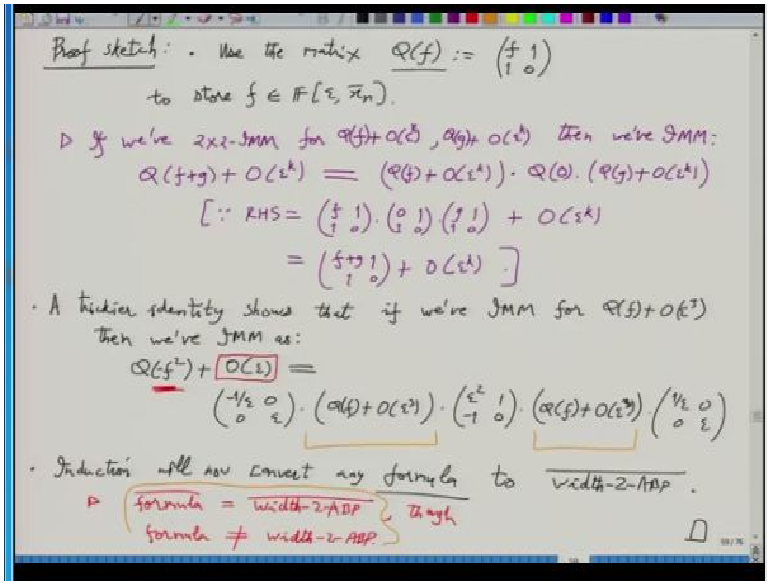
Arithmetic Circuit Complexity
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology-Kanpur

Lecture-13
Grigoriev-Karpinski Measure

(Refer Slide Time: 00:15)



Proof sketch: • Use the matrix $Q(f) := \begin{pmatrix} f & 1 \\ 1 & 0 \end{pmatrix}$ to store $f \in \mathbb{F}[\varepsilon, \overline{x}_n]$.

▷ If we've 2x2-IMM for $Q(f)+O(\varepsilon^k)$, $Q(g)+O(\varepsilon^k)$ then we've IMM:

$$Q(f+g)+O(\varepsilon^k) = (Q(f)+O(\varepsilon^k))\cdot Q(0)\cdot(Q(g)+O(\varepsilon^k))$$

$$\Big[\because \text{RHS} = \begin{pmatrix} f & 1 \\ 1 & 0 \end{pmatrix}\cdot\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} g & 1 \\ 1 & 0 \end{pmatrix} + O(\varepsilon^k) = \begin{pmatrix} f+g & 1 \\ 1 & 0 \end{pmatrix} + O(\varepsilon^k)\ .\Big]$$

• A trickier identity shows that if we've IMM for $Q(f)+O(\varepsilon^3)$ then we've IMM as:

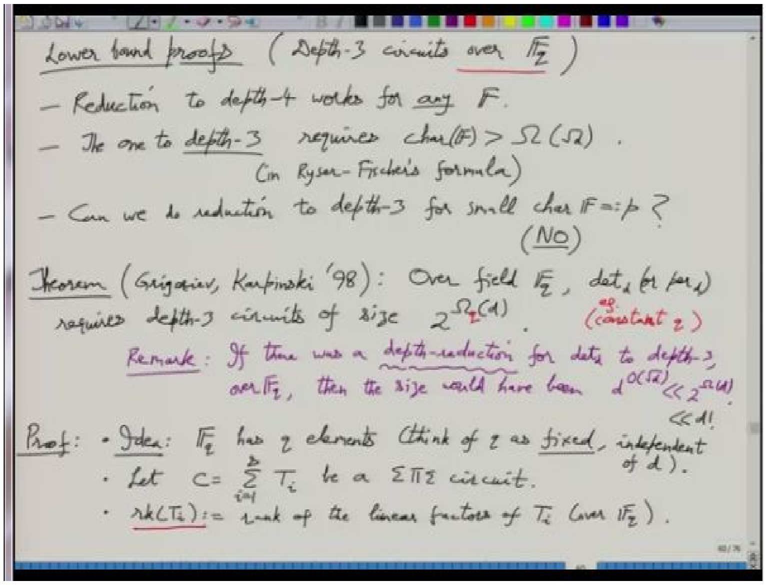
$$Q(-f^2)+\boxed{O(\varepsilon)} = \begin{pmatrix} -1/\varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}\cdot\underbrace{(Q(f)+O(\varepsilon^3))}\cdot\begin{pmatrix} \varepsilon^2 & 1 \\ -1 & 0 \end{pmatrix}\cdot\underbrace{(Q(f)+O(\varepsilon^3))}\begin{pmatrix} 1/\varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}$$

• Induction will now convert any formula to $\overline{\text{width-2-ABP}}$.

▷ $\left.\begin{array}{l} \text{formula} = \overline{\text{width-2-ABP}} \\ \text{formula} \neq \text{width-2-ABP} \end{array}\right\}$ though

Ok, so before the mid sem we finished this proof that formulas and width 3 ABP's are equivalent models. But you cannot reduce formula to width 2 although width 2 ABP you can convert to a formula. And interestingly, if you relax the computation definition to approximative complexity for approximative circuits or approximative models in general, then you can get even formula in width 2 ABP as the same model. But with this bar, so this denotes the border of your algebraic model. Any questions about that?

(Refer Slide Time: 01:17)



Lower bound proofs (Depth-3 circuits over $\mathbb{F}_q$)

- Reduction to depth-4 works for any $\mathbb{F}$.
- The one to depth-3 requires $\mathrm{char}(\mathbb{F}) > \Omega(\sqrt{d})$.
  (in Ryser-Fischer's formula)
- Can we do reduction to depth-3 for small char $\mathbb{F} =: p$?
  (NO)

Theorem (Grigoriev, Karpinski '98): Over field $\mathbb{F}_q$, $\det_d$ (or per_d) requires depth-3 circuits of size $2^{\Omega_q(d)}$. (eg. constant q)

Remark: If there was a depth-reduction for $\det_d$ to depth-3, over $\mathbb{F}_q$, then the size would have been $d^{O(\sqrt{d})} \ll 2^{\Omega(d)}$. $\ll d!$

Proof: • Idea: $\mathbb{F}_q$ has q elements (think of q as fixed, independent of d).
• Let $C = \sum_{i=1}^{s} T_i$ be a $\Sigma\Pi\Sigma$ circuit.
• $rk(T_i) :=$ rank of the linear factors of T_i (over $\mathbb{F}_q$).

Then we will start a new topic which is proving lower bounds. There are several theorems that prove interesting polynomials like determinant or permanent cannot be computed if you have a restricted model determinant of course, you know can be computed in VP computation of permanent is still an open question both algorithmically and in the circuit sense. It is believed to be not computable by circuits of small size.

But there is no proof yet, so the proves that I will present will be for restricted models , some resource will be restricted and then we will show that determinant, permanent are impossible to compute. Usually in these proofs, there will be no difference between determinant and permanent. The same proof will hold for both which will also be a bad sign that these proofs cannot separate permanent from determinant .

That is the current status, but there are many techniques that go into this. We will do that; the current result that we will discuss is also to do with depth reduction. we will in particular prove something about depth 3 circuits over finite fields . So in particular, we will show that determinant complexity in the depth 3 model over finite fields F_q, where think of q as a constant, q is not growing.

So field size is actually fixed to an absolute constant size, in this model determinant actually requires eventually bruteforce size which is $2^{degree} = 2^{d}$, what happens if you change the field to complex, what is the complexity of determinant in the depth 3 model? That was around $d^{\sqrt{d}}$. So that shows you a gap between the fields, so when the field is very small, then the determinant will need bruteforce size.

But if the field is large, then determinant can be computed by depth 3 which in particular means that depth reduction cannot be done to depth 3 for small fields . This actually will be saying something about the optimality of the depth reduction as far as the field is concerned which is a bit surprising. So which is surprising that coming depth 3 really requires the field to be large otherwise that prove does not work provably.

So background is we call that reduction to depth 4 always works. It works for any field F but as seen in the proof reduction to depth 3 requires at least last characteristic. It requires the characteristic of F to be bigger than the multiplicative fanins involved in depth 3. That was around $\sqrt{d}$ I think, was it, when you come down from depth 4 to depth 3 I think it was $\sqrt{d}$, $\Omega(\sqrt{d})$.

This was because of the Ryser-Fischer formula, so to apply this Fischer's trick converting $x_1 \cdots x_n$ to sum of powers of linear polynomials, or linear forms even. So that actually you get $1/n!$ there, we need the characteristics to be bigger than n which actually boils down to $\sqrt{d}$. Because that was the $\Sigma\Pi\ \Sigma\Pi$ model, the 2 Π 's were around $\sqrt{d}$.

we had asked this question can we do reduction for other fields for small characteristic. At least when the size is small, then we will give a proof that this cannot, this is impossible . So that is what we will show which will suggest that you have to be mindful of the field you cannot hope to could get to depth 3 without field assumptions. This seems to be smooth development, but the reduction the lower bond proof is actually very old; it is much older than the depth 3 reduction .

So this was done by Grigoriev Karpinski, so this will say that over the field, F_q det_d or per_d, well or as in this is and it holds for both. So determinant requires depth 3 circuits of size $2^{\Omega_q(d)}$, where the constants involved in Ω, Ω depend on q. So there is a q here that refers to the constants which are involved. So as long as q is constant, it does not matter that this is just $2^{\Omega(d)}$

But when q starts to grow with d then this result actually has no meaning, it is not interesting . So it is interesting only for constant q because the constant involvement is as you will see, it will be large the dependence on q will be large. But think of constant q and then it is fine or q may grow very slowly with d till that it will be fine but not for general q. So it is really a constant q result.

So before we go into the theorem proof, note that if there was a depth reduction, note that for permanent already you know a 2^d size circuit depth 3 by Ryser's formula, for determinant we do not know. The best we know is well we know because it is this for other fields we know now also, for permanent for small fields we do not know, do we know 2^d sized.

Originally when we had shown permanent is in VNP or determinant is in VNP that was for we dividing by some constant like $1/d!$. No, no not Newton formula that is much more advanced just the simple fact that determinant is in VNP just that fact or permanent is in VNP. So did we divide, probably that was not used, division was not used.

So for one of them we have this 2^d depth 3 circuit for any field. But anything better than that like for determinant better than 2^d that only came through depth reduction that did not come algorithmically. So if there was a depth reduction for det_d to depth 3 over F_q. Then the way our depth reduction result was the size would have been $d^{O(\sqrt{d})}$ which is much, much smaller than the above.

So this $d^{O(\sqrt{d})} \ll 2^{\Omega(d)}$, that would have contradicted this theorem. So in simple words the theorem rules out any kind of interesting depth 3 circuit for determinant or permanent 2^d is possible but nothing below 2^d although the monomials are around 2^d. So 2^d slightly smaller than d! which comes by definition.

So in this sense it is an optimal lower bound right this 2^d we would say the nearly optimal lower bound for depth 3 circuit size for determinant or permanent. So the proof will be long and it will be a long and hard road , it will take several lectures maybe we can cover in 2 lectures, but there are many ideas. Do you understand the statement and the comparison? All these lower bound proofs will be almost all of them will be fairly technical and several kinds of ideas will be thrown at you together with calculations, but that is expected.

We do not expect this to come cheap if it comes at all, here the basic idea will be well so what do you think would be the idea to prove such a thing. The determinant or permanent requires brute force sized depth 3 circuit, what do you know about depth 3 circuits to be able to wish such a big result, will you know that it is a sum of products of linear polynomials.

You have to prove something, if you can prove some strong result about just product of linear polynomials in a way that generalizes to the sum also, did it we can find such a property then maybe you can use that property to prove the determinant and permanent are hard . So that property would be related to rank of some associated vector space. So for a product of linear polynomials, we will associate a vector space.

And we look at the rank of that and this will actually generalize naturally also to sum of products of linear polynomials. So it will be then rank of a depth 3 circuit rank of an associated vector space. So we have to show that this rank for the depth 3 circuit is large or small. So you want idea would be to show that this will first you have to define this vector space and it is then rank will be defined.

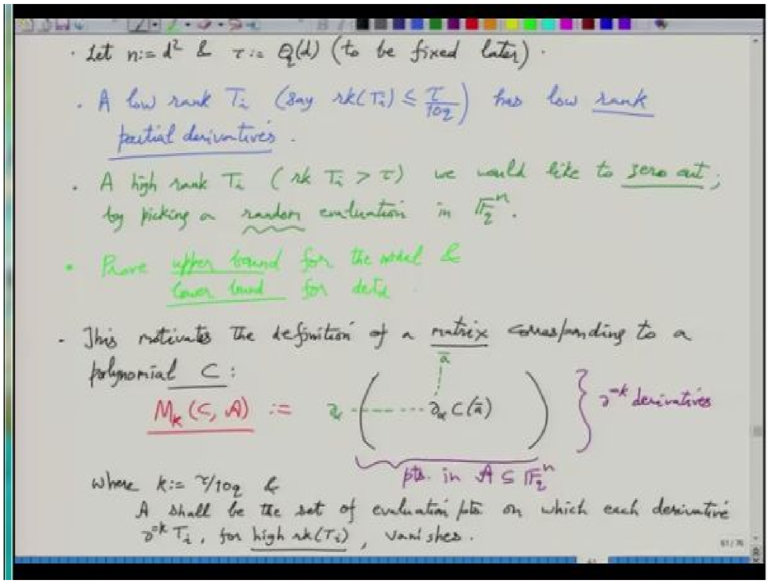
And you have to show that for those computational model it is small, while for the target polynomial it is large. And if there is sufficient gap then asymptotically you have proved you have a lower bound you have an impossibility result. So in somewhere magically this q being constant is used, so F_q will be used in this implementation in a very crucial way. So F_q has q elements think of q fixed and independent of d that is what we mean by fixed with respect to d it is constant.

To define determinant or permanent, you only need d and F_q. So think of these 2 as independent parameters where q is fixed and d is growing, but d grows to infinity, d think of d as the input size and let C be the model. So $C = \sum_{i=1}^{s} T_i$ where each T_i is a product of linear polynomials. So, what could be the notion of rank of T_i, so for now I am not defining the vector space that we will do later.

Let us just define rank to be rank of the linear polynomials that define T_i, whose product is T_i. So rank of T_i is just rank of the linear factors over the field F_q. When you look at a product gate, where this rank is very small wait what does it mean, this rank is r. That means, in some sense T_i is r variate because these so look at the basis of the linear factors r of them are independent but other ones actually just a linear they are just a linear combination of these $l_1 \cdots l_r$.

It is as if T_i is r variate, where the variables are $l_1, .., l_r$. So if r is very small then T_i is a low variate polynomial, if r is large then T_i is more of a general number of variables. So, small rank T_i's are actually in this sense, they will not pose a problem they seem to be quite weak when the rank of T_i is small. What should we do when the rank is large, those are the T_i's where there is more complication. We are proof will always be in these 2 cases T_i's with low rank and T_i's with high rank.

(Refer Slide Time: 20:00)



- let $n := d^2$ & $\tau := \Theta_q(d)$ (to be fixed later).
- A low rank T_i (say $rk(T_i) \le \frac{\tau}{10q}$) has low rank partial derivatives.
- A high rank T_i ($rk\ T_i > \tau$) we would like to zero out; by picking a random evaluation in $\mathbb{F}_2^n$.
- Prove upper bound for the model & lower bound for det_d.
- This motivates the definition of a matrix corresponding to a polynomial C:

$$M_k(C, A) := \begin{array}{c} \\ \partial \end{array} \underbrace{\begin{pmatrix} & \bar{a} & \\ \cdots & \partial_x C(\bar{a}) & \\ & & \end{pmatrix}}_{\text{pts. in } A \subseteq \mathbb{F}_2^n} \left.\vphantom{\begin{pmatrix} \\ \\ \\ \end{pmatrix}}\right\} \partial^{=k} \text{ derivatives}$$

where $k := \tau/10q$ &
A shall be the set of evaluation pts on which each derivative $\partial^{=k} T_i$, for high $rk(T_i)$, vanishes.

Number of variables is d^2, d is the degree of the determinant d × d determinant and let us use a parameter τ it will be around d constants depending on q will make τ specific towards the end. The cases are like this, a low rank T_i say rank of T_i is sufficiently small, so (τ / 10 q). So this has low rank partial derivatives. Partial well when you look at the partial derivatives of a low rank T_i since T_i is just a polynomial in $l_1, .., l_r$ where r is small.

If you look at the set of all the derivatives right, their rank will also be the rank of that space and will also be small, where small is really being overused. I mean r small and the dimension of this derivative space are different. We are talking about low in 2 different ways, but it will be low compared to when r is large. If r is large then, this partial derivative space has actually much bigger dimension.

We will ultimately be looking at the vector space will be the vector space generated by the derivatives that is the let us say that is our measure. So for the algebraic model we will define a measure which will basically be the dimension of the partial derivative space. So when r is when rank of T_i is small then this measure is small and when this is high, so a high rank T_i (rank of $T_i > \tau$).

Naturally it is partial derivative space will be large, so is there a way to force this to become small. I mean, we are looking at a polynomial, where the derivative space is large. What should you do to make it small, you kill some of the extra variables. How do you kill them, you kill them by evaluating them at points. To handle these, what we will do is, we would like to 0 them out.

High rank T_i we would like to just completely 0 out which will be done by picking a random evaluation. Why would a random evaluation work to 0 this out, to see this in the simplest possible we think of $x_1 \cdots x_n$. So what is the probability that this product vanishes, you just have to pick 0 somewhere. So what is the chance that $x_1 \cdots x_n$ is picked to be 0, so x_1 is picked to be 0 with probability 1 /q.

Accordingly you can calculate the probability and as n grows you will get more opportunities to set the monomial to 0. This is the intuition that these high rank T_i's since they are have many they depend on kind of many variables. So this $l_1 \cdots l_r$ this r is big, since it depends on r many variables where r is large, you have more chance to actually just 0 this out by random substitutions.

This chance is actually smaller if it had lower fewer variables. But those cases are anyways easy for us because those T_i's will have low rank partial derivatives. And on the other side when the variables are many, then actually by evaluating them randomly, we will set them to 0. So this green things will be gone only the blue will be left and so we would have essentially proved that in some sense, the circuit has low dimension of the partial derivative space associated to it.

This will be the upper bound on the measure of the model. So and then once you have done this, we have to show the opposite for determinant that if you do this rank of partial derivative space analysis for the determinant that comes out to be large. Determinant cannot be computed by this model, that is the overall proof idea, prove upper bound on the model.

And lower bound for determinant, so that will be the overall strategy but implementation of this will really go into gory details. This is not an easy proof, but the idea, do you understand this overall this very vague idea. So as you basically we are associating defining a measure, so the measure for determinant polynomial will be show is large and the same measure for the circuit model will show a small.

So the measure will distinguish determinant permanent from the model So this motivates, let us do this now explicitly This description motivates the definition of a matrix corresponding to, actually it will be for any polynomial, to any polynomial. We want to define a matrix for a given polynomials C, this matrix is essentially we will just look at the derivatives of C make them r rows of the matrix.

And in every row that particular derivative will just be written as, what, we will write the evaluations of that polynomial picking some points in the space $F_q{}^n$. So the matrix will be basically the rows are indexed by derivative operators and the columns are indexed by evaluation points. So because the above proof suggests both of them, so the blue part suggests that we have to use derivative operators and this dark green part suggests evaluation points have to be used from F_q to n.

So we will use both of them to define the matrix. This is the following matrix for C : $M_k(C,A)$ is the subset of points. So as promised, the entry here is $\partial_\alpha C(\overline{a})$ where the row is derivative with respect to α and the column is point $\overline{a}$. So the rows are basically k order derivatives, this 1 think of α as the exponent vector.

So variables are $x_1......x_n$, so on the variables there is an exponent vector $\overline{\alpha}$. And you want to only look at k order derivatives which means that the sum of these exponents is exactly k, k also will fixed later. And the columns are they denote some points in the space. So points in some subset A of $F_q{}^n$ just looking at this definition you can see that the space has to be finite at least the subset A should be finite, why should q be, why should the field be finite.

So that actually comes from this intuition of $x_1 \cdots x_n$ vanishing. So if the field is quite large, then $x_1 \cdots x_n$ will not tend to vanish because the chance of hitting 0 in a big hay of points is negligible. So that is the problem, so you do not want the field to be too big, if it is too big then you $x_1 \cdots x_n$ is negligibly I mean it vanishes with a negligible probability. So the probability looks interesting only when the field is small.

So in that finite space we will be picking some subset A and then this matrix gets fully defined. This matrix is just derivatives at points, that is all we are doing. Where we will take k:=($\tau/$ 10q) it is once we have fixed τ this will also get fixed, so this k is are measure of basically low rank product T_i. So we had used this before also, this will be the k and A shall be the set of evaluations or evaluation points on which each derivative for high rank T_i vanishes.

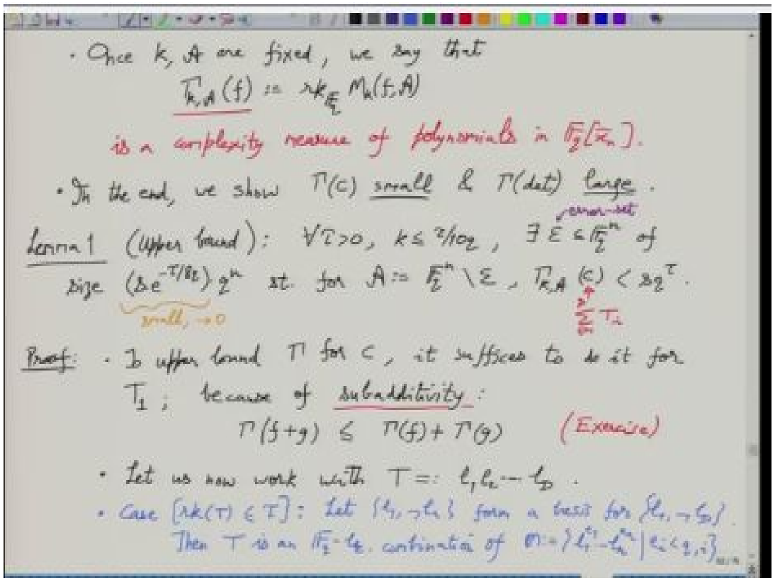
We want this points $\overline{a}$ to be such that this high rank T_i's do not contribute in the calculation. But there will be more restrictions on A which we will put as we go through the proof. No, so, in the proof we will be picking them we will show the existence of A such that upper bounds and lower bound properties hold. But in the end, A will be it will be a defined subset it is not a probabilistic object.

So probability will be used in the proof to show the existence of it, I mean the proof can remain existential it is a lower bound proof. So existential proof is fine,. You want upper bounds and lower you want a measure, the measure will be defined by A because this $M_K(C, A)$, is the object you are measuring refer to again yes, exactly right.

So this measure will be I mean we could say safely at this point that the measure will be existential. Because at least in the proof, we will not work out the details what exactly A is, we will be happy with an existential measure because as long as a measure exists that separates model from determinant we are fine. You only want a lower bound proof there is no demand for an algorithm here.

We will be preferring probability in the proof but that may not be very important mathematically just for convenience I think right.

(Refer Slide Time: 36:58)



- Once k, A are fixed, we say that $\Gamma_{k,A}(f) := rk_{\mathbb{F}_q} M_k(f,A)$ is a complexity measure of polynomials in $\mathbb{F}_q[\bar{x}_n]$.
- In the end, we show $\Gamma(C)$ small & $\Gamma(\det)$ large.

Lemma 1 (upper bound): $\forall \tau > 0,\ k \leq \tau/10q,\ \exists \overset{\text{error-set}}{E} \subseteq \mathbb{F}_q^n$ of size $\underbrace{(3e^{-\tau/8k})}_{\text{small}, \to 0} q^n$ s.t. for $A := \mathbb{F}_q^n \setminus E$, $\Gamma_{k,A}(\underset{\sum_{i=1}^{s} T_i}{C}) < 3q^{\tau}$.

Proof: • To upper bound Γ for C, it suffices to do it for T_1; because of subadditivity:

$$\Gamma(f+g) \leq \Gamma(f) + \Gamma(g) \quad \text{(Exercise)}$$

- Let us now work with $T =: \ell_1 \ell_2 \cdots \ell_D$.
- Case $[rk(T) \leq \tau]$: Let $\{\ell_1, \ldots, \ell_r\}$ form a basis for $\{\ell_1, \ldots, \ell_D\}$. Then T is an $\mathbb{F}_q$-lin. combination of $M := \{\ell_1^{e_1} \cdots \ell_r^{e_r} \mid e_i < q, i\}$

So once we have fixed k, A, K and A are fixed or defined we say that we have a measure on polynomials f which is the rank of this matrix over the finite field.We will say that this Γ which is the number associated with the matrix. This matrix is huge exponentially large, but it there is a single number attached to this which is Γ(f), so this is a complexity measure of polynomials.

So it will be defined for all the polynomials with coefficients in F_q and again, so in the end we want to show that Γ(C) is small and Γ(det) is large. So which means that c cannot be equal to determinant that is the strategy, First we will do the small part, later on we will do the largeness part. Let us first do this upper bound or the smallness property.

Already this lemma 1 showing that Γ(C) is small that already is an achievement because you have identified a measure that is small for the circuits for depth 3 circuits, this is a measure that is small already finding such measures is a challenge. By definition 1 measure is the size but that is not very helpful because that is exactly the thing you want to prove is large for determinant.

You have to find an alternative measure and finding this alternative measure is a challenge. It requires creativity to come up with an alternative measure that is small and that is not size. So it should be more mathematical, what we will show is that $\forall\ \tau > 0,\ k \leq \tau\ /10\, q$, $\exists$ subset E of $F_q{}^n$ of size almost everything. So the whole spaces of size q^n and this subset E is of size q^n time something this something will I will want to make it small.

So it is not nearly everything but much smaller, so the basically subset E over q^n is quite small it actually is tending to 0. So we identify a small set E in the space such that the definition of A will be the complement, so think of E as the error set, these are the bad points.Apart from these bad points or these error points A is almost everything in the space and for this you use the measure.

So $\Gamma_{K,A}(C) < sq^{\tau}$ c is your depth 3 circuit is the $\sum_{i=1}^{s} T_i$ So the measure on $\sum_{i=1}^{s} T_i$, well I am saying small, but it seems to be sq^{τ} , so it will really depend in the end, what does small and large mean. But let me continue with highlight that this is small, so because it is certainly less than it is a inequality in the correct direction.

So we will show that $\Gamma\,(\mathrm{C}) < sq^{\tau}$ later on we will show that for determinant this is far more than sq^{τ} . So the 2 things will be enough to finish the proof, actually looking at this part I want this to be very small, in fact tending to 0. So this suggests that in the end τ should be picked as let us say $q\ log(s)$ or $8q\ log(s)$. So you pick that if you pick τ to be more than that, then this multiplier is very small.

And correspondingly the set A is nearly everything the only thing changing is n or d that is the only variable allowed to grow sure. So $q\ log(s)$ means that it is growing s is the parameter it has nothing, s is a function of d, s is a function of n, s is the size of the depth 3 circuit. I mean s is at least d^2 because d^2 is a number of variables well ok maybe there are some inaccuracies because s I am saying is the number of T_i 's.

But I mean truth be told s will be at least n you can show that you cannot there is an easy proof that if you use less than n product gates. You cannot the sum cannot be computing determinant or permanent those are easy proofs. So even in this definition s will be a function of n growing function of n, it cannot be too small and then τ is a growing function of n. But what may not be clear is, why we need q to be constant that I think is not visible here. But we will see that at some point ok any questions on the statement.

Let us go through this, to upper bound Γ for c it suffices to do it for T_1 is that clear. If you show that the measure on T_1 is q^τ then we get that measure on C is atmost s times that why oh sub additive I have to define well first of all. So if you know that rank of T_1 and rank of T_2, well if this measure gamma actually, if $\Gamma(T_1)$ and $\Gamma(T_2)$ you show to be small, then you can deduce that $\Gamma(T_1 + T_2)$ is also small this happens.

Because if you look at the definition of gamma it is rank of a matrix. So you have basically a matrix for T_1 and a matrix for T_2 and both are low rank. So the sum is also correspondingly low rank. We will only work with product then which will simplify things by a large amount. So this is because of what I just said is called in one word sub additivity, why is it not called additivity.

Just because of the unfortunate fact that rank of A + B maybe matrices A and B rank of A + B will be much smaller an example is exactly, A and -A is a good example. So but we only need sub additivity we do not need strong additivity. So, because if you look at f and g polynomials $\Gamma(f+g) \leq \Gamma(f) + \Gamma(g)$. Let us leave this as a linear algebra exercise.

Let us now work with $T_1 := l_1 l_2 \cdots l_D$ and we have so now you want to study $\Gamma(T_1)$ basically. And we will this proof will do in 2 cases as already suggested in the original idea that 1 when $l_1 l_2 \cdots l_D$ their rank of the linear polynomials is small versus when it is large. So in both the cases we will show that the measure is upper bounded by q^τ.

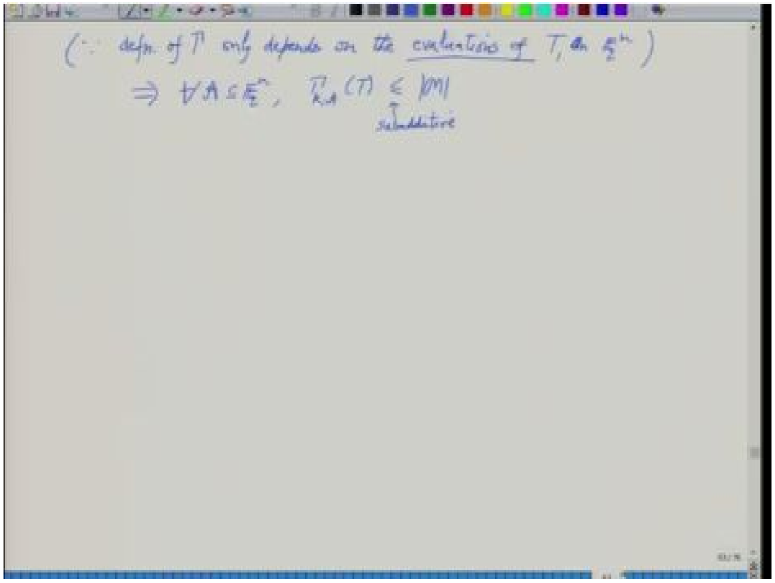
So far low rank it was blue, so the case when rank of it is not use T_1, so T , so rank (T)$\leq$ τ in this case. Let $\{l_1....l_r\}$ form a basis for $\{l_1....l_D\}$ then T is an F_q linear combination of these products, $\{l_1^{e_1}l_2^{e_2}\cdots l_r^{e_r}\}$ $e_1, e_2, .., e_r$ are the exponents and how big can e_i is be, less than yes. But if I allow that, then your upper bound will come out to be d^r it will come out to be d^τ which will be which is hopeless because Q is constant you want q^τ not d^τ .

So that is a problem even in this case when the rank of T is small just because the exponents in $\{l_1....l_r\}$ may be large. There are just the set M has too many monomials, has too many products. So what should you do, what should be the correct definition of M, so you actually have to look at e_i 's $mod\, q$, the so e_i's will be actually less than q. Why is this allowed, what happens if you have $l_1{}^q$, will I say that $l_1{}^q$ is l_1 . Is that fine?

So the definition of Γ is based on the matrix and in the matrix M_K the you are actually not looking at the entries are not polynomials, they are just evaluations. And the evaluations are happening in the space $F_q{}^n$. So you should not think of l_1 as a linear polynomial, it actually when it is used in the matrix. It only gives field element and for the field element raise to e_1 you will just get l_1

So that actually is one important part which was not discussed in the basic idea that when the rank is small, you also have to use the fact that the exponents are small. So that the overall monomials or these products in M are few, so now you can get q^τ , that is the upper bound.

(Refer Slide Time: 53:35)



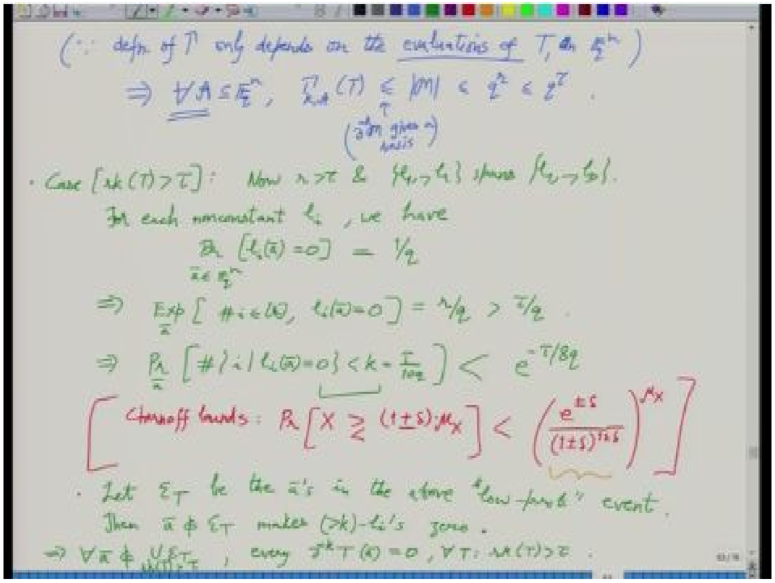
(∵ defn. of Γ only depends on the $\underline{\text{evaluations of}}$ T, on $\mathbb{F}_q^n$)

$\Rightarrow \forall A \subseteq \mathbb{F}_q^n,\ \Gamma_{k,A}(T) \underset{\text{subadditive}}{\leq} |M|$

So because since definition of Γ only depends on the evaluations of T, in $F_q{}^n$ or on fine. So what this means is that irrespective of E that you pick. So irrespective of measures the $_{K,A}(T)$ is upper bounded by the size of M this should be clear. Because of number of columns in the matrix no not number of columns. How do you see this M is the set of possible products of $l_1......l_r$ all possible products.

T is actually a linear combination of these monomials these products. So I guess you again use the sub additivity property. And so this is again the sub additivity property sub additivity again. This is again the sub additivity property that is why you get this. So sub additivity will reduce to just this product and of this product is what, is it 1. So you differentiate it k times sure, that is needed.

(Refer Slide Time: 56:24)



($\because$ defn of Γ only depends on the evaluations of T, on $\mathbb{F}_q^n$)

$\Rightarrow \forall A \subseteq \mathbb{F}_q^n,\ \Gamma_{k,A}(T) \underset{\uparrow}{\le} |M| \le q^r \le q^\tau$.

($\partial^{=k}M$ gives a basis)

- Case $[rk(T) > \tau]$: Now $r > \tau$ & $\{l_1, \ldots, l_r\}$ spans $\{l_1, \ldots, l_D\}$.

For each nonconstant l_i, we have

$$\Pr_{\bar{a} \in \mathbb{F}_q^n} [l_i(\bar{a}) = 0] = 1/q$$

$\Rightarrow \underset{\bar{a}}{Exp} [\# i \in [D],\ l_i(\bar{a}) = 0] = r/q > \tau/q$.

$\Rightarrow \Pr_{\bar{a}} [\# \{ i \mid l_i(\bar{a}) = 0 \} < \underbrace{k = \tfrac{\tau}{10q}}] < e^{-\tau/8q}$

[Chernoff bounds: $\Pr [X \gtrless (1 \pm \delta) \mu_X] < \left(\frac{e^{\pm\delta}}{(1 \pm \delta)^{1 \pm \delta}} \right)^{\mu_X}$]

- Let E_T be the $\bar{a}$'s in the above "low-prob" event.

Then $\bar{a} \notin E_T$ makes $(\ge k)$-l_i's zero.

$\Rightarrow \forall \bar{a} \notin \bigcup_{rk(T) > \tau} E_T$, every $\partial^{=k} T(\bar{a}) = 0$, $\forall T: rk(T) > \tau$.

Even with the sub additivity that is needed that M is this $\partial^{=k}M$ a basis, it gives a. And so the basis size cannot exceed size of M, is that clear and size of M is simply q^r which is q^τ . So just what we wanted to show, next case is high rank case. Now, r > τ and $\{l_1 ...... l_r\}$ spans $\{l_1 ... l_D\}$, so we cannot use the above proof. Because there's proof in blue, use the fact that $q^r \le q^\tau$ which now does not hold much more than q^τ now.

So this is the part where we will use the specialty of A, this subset A which will require definition of the set error set E. So till now this blue part holds for any, so that luxury will not have any more, so we will actually fix it probabilistically. So for each non constant l_i we have, so we I mean as discussed in initial idea we actually want to just kill T. So when T is evaluated on points in A then it should just vanish, that will boil down to studying whether l_i vanishes or not.

So, $\Pr_{a \in F_q{}^n}\ [l_i(\overline{a}) = 0] = 1/q$. Because l_i is non constant hence nonzero also, did what is the proof of this exactly. So you know that there is at least some variable in the support of l_i say x_1 appears. So just keep x_1 free and fix everything else and that will put that will fix x_1 , so except this value of x_1 , l_i will not vanish.

So the others were fixed randomly, so ultimately the probability comes out to be only 1 /q. So here we are using the fact that F_q is a finite field s actually previously in the blue part also we used. This when we reduce the exponents |q| there also used the fact that evaluations will happen from in this finite field. So this immediately gives you that expectation over choices random case number of i's such that $l_i\ (\overline{a})$ =0, what is this expectation, how many l_i 's will vanish.

There are r many l_i's, so r / q which is greater than τ / q, . So now r $>$ τ and that actually is what helps us in saying that if you have a random point chosen then expectation of something vanishing is pretty good well or the probability of something vanishing is pretty good and the expectation because the expectation is good, expectation is far more than 1.

So let us look at this probability then, so the probability over $\overline{a}$ that the number of i's for which a $l_i\ (\overline{a})$ vanishes, this number is smaller than k. So basically what is the chance looking at the expectation what is the chance that this count on i a for which l_i is vanishing is less than k. So Chernoff is the correct bound. If you have not seen it, then let us write down the statement of Chernoff bound.

So Chernoff bounds the probability of a random variable X which is a sum of 0-1 random variables X_i's, so you do not need the sum here actually. I just need mean and, I just need the mean of X which is μ let us say. So probability that it is random variable is far away from the mean. So this is μ_X mean of X probability it is some multiple ahead of the mean or it is sufficiently smaller than the mean, so minus.

Both these events or any of these events, that exists either on the side of the mean or on the left side of the mean, these probabilities respectively are of small. So how small, it is like $e^{\delta\mu}$, let me give the expression first:

$$\Pr\ [X \gtrless (1\ \pm \delta) \cdot \mu_X\ ]\ < (\frac{e^{\pm\delta}}{(1\pm\delta)^{1\pm\delta}})^{\mu_X}$$

So in this expression here,we are anyways interested in the things happening below the mean. So in that case what this is saying is that the probability that the random variable is much smaller than the mean.

This is kind of $e^{-\delta\mu}$, $e^{-\delta\mu}$ is a fraction, so the probability is very small. So if you keep increasing δ or this $\delta\mu$ is the difference. If you keep going away from the mean backwards below the mean then that the probability will be also fall in accordingly to the minus difference. And for the $1 + \delta$ thing also this probability can be made smaller but anyways, you would not be needing that.

But Chernoff bound basically gives you bound on the on both sides of the tail being on the tail sides is the probability is very low. You can read the proof of this or you can try to prove this and easy ways just read the proof on Wikipedia will not take much long . Coming back to this green line, probability that number of i's such that l_i vanishes at randomly chosen point $\overline{a}$ is the number of i's is smaller than k this probability is small, this happens with low probability.

I mean, why are we proving such a strong thing, in the initial proof idea I said that well one of the l_i's will vanish with high probability. Why do we want here that, why are we looking at whether k of the l_i's are vanishing or not. It is not 1 but we are now actually looking at k of the l_i's, why do we do that. Because we are using derivative operators of k order, suppose there are (k + 1) l_i's that are vanishing.

When you will apply the k order derivative it will continue to vanish, we actually want the. We not only want T to vanish but since we have used k order differentiation, we actually want the derivatives to vanish. So derivative vanishing will need that many repeated zeros in T. We are actually this is why we are comparing number of i's with k and since this is low event, low probability event.

We should now be able to reduce what we wanted. Let ε_T be the $\overline{a}$'s in the above low probability event. Then $\overline{a} \notin \varepsilon_T$ makes (>k) l_i s zero, ε_T are these bad $\overline{a}$. But other than the bad a bars the k + 1 of the l_i's will vanish, for all a bar that are not in the union of ε_T and these are the T's of high rank. I mean for, there are many, if you look at the level of the circuit c there were many T_i's.

And the look at the ones for which the rank of T_i's is more than tau. So for those take the union of their bad $\overline{a}$ this is still a small set,. So outside this if you pick $\overline{a}$ then every k derivative vanishes. Every, $\tau = k$ is not a single derivative operator but all these possible k orders derivative. Pick any k order derivative and pick any such T and any point $\overline{a}$ that is not bad, this is 0. Let me stop here, we will finish the calculations tomorrow.