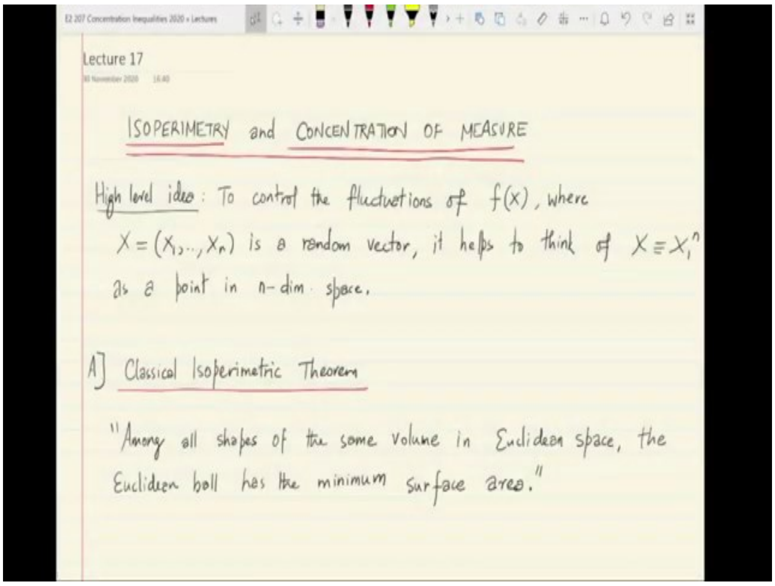
Concentration Inequalities
Prof. Aditya Gopalan
Prof. Himanshu Tyagi
Department of Electrical Communication Engineering
Indian Institute of Science, Bengaluru

Lecture - 18
Isoperimetry and Concentration of Measure

(Refer Slide Time: 00:21)



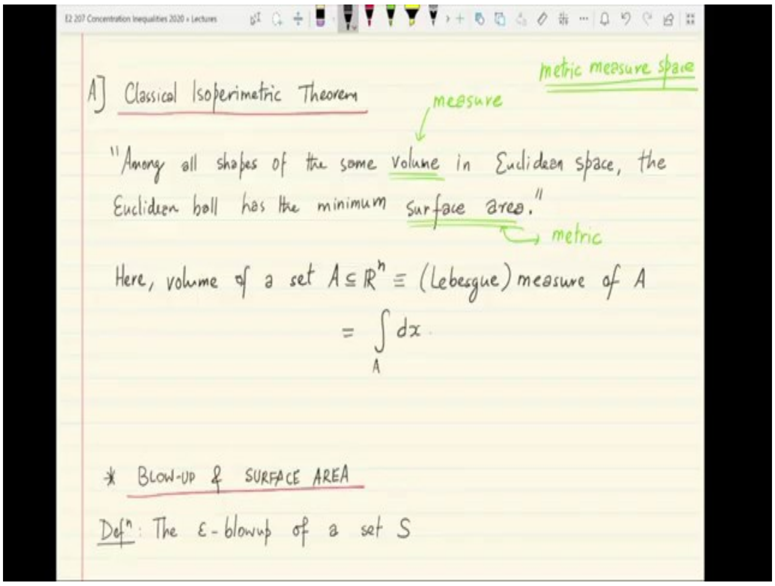
Hi all today we would like to investigate a connection between Concentration of Measure or probability and what is called Isoperimetry, which is a very classical subject in geometry. So, this gives us a new geometric viewpoint on how to understand and approach concentration of measured results and in fact we will find out that results in isoperimetry, actually imply concentration of measure results.

So, what is the high level idea behind this approach, the high level idea is that in order to control in probability the fluctuations of some function of n random variables by X where X given by X 1 through X n is a random vector it helps to view X, where X is given by X 1 to X n is a random vector it helps to view X as a point in n dimensions or in some convenient n dimensional space. So, that if we understand where this point is distributed in n dimensional space using results from geometry and isoperimetry.

Then we will have some idea about how to control the fluctuations of the function f applied to X ok. So, what is firstly the isoperimetry view here? So, here is an example of a Classical Isoperimetric Theorem ok which is perhaps a statement that you would have seen during the course of high school or college which is regarded as well known fact by now. So, it says for instance that among all shapes of the same volume in Euclidean space, let us think of 3 dimensional Euclidean space as usual.

So, if you are given a target volume then if you want to find the shape which minimizes the surface area with a given volume, then it has to be a sphere or the Euclidean ball more generally ok. So, this is essentially a typical or the most classical statement of isoperimetry there is ok. So, in some sense we are stating the dual form of the classical isoperimetric theorem, isoperimetry theory means shapes that have the have equal perimeter or surface area.

(Refer Slide Time: 02:47)

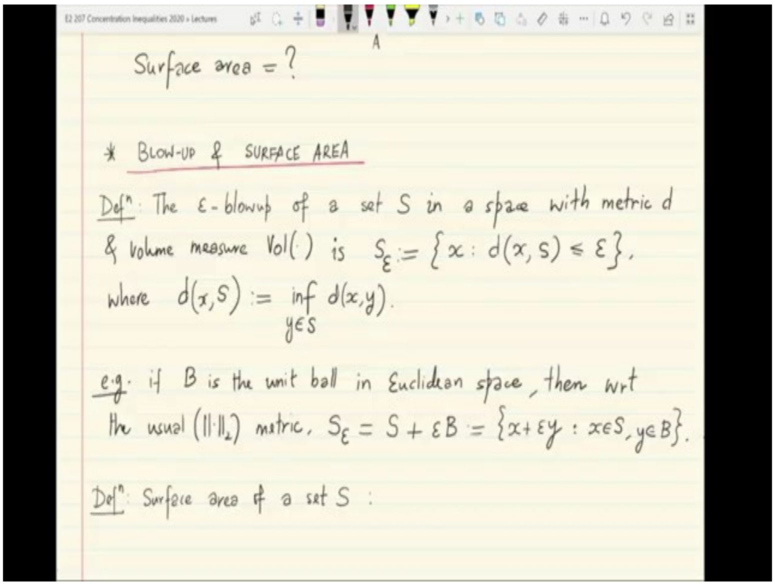


And the dual version of the statement is that among all shapes of a given surface area the Euclidean ball has the maximum volume ok. Now there are 2 terms here which are very important here to understand in such a statement there is a term which is indicative of a measure ok or volume. In the sense that it measure size and there is a term that measures what is called perimeter or surface area in some sense we will show that this can be interpreted in terms of distances ok.

So, what we need typically to frame classical like isoperimetry statements is the notion of a metric measure space. So, in all that follows in the background we will always assume that we are working in a space that has both a notion of a metric or a distance measure as well as a volume measure ok.

So, again coming back to a statement like the classical isoperimetric theorem, let us be a little more precise about what we mean by volume and surface area. So, here volume of a set A let us say line in n dimensional Euclidean space is just taken to be its standard measure or Lebesgue measure more precisely and this is just given by the integral of dx over A for instance.

(Refer Slide Time: 04:21)

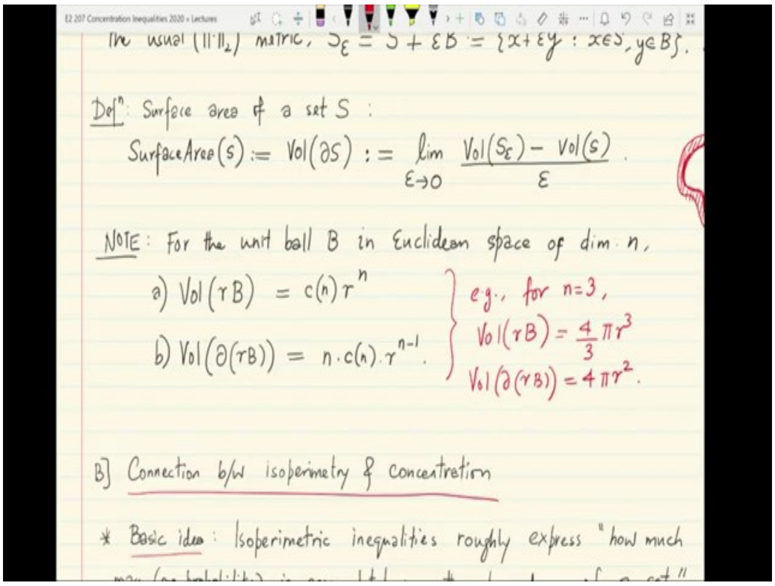


What about surface area? So, what about surface area? How do we define the surface area of a set or an object or a shape? So, one of the ways of doing it is through blow ups. So, let us define for a given set S the epsilon blow up of that set S. So, note that S is sitting in a space with metric given by d and the volume measure given by vol. So, the epsilon blow up of a set S is defined to be let us say notation S epsilon is the set of all points x such that the distance of x from S is utmost epsilon ok.

Where by the way what is the distance of a point from a set d of x comma S, is just defined to be the minimum distance of any point in S from x ok. So, S epsilon is essentially capturing the set of all points which are at a distance of utmost epsilon from the set S. So, its a the set S with a small expansion this you will. So, for instance in standard Euclidean space if B is the unit ball in a standard Euclidean space; that means, the metric is the l 2 metric.

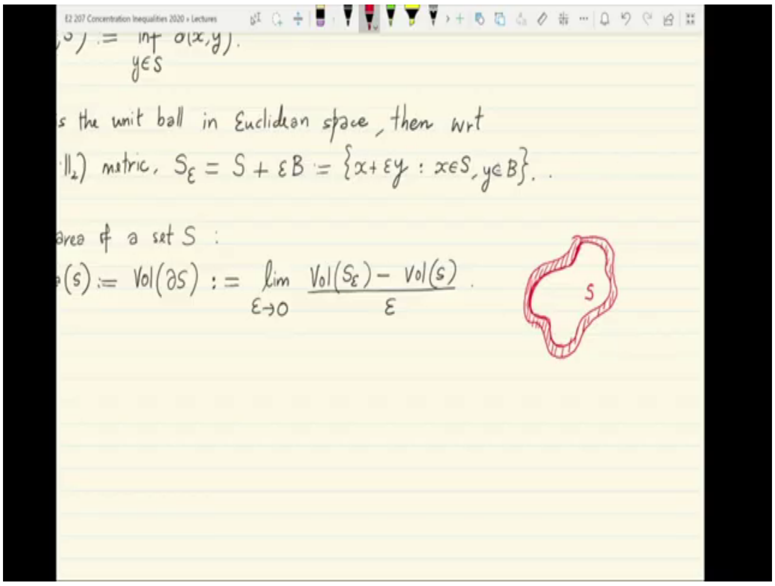
Then with respect to the usual, that means l 2 metric the epsilon blow up of any set is essentially obtained by taking S and adding to it in a set theoretic sense a scale multiple of epsilon * B. And this in turn can be explicitly defined as the set of all x + epsilon y, where x belongs to S and y belongs to B ok. Now once we have the notion of a blow up its quite natural to be able to use this to define the surface area of a set S.

(Refer Slide Time: 07:17)



So, here is one particular definition of the surface area with respect to a volume measure and a notion of distance. So, the volume so let us see. So, in the surface area in this view of S is defined to be the volume of the boundary of S ok and we define this in turn to be the limit as epsilon approach to 0 of the volume of the epsilon blow up of S - the volume of S divided by epsilon ok.

(Refer Slide Time: 08:05)



So, what this does in pictures is that if you have this set abstractly which is called S. So, how would you evaluate in some sense the measure its surface area its surface area is roughly the measure of this the size in some sense of this red curve and how you measure this in some sense is to try to take a small blow up. So, the epsilon blow up of S is essentially the set.

The volume of the epsilon blow up of S - the volume of S in some sense is going to be able to give us a measure of the difference area which is the shaded area here ok and that is what the surface area ends up measuring. So, by the way we are not going to be technically very precise and assume that these limits exist, but one can always show that for an appropriate class of bodies or sets S.

Such a surface area measure is well defined in the sense that the limit exists. So, again some examples, so one can use this to understand the usual surface area of sets in Euclidean space. So, for the unit ball B in Euclidean space of dimension n the following is well known which is volume of let us say a scaled version of the unit ball r * the unit ball, which is basically the ball of radius r is scales as r to the n with some constant that depends on the dimension.

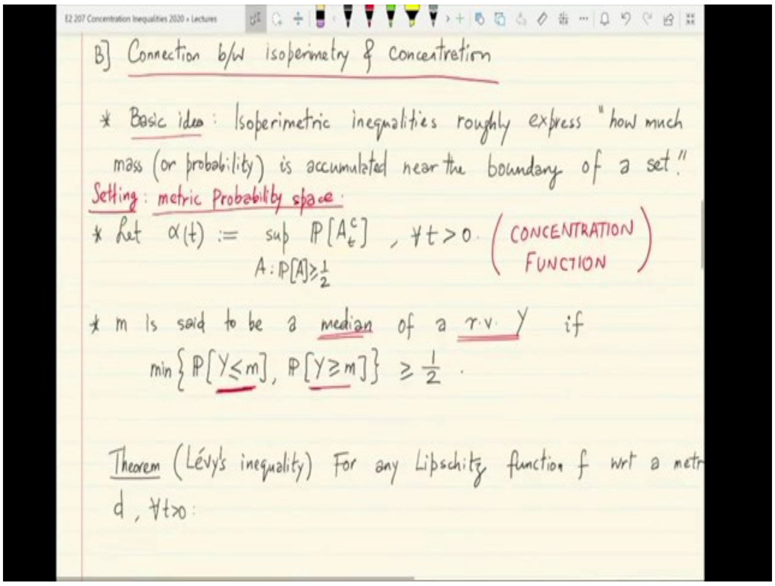
And the volume the surface area of this ball of radius r which is volume of the boundary of r B. So, recall that when we say the volume of the boundary of r B its really an n - 1 dimensional volume which is the surface area, it is known and this = n it is really in fact the derivative of the volume evaluated with respect to r. So, this is n c n r raised to n - 1 ok.

And this reduces to for 3 dimensional space we have the usual high school formulae that the volume of a sphere of radius r in 3 dimensions is some constant 4 by 3rd πr cube. So, 4/ 3 rd πis basically the c of n c of 3 and the volume of the surface area of a sphere is basically 4 π r square ok.

So, this generalizes to any dimension in the Euclidean in the standard Euclidean geometry in n dimensions with the standard Lebesgue measure which measures volume ok. Next we go to investigating the connection the exact connection between statements of isoperimetry such as this ok.

So, statements of isoperimetry will broadly be taken to mean the solution of problems where a surface area or a volume is given and you have to try to optimize over all shapes of either the volume or the surface area respectively to find sort of optimal shapes in some sense ok. So, that is an isoperimetry statement.

(Refer Slide Time: 11:59)



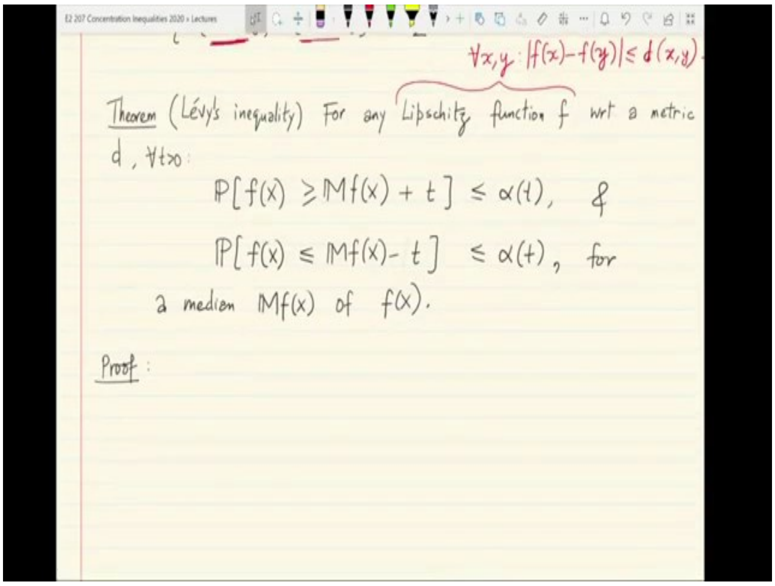
Now, what is the connection between such isoperimetry statements and concentration of measure inequalities? The basic idea here is that isoperimetric inequalities essentially in a critical way. Tell us exactly how much mass or probability is accumulated near the boundary of a set ok. So, that is what an isoperimetric statement is really trying to say ok. So, for instance in the classical isoperimetric theorem, its basically saying that the sphere sort of has a certain distribution of mass on its boundaries which is optimal compared to any other body with the same volume ok.

So, towards this let us define let us start by defining this very key quantity or function alpha of t. So, this is by the way defined for all t positive or in fact non negative. As the largest probability of a complement A t complement where A t is the t blow up of any set A of all sets A which have a minimum probability measure of half. So, imagine so by the way something that is useful to keep in mind is that the setting here is that of a metric probability space.

So, we are imagining a space that is endowed both with a probability measure and as well as a metric. So, here we are trying to define alpha of t as the largest probability of the t blow up of the complement of the t blow up of any set, which has probability at least half for every t greater than 0 and this is also called alpha of t is also called the concentration function. So basically measures you know how much mass there is in the in if you blow up A by a small amount t, when the set A itself is sizable ok.

And another definition that we will often need is that of a median of a random variable. So, what is the median of a random variable? A median of a random variable is any value such that the probability that the random variable is either utmost that value or at least that value are both at least half ok. There can be many medians for a single random variable its not unique, but we will call M to be a median if this property satisfied ok.

(Refer Slide Time: 14:45)



So, yeah now the main bridge between isoperimetry or concentration functions and concentration inequalities themselves is what is called Levy's inequality. Which says that consider any Lipschitz function f with respect to the metric d ok, so let us say there is a metric probability space on which there is a function f defined which is a Lipschitz function.

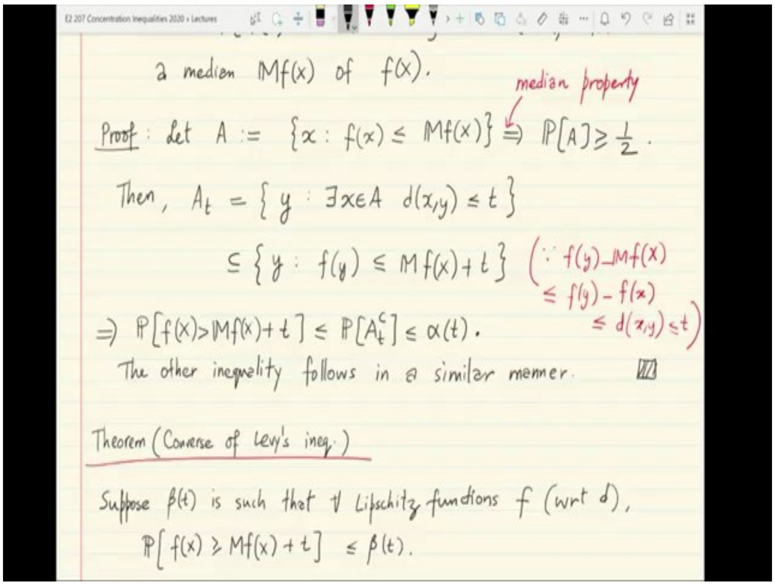
Now, by Lipschitz function what do we mean? We mean exactly that for all x and y in the space we have f(x) - f(y) to be bounded in absolute value by the distance between x and y the metric distance between x and y, that is what it means to be a Lipschitz function. So in fact, for any Lipschitz function any such Lipschitz function f on this metric probability space for all positive t we have these inequalities probability that f(x) is larger than any median of it M f(x) + t less than equal to alpha t.

Let us write M f(x) in boldface just in analogy with expected value, even though this is not unique and on the other hand as well probability f(x) is short by - t is short by t from it is median is also bounded by the concentration function t ok, for any median M f(x) of f(x) of this random variable that would x.

So, Levy's inequality basically says that if you can bound the concentration function which is a purely geometric quantity, then that at once gives you concentration results of a function of Lipschitz functions about their median values about their medians ok.

So, this is not concentration about the expected value, but this is very differently concentration about the median although in many cases one can show that the median and expected value are fairly close by themselves. But the natural spirit of Levy's inequality is to state it in the sense of the median concentration about the median ok.

(Refer Slide Time: 17:19)



So, let us go ahead and show the proof of this inequality. So, define A as the set of all x let us show the first inequality.

So, if you define A as the set of all vectors or points x where f(x) is utmost it is median sorry M f of capital X as a random variable, then by definition we know that the probability of A is at least half this is by the median property or median definition. And then we have that the t blow up of A which is defined as the set of all y S in the space where for which there exists an A exists an x in A with d x y utmost t belong.

So, this itself is a subset of all points y where. So, recall that if x and y are such that d x y less than equal to t then. So, we can say that the set of all such ys is a subset of ys where M f(x) is less than equal to f(y). So, any x belong any such x is assumed to belong to A, so we know that for X f(x) is upper bounded by M f(x).

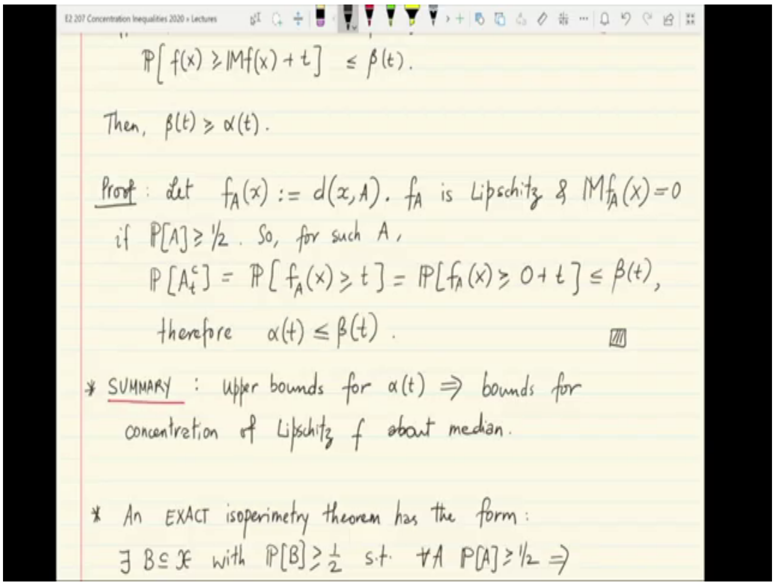
And we can say that f this we can replace M f(x) by f(y) sorry by sorry. We can replace M f(x) by f of small x ok, because of precisely this reason and we know that f is

Lipschitz. So, basically this is upper bounded by d x y and we know that d x y is upper bounded by t ok.

So, this is what gives us this inequality A t is a subset of y such that f(y) is utmost M f(x) + t and this in turn implies that the probability that f(x) is larger than M f(x) + t must be bounded by the probability of A t complement and by definition probability of A t complement can be utmost it is concentration function alpha of t.

So, we have the first inequality and the other inequality follows in a very similar manner and it is a symmetric counterpart of this one. So, that completes the proof of Levy's inequalities. In fact, Levy's inequalities are tied in a certain sense.

(Refer Slide Time: 20:37)



In the sense that there is this following converse, so the converse reads if beta of t is another function of positive t, such that for any Lipschitz function f where we have the same definition of Lipschitz functions with respect to this metric d. The if always a probability of f(x) larger than it is median + t is bounded by beta of t, then beta of t must be at least alpha of t ok.

So, this theorem Levy's inequalities are essentially tied over all Lipschitz function f and the reason why this is true is because of the proof of this converse which basically says that you take define this function f A of x for any given set a we can define a function f A of x as the distance of x from the set A.

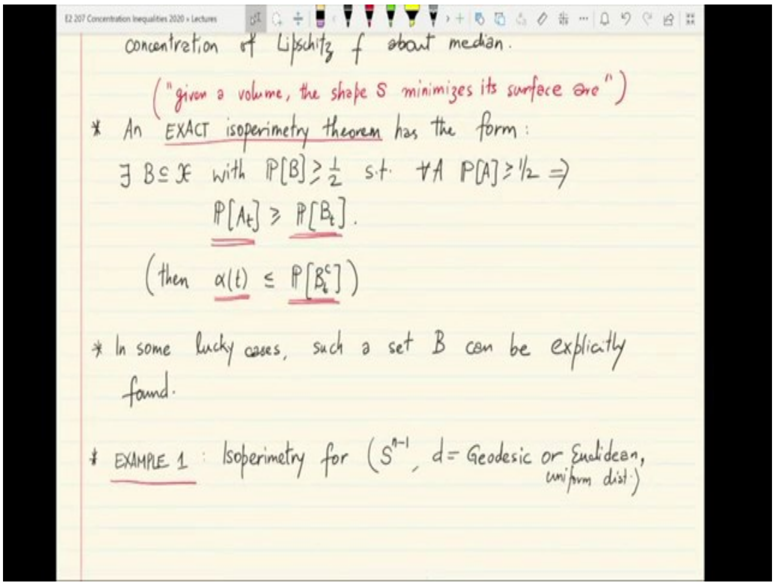
It can be easily checked that f A is Lipschitz continuous this is a Lipschitz function and the median of well the median of f A of capital X ok as a random variable is 0 if the probability of the set A is at least half ok. So, 0 is a valid median for f A of capital X, if the probability of A is at least half. So, for any such A the probability that probability of A t complement is the probability that f A of X this is just by definition greater than equal to t.

And we know by the hypothesis that so this is exactly P f A X greater than equal to 0 + t, 0 plays the role of a median in by hypothesis this is at least beta of t right. And so if we just take the supermom on both sides with respect to all such sets A with probability at least half, we get therefore that alpha of t is upper bounded by beta of t ok.

So, this basically establishes the other side of Levy's inequality and so the summary from this connection between ISO between the concentration function and fluctuations about the median is that upper bounds for alpha of t imply upper bounds for the concentration of Lipschitz functions about their medians ok.

Now, the so with this essentially the effort shifts to trying to find good ways to upper bound the concentration function in general or given metric probability spaces.

(Refer Slide Time: 23:57)

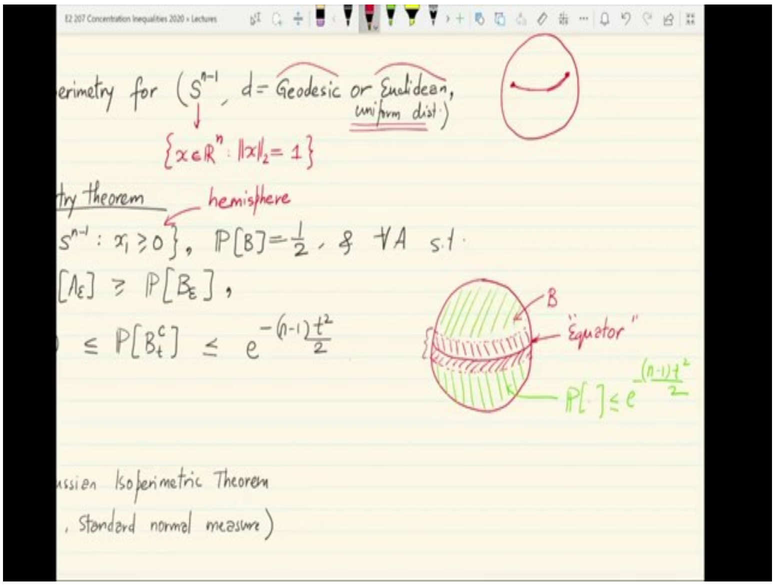


So, what we know so far is that in fact if you have an exact ioperimetry theorem. So, what do you mean by an exact isoperimetry theorem, it is a statement that says that in some sense given a volume the shape or body or subset S minimizes it is surface area. So, these are essentially called exact isoperimetry statements.

So, mathematically it has the form that there exists a B a subset B or a shape B the probability of B at least half, such that for every other shape A with probability at least half. The volume of the blow up of A is at least the volume or probability of the blow up of B. And this in fact implies that the concentration function alpha of t can just be bounded above by the probability of the complement of the blow up of P for every value positive t of positive t.

So, this is what an exact isoperimetry theorem gets you. So, in some lucky cases a set like this is explicitly definable and so here are 2 classical examples the first one is Isoperimetry statements of isoperimetry for the uniform measure or uniform probability distribution on the surface of the unit ball in n dimensions.

(Refer Slide Time: 25:35)

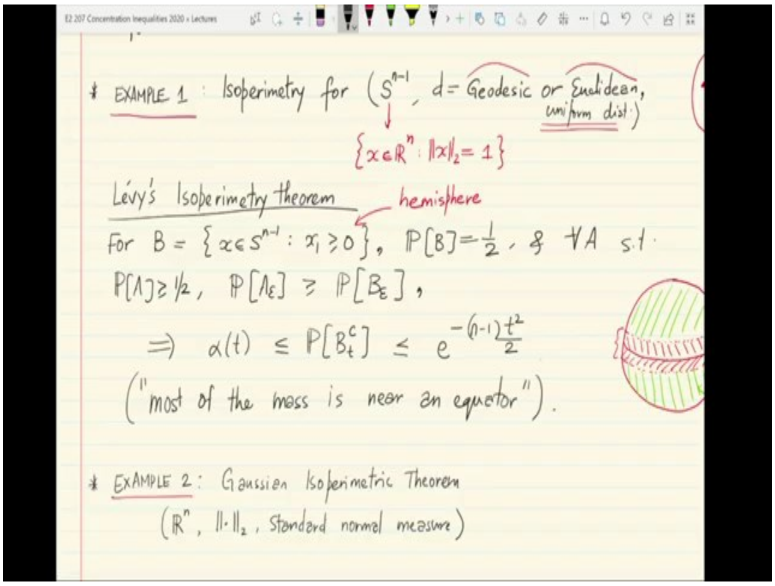


So, S n - 1 is notation for the surface of the unit ball in n dimensions is the set of all x in R n, such that their l 2 norm is exactly equal to 1 and imagine that there is a distance measure a metric on the surface of the unit ball.

So, one can have two different metrics which both work for this example, we can have what is called geodesic distance between two points on the surface. So, roughly speaking in 3 dimensions if this is the ball in 3 dimensions, then if you want to measure the distance between 2 points on the surface of the unit ball all you say is that you find the equator that connects them. So, if this is the segment of the equator that connects them on the surface of the unit sphere.

Then the length of that segment of the equator is their geodesic distance, equivalently one can also use a notion of distance like the Euclidean distance to measure distance on the surface of the sphere. But in any case imagine that there is this distance metric defined on the surface of the sphere which is symmetric and translation in variant over the surface of the sphere and you have the uniform probability distribution on the surface of the sphere ok. So, that defines the metric probability space for this example.

(Refer Slide Time: 27:07)



So, there is a famous isoperimetry result called again Levy's isoperimetry theorem, that says that that essentially specifies the sets of least sets of least volume given a surface area on the surface of the of the n dimensional units sphere. So, for B equal to let us say all points in on the surface of the sphere with the their first coordinate being non negative which is exactly a hemisphere. So, in the usual sense this would mean the northern hemisphere for instance any hemisphere will work for the purposes of this theorem.

We must have that your P of B = half and for all A such that probability of A is at least half the probability of the epsilon blow up of a is lower bounded by the probability of the epsilon blow up of this northern hemisphere or any hemisphere ok. And this finally implies that alpha of t is upper bounded by the probability of B t complement ok. And one can also do explicit calculations to show that the measure of the complement of B t can be upper bounded by something like e to the - n - 1 t square by 2 ok.

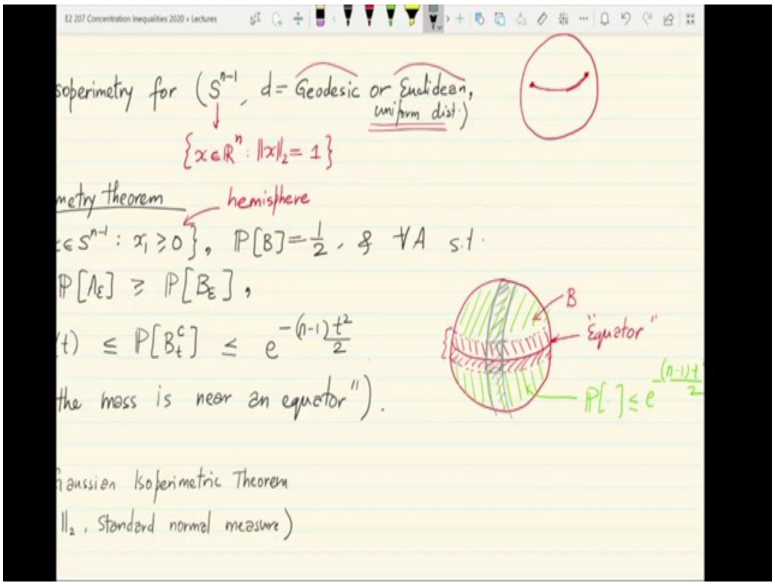
So, in other words what this says is that if you again go back to the unit sphere I mean let us say 3 dimensions. So, S 3 is S 2 is what we are talking about the surface. So, the probability metric probability space is the surface of the 3 dimensional sphere, then what it says essentially is that let us say there is this hemisphere which is let us say one hemisphere. Let us call this set B and the probability of the complement of B.

So, if you take a blow up so if this is B let us define a small blow up of it is so that the t blow up of B is the Northern hemisphere union the strip which is shaded here. So, go a little South of the equator. So, the solid line is the equator is an equator, if you go a little South of the equator by an amount proportional to t there is extremely there is very little mass in the complement ok. So, the probability of this part the remaining part falls off very rapidly with t so it is bounded above e raise to n - 1 t square by 2 ok.

And in fact, one can run the same argument with the southern hemisphere to basically say that you know you could form you go a little north of the equator ok and whatever is remaining mass is has extremely low probability. So, all it so finally by sort of basic probability it says that almost all the mass is near an equator ok.

So, most of the probability mass is near an equator and this in fact in high dimensions gives rise to very surprising consequences or non intuitive consequences, because you know if one argued that all the mass there is a lot of probability mass almost probability one on this equator.

(Refer Slide Time: 31:23)

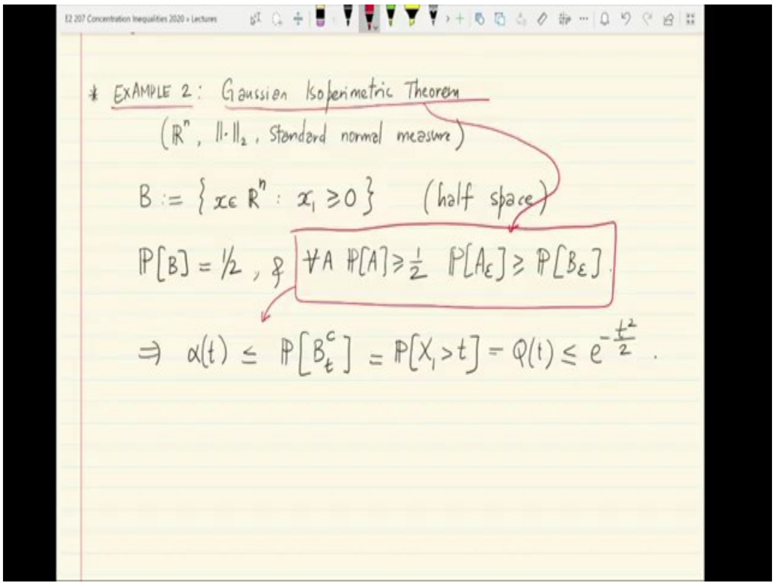


Well you could ask you could run the same argument for any equator, so this is another equator ok. So, this equator also has mass close to 1 probability mass close to 1. And so it must mean that their intersection also has probability mass close to 1 ok.

So, essentially it means that any two equators in high dimensional spaces in the surface of spheres in high dimensional spaces essentially have a lot in common ok; which is somewhat counterintuitive given our understanding 3 dimensional geometry. But in any case this is true of such metric measures basis.

The second well known example of an isoperimetric theorem is what is known for standard Gaussian measure. In n dimensions endowed with usual metric and we have standard normal measure which is essentially independent measure Gaussian measure in all coordinates, so here the statement of isoperimetry. So what are the optimal sets?

(Refer Slide Time: 32:31)



The optimal sets essentially turn out to be half spaces. So, take any B which looks like the set of all x S in R n with let us say one coordinate being non negative, this is basically a half space. So, its known so its obvious that probability of B = half because of the definition of independent Gaussian measure.

And for all a with probability of a at least half, we have that probability of the epsilon blow up of any such a is at least lower bounded by probability of epsilon blow up.

So, this is the Gaussian isoperimetric theorem. So, this statement is the Gaussian isoperimetric theorem that is the content of the Gaussian isoperimetric theorem and as a consequence of it we get that alpha of t is upper bounded by the probability of the t blow up the complement of the t blow up of any half space passing through the origin. And this because of properties of independent Gaussian measure is just P of X 1 exceeding t ok.

Which is the probability of a standard normal exceeding a value t which is what is defined to be the q function of a Gaussian which can be for instance upper bounded by e raise to - t square by 2 ok. So, alpha of t is it is possible to bound alpha of t by probability of B complement just by virtue of using this very powerful Gaussian isoperimetric theorem and as we know bounds on alpha of t can easily translate to bounds on the concentration of Lipschitz functions about their medians ok.

So, in this lecture we have seen how isoperimetry results or very strong or exact isoperimetry results help us to bound the concentration function alpha of t. But what if so such occurrences are more the exception rather than the norm, often its not explicitly possible to prove such exact isoperimetric results.

So, what happens when we do not have access to isoperimetric results for our given metric probability space of interest is there are there other ways to sort of understand bounds on alpha of t and that is what we will explore in the next lecture.

Thank you.