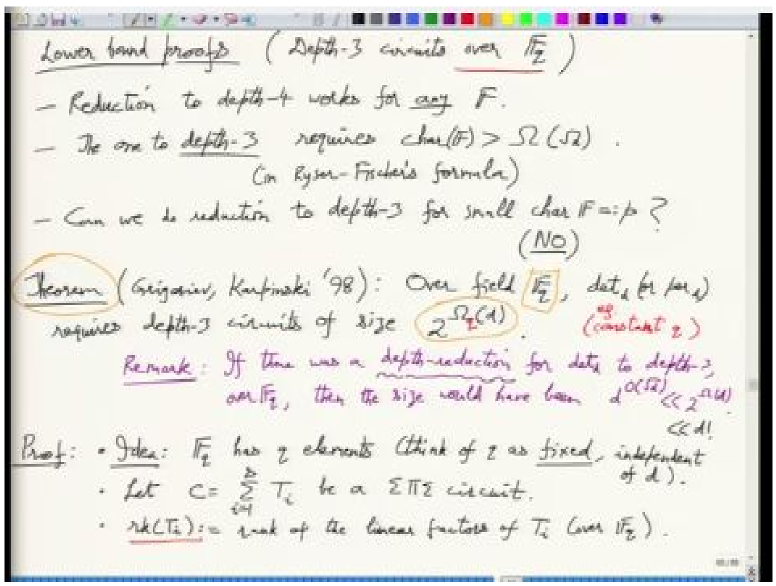
Arithmetic Circuit Complexity
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology-Kanpur

Module-12
Lecture-14
Lower Bound of Depth-3 Circuit Over Finite Fields

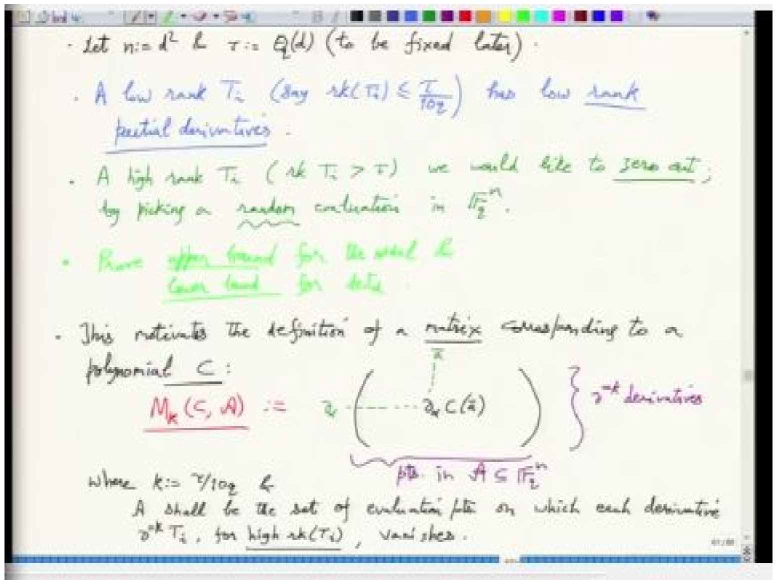
(Refer Slide Time: 00:17)



So last day we started this theorem which will basically be a depth 3 circuit lower bound for determinant and permanent and it will be exponential. So this is an exponential depth 3 circuit lower bound for determinant and permanent. The only caveat is that we have to assume the field to be a small field, a very small field. So we will in fact assume that the field size is constant.

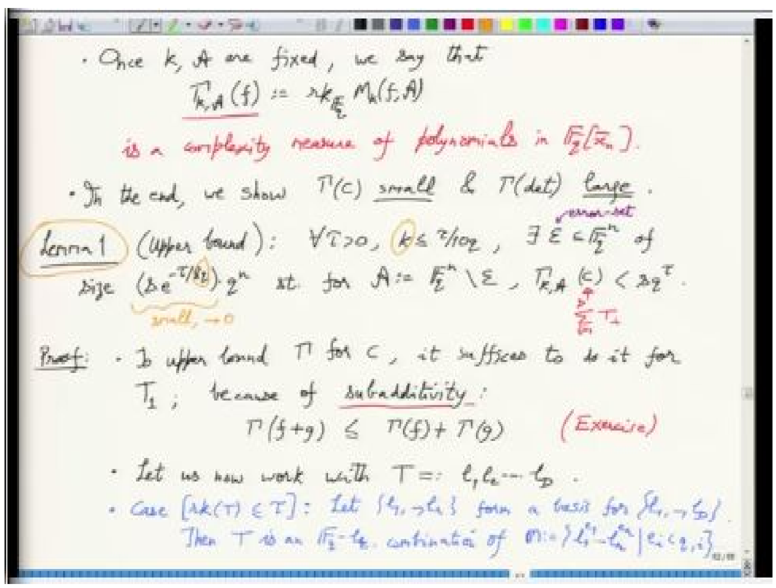
We will see in the end of the proof that proof can afford q to be slightly smaller than $\log d$ but not much more. So certainly when q is $\sqrt{\log q}$ will still get a lower bound but if q becomes larger, then the proof will break down, it will give a trivial lower bound. So the idea will be to consider a matrix and the rank of that will be a measure for polynomials.

(Refer Slide Time: 01:34)



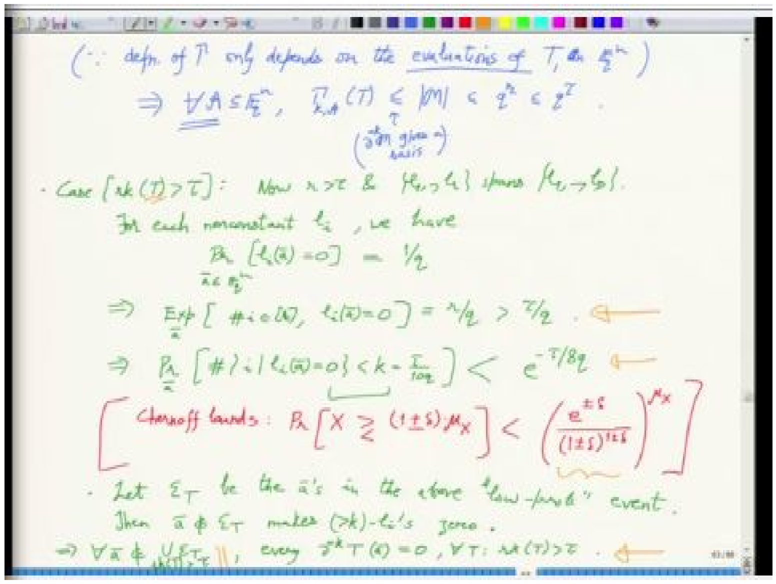
And, so that matrix is $M_l(C,A)$ is the polynomial, so the matrix has rows index by key order derivatives and columns index by points in a subset A of the space $F_q{}^n$.

(Refer Slide Time: 01:57)



So we are just looking at derivatives at points of a given polynomial and looking at the rank of that matrix. So the upper bound lemma that we proved, nearly proved, is basically saying that there is an A and there is a measure of this type which is small on the circuit, depth 3 circuits, it is a measure will be at most sq^{τ}. This was from Chernoff bound.

(Refer Slide Time: 02:50)



This should come from, so we have taken taken k to be $\tau/10q$, so that 10 actually gives you 8 in the probability. So this value of k, 10q is actually what is giving you 8q in the exponent here. So you can ask why are we using 10q, so that will be just to suitably set tau and k to define the measure we need a k and an A. So how we will pick it will be clear in the end because in the end we want lemma 1 upper bound and lemma 2 lower bound will invoke them together to get the final exponential lower bound.

So we have to take care of both the parameters, parameters in lemma 1 and parameters in lemma 2. So these constants and even this form τ varied by 10q is so that all that comparison works. So we have completed lemma 1, almost completed, let us finish it now the remaining small part. So we were looking at the case when we are just trying to study the measure on a product of linear polynomials which is this T.

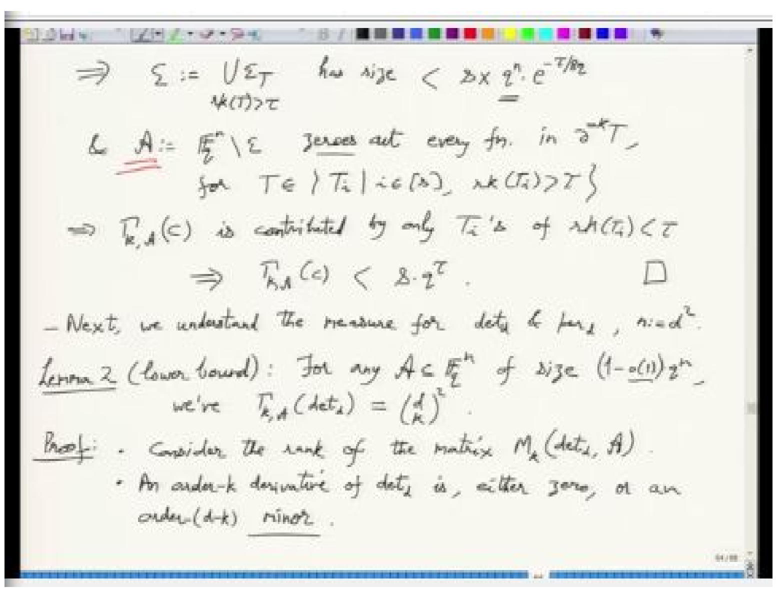
Moreover the hard case is when the rank of these linear factors is large, bigger than τ . So in that case we have to argue that so in that case the set A will be used, the reason to have the set A in the measure was exactly for this case. So we will argue that if we had picked a random point in the space square F_q^n then with the descent probability not only will T vanish but even it's k[th] derivative will vanish.

So basically T will have roots of multiplicity k, in fact more than k, so that basically follows from Chernoff bound. Because what we so look at this line, so we have shown that the number of l_i's, the factors of T, that vanish at a random point simultaneously, this is expected to be r/q which is τ/q, so the probability that it is 10% of that is this probability.

So the probabilities are this actual number in an experiment comes out to be 10% of the mean expected to be very small and quantitatively it is this e raise to – 1/8th of the mean. So this 1/8th is coming from this 10% that we are chosen of the mean. So from all that we did use this line that these ε_T's are the eranius or the bad a's on which the number of l_i's that vanish or less than 10% of the mean.

So those bad a's defined ε_T and since you have many T's you have at most s many T's, so take the union of all these bad sets. So that is being done here, so if you pick an A that is not bad then all these T's k^{th} derivative will vanish at the point. And so now once you notice that the union of it is small the remaining points are many ok, there is a huge number of points such that all the derivatives vanish of all the T's ok.

(Refer Slide Time: 07:44)



So let us just work that out, so let us call this $\varepsilon := \cup_{rk(T)>\tau} \varepsilon_T$, of course in this case here only considering multiplication of product gets T with high rank others we have already handled. So this has size at most so there s many times, what is the size going down ε_T? Well q^n is the whole space times the probability. So probability is essentially falling as 1/8th of the mean it is – $\tau/8q$ in the exponential, e is the natural log.

So it is around 2.8, certainly this is falling exponentially, it is half raise to $\tau/8q$. So if you look at the density of this so divide this by q^n so that density can be made extremely small. It can be made tending to 0, you just have to pick out to be big enough as a function of log s. And the good points will be complement of this, so this zeros out every function in k^{th} order derivatives.

If any k^{th} order derivative of T that will vanish for these T's, of high rank Ts every kth order derivative will vanish for every point in the subset A right. That is the final thing this set is quite this is almost everything. So this implies that gamma k A c, so now let us look at the measure. So when you look at the measure what does it depend on? Which T i's will it is depend on?

This measure is actually just killing high rank T i's right, so it only depends on low rank T i's. So this is contributed by only T i's of rank less than tau which means what which will give you a an obvious upper bound on the measure. Because for a low rank T the measure comes out to be q raise to tau and there are atmost s many right. So you have shown the upper bound, you wanted to show this s q tau and this; what we get, right.

So now you can see that this whatever we did was actually quite natural we identified these T i's of high rank we kill them by evaluation. And what remains is low rank and their contribution cannot exit (()) (12:13) s times tau ok, any question. So now let us move to the lower bound lemma which will be for a very specific polynomial, what we currently did is for any polynomial computable by depth 3 circuit sizes.

So next we understand the measure for determinant and permanent d cross d, so number of variables is d square. So this will be lemma 2 for any subset, so we have to see something about the size in the hypothesis. So what we will we would want to see ideally is that it is the whole space but since in our lemma 1 we have seen that E could also be could very well be smaller than the whole space.

So we will at least need this promise there it is sufficiently large ok, so this little o of 1 is saying that the difference of I mean this if you look at the density then 1minus for density is tending to 0 for increasing n ok. So the density is nearly 1 so with this promise we can show that measure is huge, so measure of determinant is d choose k square, is this statement clear, d choose k ok.

So since we have not really fix k A, so it may not be clear why we are saying this we are saying that this is large. So think of k is d by 2 alright, if you think of k is d by 2 then this comes out to be very close to 2 raise d. So that essentially the measure is 2 raise to d and if you remember the original theorem the final theorem which you want to show. We want to show that s is something like 2 raise to d, so that we will actually get because of this d choose k square ok, this is the reason for 2 raise to d.

So this is a property of another property of determinant that this matrix of derivative at points has large rank as biggest 2 raise to d. If we could have made this d raise to d if this expression was something like d raise to d then we would have gotten a really optimal lower bound on s. But we cannot improve this, so for determinant we cannot improve and also for permanent.

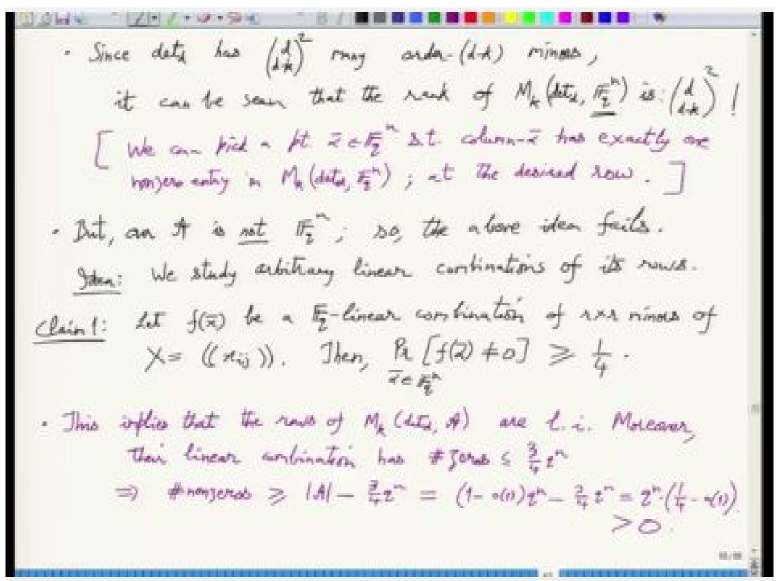
So right, so what we have to do is consider the rank of the matrix M k determinant comma A. So it so again recall that the rows order derivative operators and column for points and the points are nearly all of the space q raise to n. So how do you lower bound this rank it is a huge matrix correct. But that is not enough. We have to say something about the rank of this matrix.

So even if this matrix every entry is non 0, the rank would be 0. So well it cannot be 0, it is 1 rank can be as small as 1. So how do you show that actually the rank is large, so the best way known towards is identify a triangular structure? So in your matrix identify a triangular matrix and then the size of the triangular matrix is a lower bound on the rank right. So which derivatives should you take which rows should you take basically.

So that some matrix looks triangular right, so that is the thing to think about. So an order k derivative which will obviously we have partial derivative of determinant d is, so it can happen that the formally the derivative vanishes right. So, for example if you differentiate your matrix by first by x 1 1 and then by x 1 2 or by x 1 1 and again by x 1 1. It is these will just formally vanish, so there is no point looking at evaluations.

Because formally there is vanished otherwise what will you get if it is formally not vanishing then what is this, it is a minor right, so it is a minor. It is a determinant of a sub matrix it is an order d – k minor, so we will focus on this right. So basically a row corresponding to this minor I mean row is then just this minor being evaluated at almost nearly the whole space. But actually unfortunately still there is no signs of a triangular matrix triangular sub matrix in this right. So probably that will not work we will do something else.

(Refer Slide Time: 20:19)



So since determinant d, so how many d – k order minus are there, it is a d cross d matrix. So d choose d by k, d choose d – k each for row and column choices, so it is square that the rank of if we took A to be everything the whole space. Then due you know the rank of this measure matrix, so we are taking A to be everything, so my claim is that then it is this numbered of minus ok.

Then it will be just these exactly, so all these ordered d – k minus when you consider the rows in this measure matrix, they are linearly independent. Assuming that low point is missing in the column, column is all the points, the point where only that minor does not manage, exactly. So because if A was everything then you can identify a triangular sub matrix in fact diagonally right.

So this will be the proof, so it is not very important but let us anyways sketch this. So can pick a point alpha bar in q to the n such that column alpha has exactly one non 0 entry in this matrix at the desired row is. At this row which corresponds to order d – k minor, so you pick your desired or your favorite row. And for this minor you can choose a column alpha bar.

So (()) (24:02) consist means if you have make a minor can you make all of minors that are contained with s – 0 should make this not 0. Why do you say contained with all of the same order they have all, exactly case not as there, right. So this alpha bar will just be I mean you look at a M cross m matrix symbolic matrix just set everything to 0 except the I mean some variable 1 right, so except 1 set everything to no, no.

So or just set the diagonal entries make it basically identity matrix to put it simply, just set it to identity matrix and then, so that will give you a standard alpha bar when you extend this idea you will get alpha bar. And then as a small exercise show that for this setting of variables no other order d – k minor survives right. So it is a small exercises by the definition of determinant, so that is it this is the idea. So you have using alpha bars, you have this way to get a sub matrix which is diagonal and it is full rank, so you get d choose d – k whole square, right.

So this is a simple idea but this has no utility for us because we do not have the whole space right. In fact we might be missing a fraction of q raise to n many points. So if we miss those the good alpha bars have been missed then this proof idea does not apply at all. But our A is not F q to the n, so the above idea fails, so what do you do, so what we will do is the next best thing which is study the linear combinations of the rules ok.

So basically try to show that first of all the dimension of this matrix the number of rows in this matrix is how much d choose d – k whole square right. So you want to show that these all these rows are independent. So if you take an arbitrary linear combination non 0 linear combination and you show that there is some alpha bar at which this linear combination does not vanish right.

So if we mean that the rows are independent, it because you have you started with an arbitrary linear combination. So for an arbitrary linear combination if you are if there is some alpha bars, so that this linear combination is non vanishing. Then it means that the linear combination is a non 0 row which means that the rows are independent. So we study we have to study arbitrary linear combinations of the rows right so.

By the way this easiest thing that we did this proves an interesting property of determinant that if you take equalized minors all of them like as formal polynomials. Then they are linearly independent polynomials that which is our non trivial fact right. So if you take all the for example first order all the d – 1 order d – 1 minors I mean even showing that these d choose d – 1 which is d many polynomials they are linearly independent right, this seems non trivial.

And we are showing it actually for any order but we have to show more we have to actually talk about it is evaluations in a smaller subset of points. So ok we will prove a stronger property than what I just said ok, so we will show that if. So you have a symbolic matrix x or big X and look at all the r cross r minors of this ok and take an arbitrary linear combination of this minor polynomials let us call it f.

Then if we want to look at the probability that f alpha bar is non 0 for a random point, so what is this probability right, no it is a non trivial fact may not be hard to show. So you are saying that the support disjoints, no, no is it probably or are you saying that the support is disjoint. To form a monomial you can did use. So that could be a possible proof yes, to check it what.

So using such things yes which I think is also what we did here by picking alpha bar but then. So if you have indeed shown that the minors are linearly independent then this f cannot be a 0 polynomial ever, right. If you are taking a non 0 linear combination of the minors f will always be a non 0 polynomial it is also multi linear. So then by (()) (32:49) we already get this right. So even when q is 2 the worst possible case if field is g f 2 even then we will get this probability to be half right.

Actually that will enough I think for the calculations maybe I do not need this claim I will prove something better 1 by 4 but I do not think it is this is important then because already at this point I could have done claim 1 with probability half no, no, no some. So the short zipple that we have discuss before is for the total degree you mean r r by q. in claim 1 r cross r.

But individual degree is only 1, so cannot you get 1 over q that is what I am not sure, total degree so the version we have proved of short zipple requires total degree. But individual degree individual degree is the ok, so maybe it is not what I think. So we will do this formally we will show that the probability of f non vanishing is at least 1 by 4 ok. Before proving that claim what does the claim imply, so this implies that the rows of your matrix M k now with this smaller subset A.

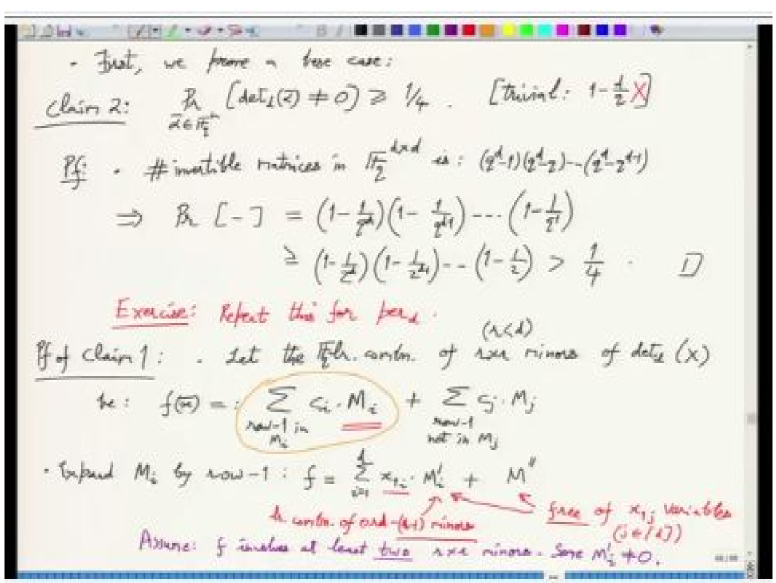
That we will first of all they are linearly independent moreover their linear combinations has number of zeros less than equal to 3 4th of q raise to n. With this is just because of this probability calculations, probability saying that being 0 is less than equal to 3 4. So when you take an arbitrary combination of the rows of matrix M k what do you see, you see basically evaluations at alpha bar in A right, A is a subset of whole space.

So it remains true that the number of zeros cannot 3 4th q raise to n. So number of zeros is less than equal to 3 by 4 q n which implies that number of non zeros is at least the size of A – this worst case 3 4 q raise to n. And here you use that A is large right, so A is 1 – small o of 1 q raise to n – 3 4th q raise to n which is q raise to n times 1 4th – small o of 1. But small o of 1 this tending to 0, it is in particular smaller than 1 4th, that is all actually we needed.

So this is positive ok so in other words any linear combinations of the rows of the matrix is a non 0 vector which means that the rows are independent which means that the rank is as claimed d choose k square right is that clear. So we will just show claim 1 now, well so the base case of claim 1 is when r is maximum, so when you are looking determinant itself right.

So what is the probability that determinant at alpha bar vanishes where vanishes or does not vanish. So short zipple will give you as you are claiming d over q and q is very small so it is meaningless it how do you get 1 4th or 3 4th. So that needs a proof no, no, no there is no possibility of boosting this is a.

(Refer Slide Time: 38:44)



So let us do that already is nontrivial then. So all these problems are happening because q is extremely small right otherwise these things are trivial. And the other parts of the proof will fail if q is large, so we have to break this cycle. So first we prove a base case it is call it claim 2, so here we will show that the probability when f is determinant d non 0 is 1 4th. So trivial would have been d = 0 is d over q so 1 – d over q, this would have been trivial but this is actually now a negative number because q is much smaller than d.

So this does not work, so short zipple is actually not applicable then we have to work harder. In fact we will use the use properties of determinant and permanent well so determinant of alpha bar being non 0 means that alpha bar represents the non similar matrix right. So how many matrixes are there just do the count, so number of invertible this is alpha bar represents an invertible matrix.

So number of invertible matrixes in f q d cross d is so who knows this count there is a formula no there is no P. So q raise d – 1 q raise to d – q. So the this is just counting the rows, so first row has q raise to d – 1 possibilities and once you have fixed the first row then the second row is remove the q multiples of the first row. So second row possibilities become q raise to d – q and so on that is it right, so this gives you the probability now.

So the probability is equal to this number divided by q raise to d square right, so it is 1 – 1 over q then 1 – 1 over q square and so on which now we want to lower bound by 1 4th, so what is that why is that q is at least 2 right. So you get 1 – 1 over 2 raise to d, so in fact fix q to be 2, so what do you get, that is the worst case. . This is 1 4th lower bounded by 1 4th in fact strictly greater than right strictly greater than 1 4th is that clear that is the proof.

But what will you do with permanent what is the permanent analogue of this, permanent of alpha bar non 0 has no good interpretation. So what will you count, so let us leave as an exercise for permanent. So I think for permanent the number of alpha bars so does permanent does not vanish comes out to be the same number. You have to use the definition of permanent to get to this, let me skip that.

So now with this base case and mind let us go back to the proof of claim 1. So now let the linear combination f q linear combination of r cross r minors determinant d of a formal matrix x, r here you will take now less than d, d f x bar equal to so we will take we are taking a linear combination of the minors M i's of the matrix x. And we want to divide this into 2 parts, so in one part the minors are respect to I mean the ok the first row is used.

So these are the minors with row 1, so row 1 in M i and these are the remaining ones row 1 not in M i in M j ok. So since row 1 is in M i we can expand M i the respect to the first row ok with respect to this x 1 start variables expand M i by row 1. So what will you get f is sigma x 1 start so let us say x 1 i times M i prime plus and let us remaining is some kind of garbage part M prime**.**

So these M i primes are since we get M i prime when we expand this right, we expand M i with respect to the first row of determinant of x. So this expansion will reduce the order of the minor right, so these will be should be now r – 1 minors, so order r – 1 minors. And what is M prime actually M i prime is not itself a minor it is a linear combination. So linear combination of order r -1 minors right.

So this C i's will contribute and we have taken out x 1 i. This M prime are those is the part which is free of the variables that appear in the first row. So this is free of x 1 j variables j 1 to d, so this is the decomposition, so what can you do with this, this is the nature of your f. And so you want to show that when you evaluate f at a random point it will not vanish with probability 1 4th right.

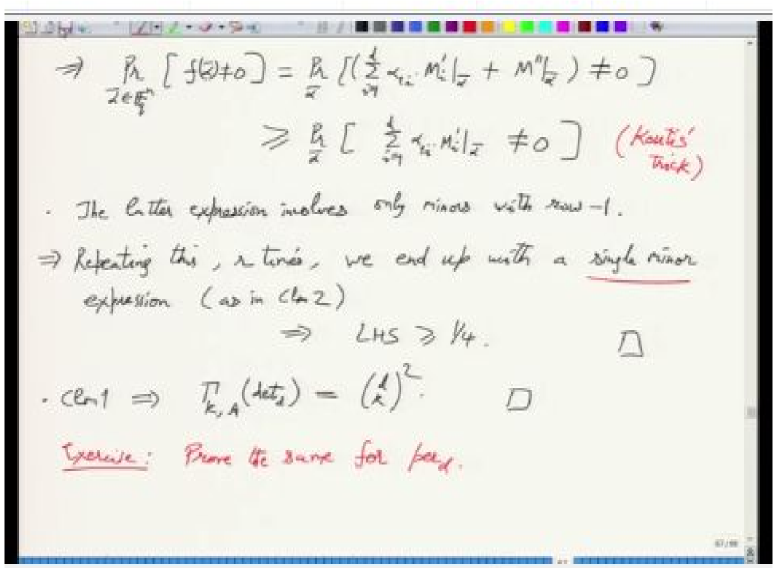
So how do you get that, we want to use the fact that M prime is somehow disjoint from this first part. This M i prime is also free of x 1 j variables right, so both of them of free of x 1 j variables. So x 1 j variables have been isolated they have been brought out so the idea would be that except the first row the other rows you fix randomly. So when you do that it is possible that all these primes and prime primes vanish that will be a bad case.

If not then you have a linear function in the first row variables and you know that it will not vanish with probability 1 – 1 over q, formally you or with after substitution, no even if all the M i primes vanish if M prime does not vanish then f does not vanish. If so what so only problem is when M i primes and M prime all of them vanish of at a point, then you have anything to argue.

I mean then obviously f is also vanishing x 1 i will not help that is the only bad case. So we have to use induction and this structure to finish the proof. Now you are picking a random points, so it could vanish it is a arbitrary point alpha bar. Sure right. One thing we can assume here is that these M i primes at least 2 of them are at least what I want to say that.

I mean since we have covered the base case, we will assume that f is a linear combination of at least 2 minors. At least 2 r cross r minors and hence we can assume that at least one of the M i primes is non 0 formally ok. So what will you do when all these M i's were this part was 0 to begin with there was no row 1 minor, no minor involving row 1, well row 1 was just arbitrarily chosen, you whatever row appears you pick that right. So hence we can assume that in this first part something is actually appearing, so one of these M i primes we can assume is non 0 I think that is all I need to assume.

(Refer Slide Time: 54:30)



So probability over the space f q to the n such that f at alpha bar is non 0 ok, this is the same as the probability that alpha 1 i M i prime at alpha bar plus. So such that decomposition at the point alpha bar we are looking at this is non 0 obviously. And I want to say that this probability is at least the probability that the first part is non 0 is that clear, why is that. No we have evaluated the variables at alpha bar but I think it uses what you are saying, that is the proof.

Well, if M prime is vanishing at alpha bar then it is these 2 probabilities are actually equal. So the only case you have to look at is what if M prime, prime is non vanishing at alpha bar right. So it comes out to be a number but it is free of these alpha 1 i's it is independent of that right. So then you should think of alpha 1 i's as not yet fixed there actually variables.

So this linear this sum has to take this particular value, so that probability will not depend on the value. So you can assume the value to be as well 0 ok that is the proof of this, did we just show that it is equal and not greater than equal to anyways this trick has name it is called Kautis' trick. So right so what have we reduce the problem to, so the latter expression involves only minors with row 1.

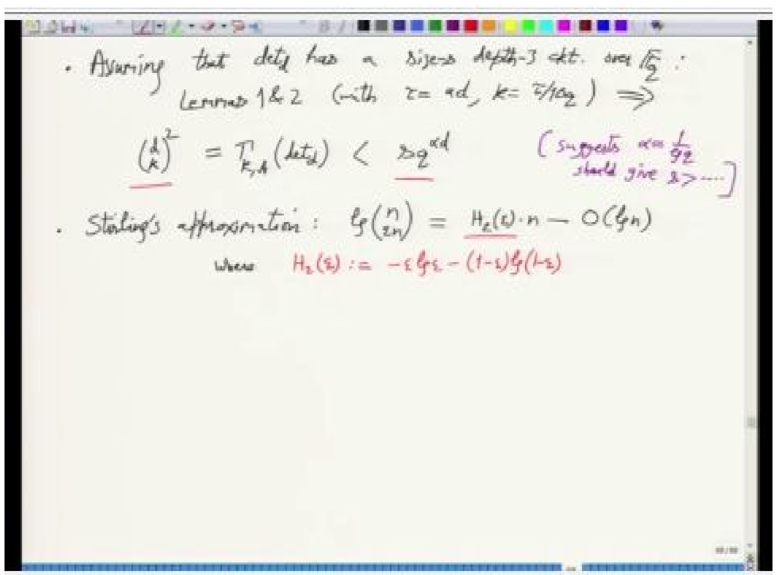
So we can repeat this trick right and reduce to an expression where there is some other rows row 1 and row 2 both are present and row 3 dot dot all the rows are now fixed. So your minor is fixed, so you have reduce to a single minor kits which is the base case, so repeating this several times. So basically I think r times, we end up with a single minor expression, so that was covered in claim 2.

So the probability is hence the same as shown there is this clear. So this trick is use to reduce the number of minors in the expression eventually go to the base. So if you look at M i prime then yes but by expression I mean the sum, the sum is still on the r + r minors, right the sum is just a reformulation. So the sum remains all we have done is we have eliminated those minors that do not involve row 1.

And the this can be a strict inequality in the case when M i primes vanish at alpha bar right. So if they vanish then because of that the probability will already become 1, so it is strict inequality but it is at least we at least get greater than equal to in all cases, right. So, claim 1 is shown which we have discussed before the measure exactly actually not just a lower bound.

So when you look at this M k matrix of the determinant for columns less than the whole space f q to the n the rows are actually linearly independent, it is a non similar matrix. So again this proof you can complete for permanent, you can repeat this for permanent so eventually these properties and the measure it is not able to distinguish determinant from permanent, right. So we have proved lemma 1 which was upper bound on circuits and lemma 2 which is lower bound on determinants right. So this is the right time to compare the 2, so that we finish the proof.

(Refer Slide Time: 1:02:43)



So assuming that determinant d has a size s depth 3 circuit over F q, so lemmas 1 and 2 imply something about the parameter with tau is alpha times d k is tau over 10 q. Till now we have not anything about tau right, so tau will be some multiple of d multiplier is alpha k v f fix to be the maximum lemma 1 which is tau over 10 q. We get that the measure on determinant because it has a depth 3 circuit it is smaller than sq to the tau which is alpha d.

And by determinants property lemma 2 this is equal to d choose k square right we get this. And now look at the end points, so you will get that s is greater than d choose k whole square times q to the – alpha d, right. In we have to see how good this is because by I mean from phase of it may not be very good because d choose k whole square is something like 2 raise to d. and q raise to alpha d is also like well actually we are.

So assume q to be constant then q is like 2, so this is 2 raise to alpha d, so this is then intuitively s has to be at least 2 raise to 1 – alpha times d right. So this suggest that if you take alpha to be a fraction thing should be fine right, you take alpha to be a fraction actually depending on log q. So you should take alpha to be 1 over log q right, this is what this suggest, right.

You take alpha to be 1 over log q should give s greater than something like what we wanted right. But this is obviously not correct because d choose k whole square is not 2 raise to d, where it has a complicate expression. So we have to do this then in more detail, so let us do that calculation. So what do you know about binomial estimates, sterling's approximation, but at least this back of the calculation tells you that there is no hope for q large.

If q is really growing then the r h s here is growing, so when it is becomes hopeless, right. Then you will just get that s is greater than 1 or s is greater than 0 you will not getting meaningful lower bound. But for constant q you have to work this out to find out what q will work. So for that we have to go into Sterling's approximation which tells us that log of n choose epsilon n is what.

Let me express in terms of Shannon's entropy where H 2 is what is H 2 epsilon exactly sorry, right. So since it is a log so basically n choose epsilon in something like 2 raise to n but if you try to pin down the factor that is it with fits in the exponent that is exactly Shannon's entropy ok. So depending on epsilon if you take epsilon to be 2 high like 1 then you do not get anything. And if you take epsilon to be very small like 0 or close to 0 then again you are in trouble.

But if you take epsilon to be somewhere in the middle like half that gives you the most mileage ok, so that takes you to 2 raise to n and there is this error term of log n which we can ignore this. Because the main term is n, so we can ignore log n I think today also we cannot finish this. There are some calculations remaining, so we will do this on Saturday or here that Saturday is a class working day Wednesday right. So let us finish this and start another lower bound, any questions.