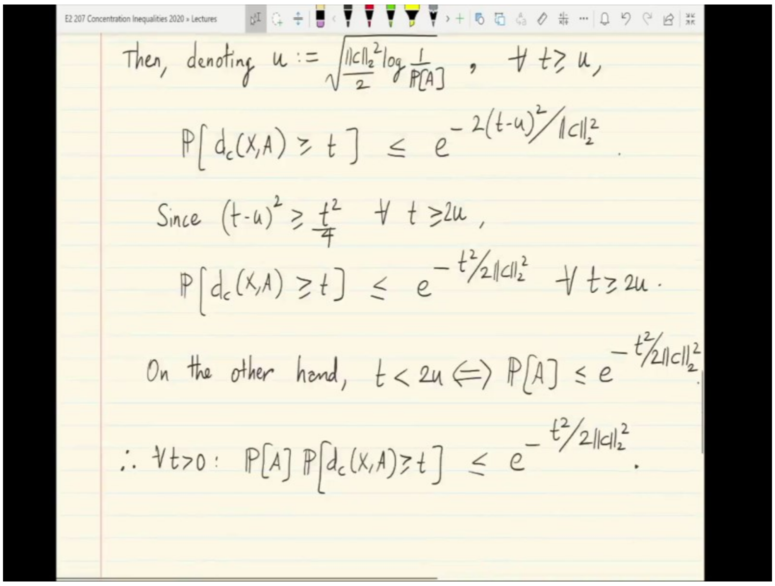


So, towards this let us first show an intermediate result. Let us assume we can show the following intermediate result that the probability of d c X A exceeding t + square root norm c square by 2 log 1 over P A is no more than e to the - 2 t square by norm c square.

So, let us assume that we have somehow shown this result that we will call star. Now, assuming that star is proven, what we can do to complete the proof of the theorem is we first denote. So, denoting u by this quantity the square root c norm c square by 2 log 1 by P A.

We have that whenever t is at least u in the above. So, we can write probability d c X A exceeding t. So, we add and subtract u from this t here t - u + u. So, we write this as sorry P + u - t sorry add and subtract u.

(Refer Slide Time: 15:20)



And we have e to the - 2 t - u the whole square by norm c square. That is what the equation that is what the inequality labeled by hash gives you. And you can also show a simple lower bound that t - u the whole square is at least t square by 4 for all t larger than 2 u.

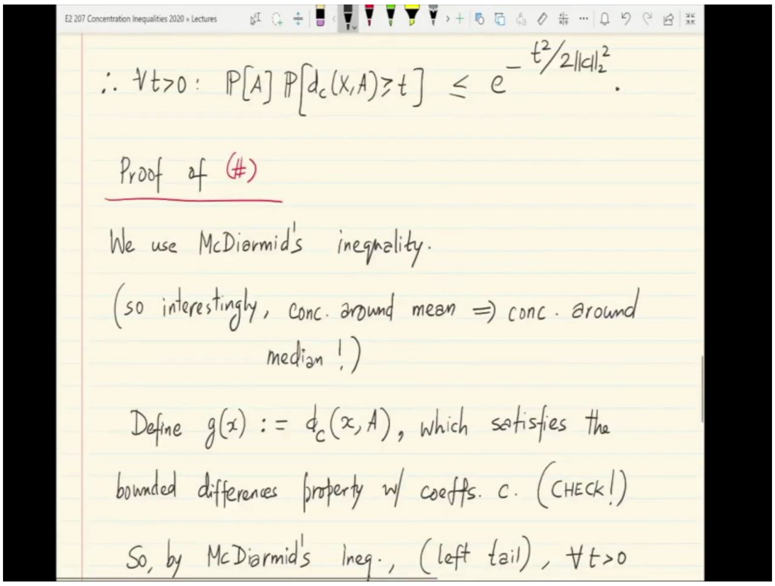
So, with this we will have that probability of d c X A exceeding t is upper bounded by e raised to - t square by 2 norm c square, the 2 has gone into the denominator in the exponent just because of a division by 4 and this is true for all t larger than 2 u.

Now, on the other hand, if t is less than 2 u then by the definition of u here, this is equivalent to the fact that the probability of A is at most e to the - t square by 2 norm c square. So, regardless of whether t is larger than 2 u or less than 2 u, for all t greater than 0, if you multiply these two probabilities out.

So, when t is larger than equal to 2 u, the second probability is bounded when is very small. When t is less than 2 u, the first one is very small. So, you can bound the other probability by 1 in each case. And hence, you get e raised to - t square by 2 c square.

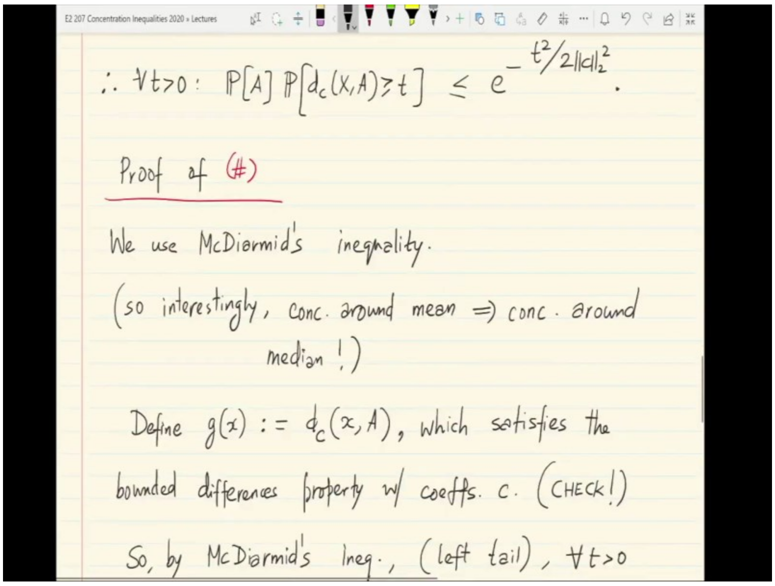
And so, that completes the proof of the theorem modulo this inequality star. Now, all that remains is to prove the inequality star for the C hamming distance C weighted hamming distance.

(Refer Slide Time: 17:33)



So, a proof of inequality star. So, how do we do this? In fact, we will accomplish the proof of star using McDiarmid inequality, which is essentially a statement about concentration about the mean. So, interestingly by the corollary of the mean theorem, concentration around the mean finally, has its consequence as one of its consequences as concentration around the median ok which is a nice fact. So, how do we use McDiarmid's inequality?

(Refer Slide Time: 18:33)



Let us first define the quantity. So, we are given a set A in the theorem in this inequality that we have to prove. So, let us define g of x; just as we did with the converse to Levy's inequality as the distance the c distance of x from the set A which in fact, you can check satisfies the bounded difference property with coefficients C ok. This is something that you can easily check the proof relies on using the triangle inequality for matrix.

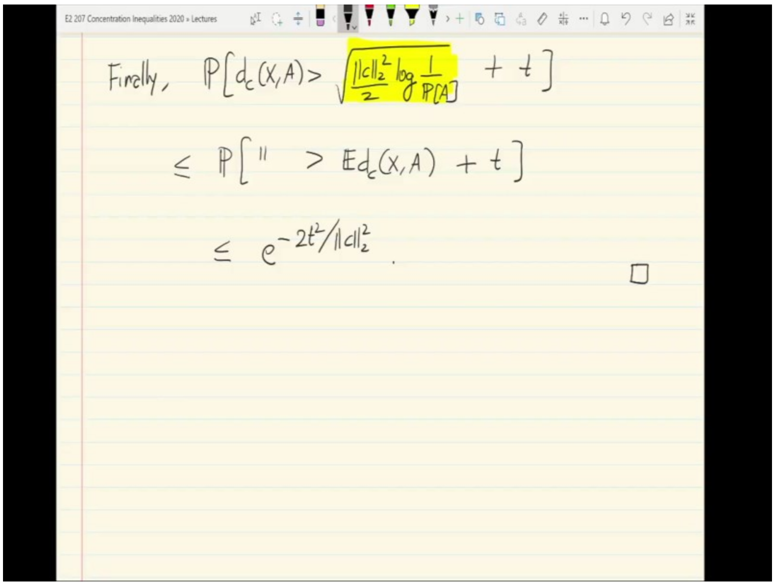
So, with this so, by McDiarmid's inequality applied to the g function for the left tail, we have that the probability that d c X A is less than its own mean - t. So, this is for all t greater than 0 is at most e raised to - 2 t square by norm c the whole square.

So, one can take in particular t as exactly equal to E d c X A expected value of d c X A, which gives you in particular that the probability that d c X A less than equal to 0 which by definition is just P A is bounded by McDiarmid by e to the - 2 twice expected value of d c X A whole square by norm c square.

This is the same as saying that the expected value of d c X A if you invert this expression is at most the square root of c square by 2 log 1 by P of A ok. And now, we can use the other direction using the right tail side of McDiarmid for the same function g.

You basically get that the probability of d c X A exceeding its expected value by an amount t is upper bounded by e to the - 2 t square by norm c square and finally, if we combine both of these results.

(Refer Slide Time: 22:00)



So, finally, we get that the so, the desired probability in star on the left-hand side is the probability that d c X A greater than square root c square by 2 log 1 by P A + t.

So, we know that this quantity here, is already an upper bound to expected value d c X A. So, we have that the probability is upper bounded by whatever you get when you replace the square root term by E d c X A. And the right tail bound is exactly what gives us the bound here from McDiarmid and that completes the proof of this theorem.

So, basically, we have understood the connection between what bounded differences functions mean and how functions of bounded differences; bounded differences functions of independent random variables enjoy both mean concentration via simple McDiarmid, as well as median concentration thanks to bounds on the concentration function and Levy's inequality. So, that finishes this lecture.

Thank you.