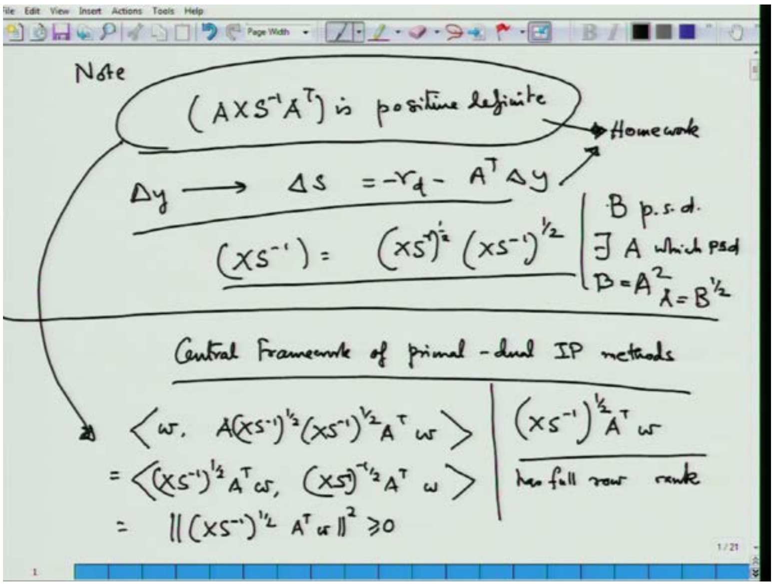
Convex Optimization
Prof. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 30

In the last lecture, we had spoken about how to compute the newton steps, that is the delta x, delta y, and delta s for solving the system, rather for solving the newton system or the system the system or equations associated with the relaxed KKT condition, this is what we had already spoken of.

(Refer Slide Time: 00:53)

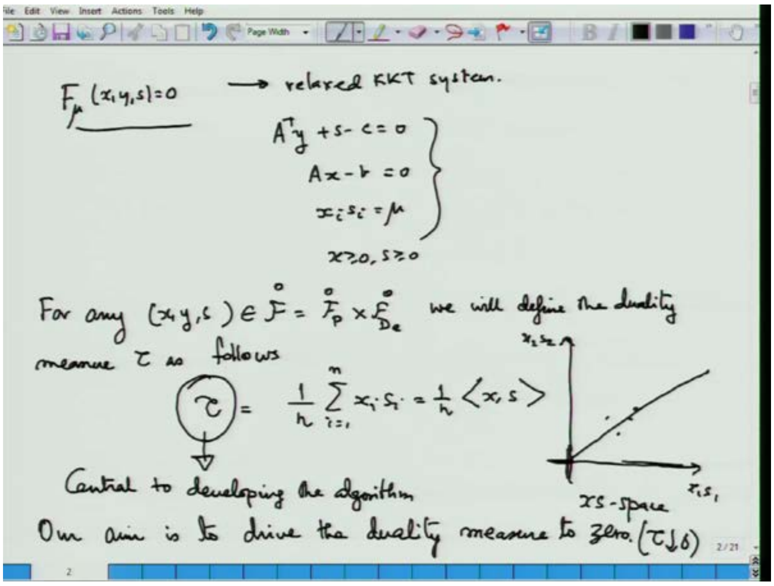


Now, what is important for us is that to know the the important step for us to know that this is positive definite so, it has an inverse, because the inverse of this was used. Now, I had told you to figure it out as an complete the rest of the computations, whether how to prove that this is positive definite.

So, one thing is that you have to first show it is positive definite look, this is I I can write XS inverse in this form, because given any positive definite matrix positive PSD any PSD matrix. So, Positive Semi Definite matrix there always exist a matrix A, which is also positive semi definite, such that B is equal to square of A or A is written as square root of B. So, there is nothing like B to the power half, but just it is the symbolism. So, here I can write this into these two parts, and I have in this way, and I put it here in this place. And then I take the transpose this is the symmetric matrix, so transpose is the same thing then A A A B transpose is B transpose, A transpose, and this is how the things happen.

So, now this is nothing but the norm square and you have this greater than equal to zero, what you can show is that this matrix this is anyway a full rank matrix into A transpose this has full column rank, has a full sorry has a full row rank sorry not column rank, row rank is same as column rank, but full row rank. So, once you can show that these are the full row rank. So, this equal to if this is equal to this is if and only if this W becomes equal to 0 if this has full row rank and it has if it is simple to show that it has a full row rank and that is exactly what you have to show to show that this matrix is positive Semi Definite. Now, once we look into the main primary dual framework

(Refer Slide Time: 03:18)



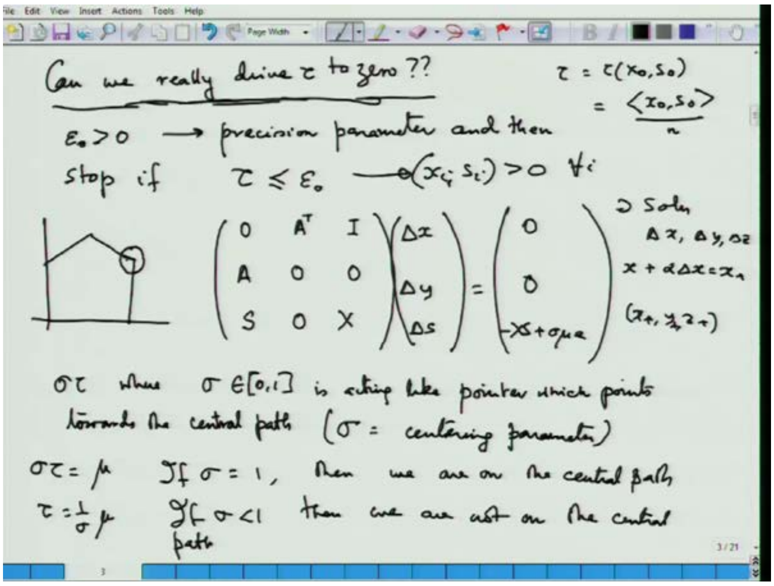
Let us, go to the story once again you know we have to solve this equation. We have to solve this equation that is the relaxed KKT system. So, you know that so, the relaxed KKT system if you have forgotten let me write down once again. So, this is the system that I have we are intending to solve of of course, we are intending to have this, but this thing automatically means that I will have this strictly bigger than z 0 this. So, as I told you that it is very difficult to really satisfy this equation.

So, you want to find some solution which will approximately satisfy this, this may be that x i s i for every I need not be equal to mu, but some number very near mu that would be enough for a purpose. So, for any (x, y, s) which is in F naught which is nothing but F naught p cross F naught D e for any any such thing we will define the duality measure, see what the the finals approximate solution that I will get will not give me x i s i is equal to mu, but they can then I will try to see on the average what is the value taken up by the product.

So, we will define the duality measure tau as follows,: tau is equal to 1 by n summation x i s i is equal to 1 to n or inner product so, this same as a inner product. Now, this tau is central to developing the algorithms, because then solutions of the approximate solutions of these equations will not give me x i s i equal to mu. So, they would not be exactly on the central path. So, what I get is instead if this was my central path what I will get is something here nearby. So, I want to look into take those x and s and try to see. So, I am looking for let me look at the whole thing in the x s space which is much more simpler x 1 s 1, x 2 s 2.

Now, any point on the central path would be projected back on this path in a x in the x s space x into s space. So, unless the point x 1 and s 1 or a x 2 s 2 x 3 s 3 whatever that x and s they lie on the central path x s 1 be equal to mu. So, the x s point would not be on the straight line. So, it is somewhere may be here, may be here and may be here. So, then what is the corresponding average value that point is satisfying that is what we call the duality measure, our aim is to drive the duality measure to 0. So, our aim is to drive tau down to 0, that is do this now, the whole all the algorithms what we will do we will plan to drive tau to 0.

(Refer Slide Time: 07:44)



But can we really drive tau to 0? That is the question that we want to answer. So, what we do at every step, we solve this approximate system approximately get the x i s, x i s i and try to develop the algorithm in such a way that every step at every each and every step, the mu values sorry the tau values would continuously decrease. Now, can we really get finally, reach up to 0 the answer basically is not really true, because it may not reach you might not reach 0.

So, what we essentially want is that we fix up a small epsilon naught greater than 0 which is called the precision parameter and then then stop the algorithms, whatever algorithms you have stop, if your tau so, that x i s i of course, they are strictly bigger than 0 you you will stop the algorithm there. Now, what you would get is an x i s i which is whose product x i s i where both of them are strictly greater than 0. So, will you will you really get a point on the boundary may be not you have not got a point on the boundary.

So, what you have done here is that if this is my so. So, you have stopped someone very near a boundary point and then there are certain approaches which will not discuss in detail here, you can move on to the boundary you, can basically if this is not the B F S you can move on to the next B F S, you can know that if I excise this is my where I had stopped. Let me go to on to the next B F s by taking of m minus n points I put 0, then what what is the solution what what is the value of x is. So, you are you can find so, you what you do you find the near there is a technique of finding the nearest vertex to that point and that is that is the required solution.

So, but all of these works in polynomial time means, all these work within a reasonable amount of time. So, what I am going to do here is to solve an equation of this form. So, is we are trying to solve an equation the in exact heraldic condition, but we know that it is not possible to find mu e in mu every time, for every x i s i cannot be equal to mu every time, for every i so, what we do is say I do not care I take the duality measure, I I take some tau which is a there is a initial starting tau. The starting point that you have take take is duality measure and then so, if you have x, x naught s naught, you you take tau equal to now basically tau depending on say x naught s naught. So, this will be your starting tau tau.

So, then you use that tau in the equation and so, for tau you do not you cannot have exactly tau exactly for every x i s i need not be tau. So, what you do you put in a centering parameter basically instead of putting mu epsilon in the equations what you will have is, you will have x s e minus sorry sigma tau. So, you have not bothered this sigma tau is acting like like your mu, the sigma tau is acting exactly like your mu. So, this is acting like your mu here, the sigma tau the sigma tau where sigma is between 0 and 1 is acting like your mu.

So, I cannot find everything exactly equal to mu, but so I am computing the system that everything is not exactly equal to mu, but some some fraction of the average, the fraction of the duality measure. So, instead of mu using mu I am using sigma tau. So, when sigma is equal to 1, then tau is exactly my mu and then sigma is 0 of course, then I am I am the exact solution. So, sigma is lying in between and that is not exactly mu, so it not giving me a point exactly on the central path, but somewhere as we showed that nearby. So, this is the basic framework, the sigma this term this is acting like a so called so, sigma is called an affine scaling direction or is acting like a pointer which points towards the central path.

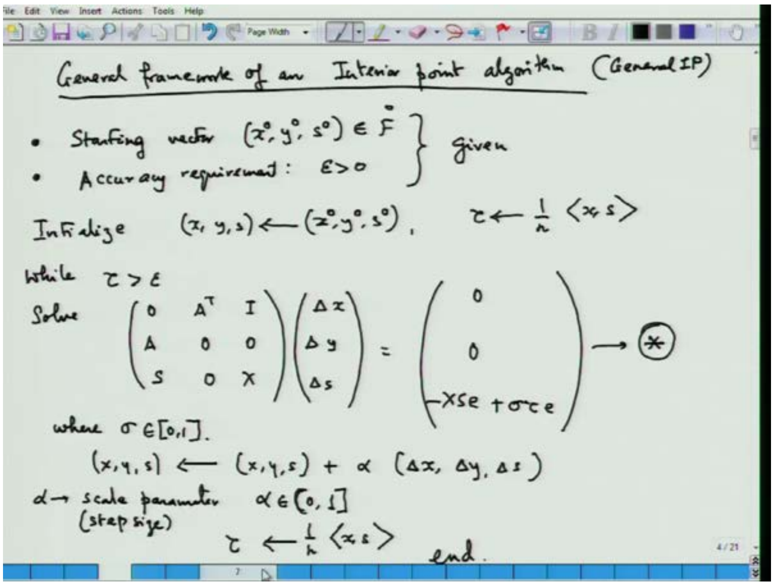
So, there put sigma tau is equal to mu. So, my sigma is 1 by times mu is not exactly mu, but tau is 1 by sigma times. So, mu is some fraction of this tau. So, I am exactly not on the central path, but somewhere else that is exactly what we are trying to say. So, this sigma tau, this sigma is 1 then at tau is mu then I am on the central path if sigma is not equal to 1, if sigma is equal to1 then we are on path, we are on the central path, if sigma is strictly less than 1 then we are not on the central path. So, this called an affine scaling direction.

So, whatever solution we have here of this system, suppose we call this solution we can write this as x sigma tau. So, whatever solution we have that is basically, there is a mistake here I should have written, just have to rewrite this a bit I think there is a small I have forgotten the Newton step, this is equal to of course, this is 0 0, because we have taken our points are all coming from my starting point also be an interior point x naught, s naught, f naught is belonging to f naught, the interiority conditions holds.

So, any solution delta x, delta y and delta z from there you can create a new solution that is adding delta x to x k, delta y to y k, delta s to s k by putting in certain controls over here, that is you for example, you take x of x plus alpha delta x is my delta x plus the new one new new thing. So, whatever you get this x plus y plus z plus which has solutions to the final solutions and after you get the delta x this is the new point, that I have these points are pointing in the direction of the Newton solution, pointing in the direction of the central path right. So, this is called sigma is also sometimes called sigma is also called the centering parameter.

So, it is trying to force the points to go towards the central path and keep them near the central path. So, let me write down the general framework of an IP algorithm. So, that will give you a fair idea of what we had just discussed.

(Refer Slide Time: 16:29)



So, the general framework of an IP algorithm of an interior point algorithm so, starting vector so you start with some we assume the starting vector, is element of f naught that is given to me, accuracy requirement I need, there is a precision, these are given initiation of the algorithm. So, we will start with initialization step that is you first initialize x, y, s with the value x naught, y naught, s naught starting values.

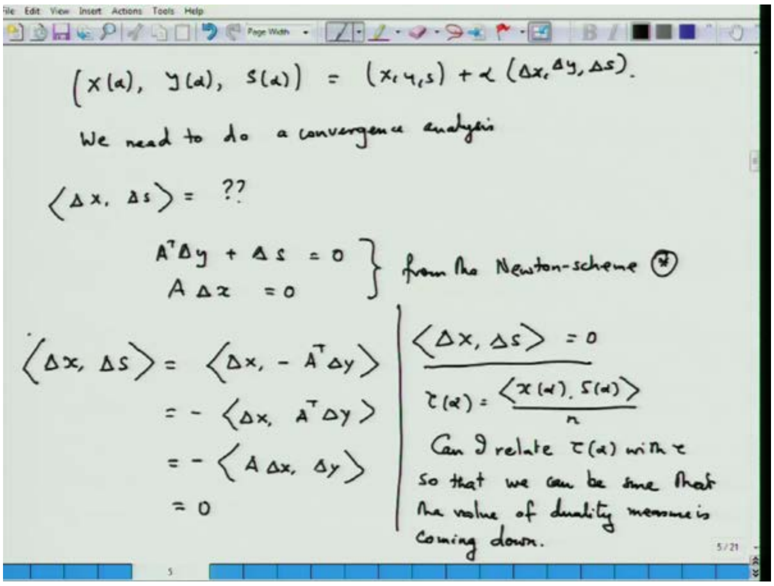
And compute the initial tau which is 1 by n, x s in a product. While so, we will have a while do loop, while tau is bigger than e epsilon, solve and you know how to solve it if if I put sigma tau equal to mu we have already showed you in the last class how to solve this system or equation how to get delta x, delta y, delta s. So, you can just write a computer program simply behind which which will solve it out somehow you can try it at home by writing a computer program, but of course, you have to take matrices of smaller sizes, because you know unless you get just you guys are working on a PC there may be lot of overflows.

Because when you try to take the inverse inverting is the very difficult operation is a very costly operation a machine. So, I am telling you from my own experiences of trying if you trying to take slightly larger matrices try to take, very small medium size matrices here 2 by 3, 3 by 3 and try to check it out. So, this is what you need to solve where sigma is the centering parameter or the pointer is between 0 and 1. Now, at every step the new x, y, s would be nothing but x, y, s into this into x, y, s is assigned the old x, y, s plus alpha times delta x, delta y, delta s.

So, the alpha is called the scaling direction, scale scale parameter or the scaling scale. So, this scale parameter alpha has to be chosen between 0 of course, it cannot be 0 that is not fair, zero means there is you are going talking about exact solution that is not possible. So, alpha is between 0 and 1 which is interval excluding zero and this denotes my suitable scale parameter or step size. So, it is also called step size. So, it is a while do loop, while do this solve this and once you get the new value x, y, s again assign to tau. I think you should have assign to tau the new one. So, end y end.

So, what you do is again you check if tau is so, my new x, y, s is this you my new tau is check whether tau is this and again go back and solve this and continue. So, when tau is less than equal to epsilon, stop the whole thing. So, what would be a suitable step size? So, that depends suppose this sometimes alpha in the short step path following method, which will start studying tomorrow you will have alpha is equal to 1 and that that will be giving you a lot of so, sometimes for when I am not writing an algorithm algorithm you can write this assignment parameters and so, it looks like program.

(Refer Slide Time: 21:56)



But in general, I would what people would do people would write x alpha or x plus does not matter, y alpha it depend on the alpha, s alpha is nothing but equal to x, y, s plus alpha times delta y delta. Now, in order to study the algorithms it is very important that we make certain estimates, this estimates will tell us whether what we were doing is true is correct. So, if I design an algorithm I have to show that if I run my algorithm I will finally, go and get what I desire, I will get tau strictly less than epsilon in certain step that has to be that is what is called the convergence analysis.

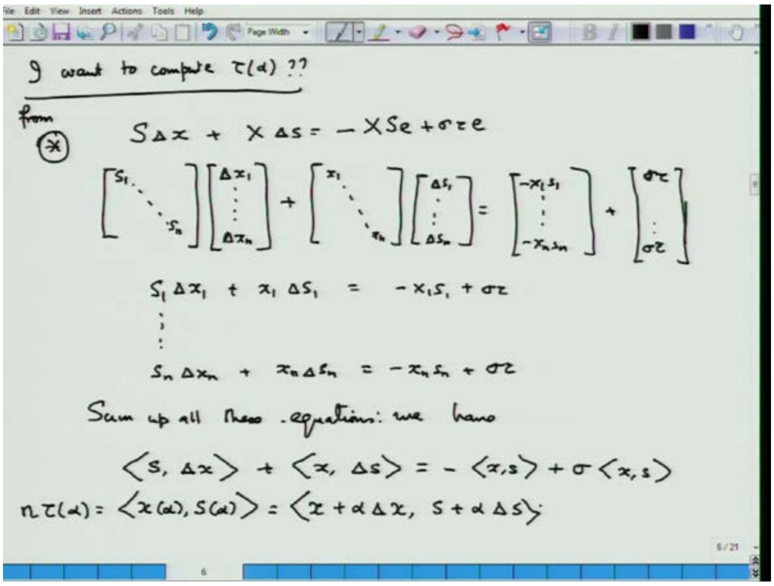
So, every algorithm we need to do we need to do a convergence analysis, in the sense that whatever algorithm I am writing it has it is making sense, that it is working has to be proved mathematically. So, let us look it and estimate that will require quiet often is this, what is this now how do I go about solving this, look if I look at the equation then what do I have if I look at the equation this one. So, I would have A transpose delta x plus delta y equal to zero and I will have A transpose delta A a a I am sorry I will have a transpose delta y plus delta s equal to 0 I will have A transpose A delta x equal to 0.

So, from the Newton's equation from the Newton system so, I will have A transpose delta y plus delta s is equal to 0, also I am having A of delta x is equal to 0. Now the from the newton scheme from the so, I will now write this as this newton scheme as the star scheme that is the main scheme, from the Newton scheme star. So, what I what would I have now, this means delta s I can write from here which is nothing but minus A transpose y. So, this A transpose delta y. So, I instead of delta s, I have written this taking the minus sign out so now, I can write this as A of delta x and what is A of delta x it is 0. So, I get 0.

So, we have this conclusion that if you take the newton steps of x and s first and third variable you get 0. Now an interesting question is, can I compute can I compute the value of tau alpha? Tau alpha of course, means so can I relate see tau alpha has to be such that at every it has to be smaller than tau, because my tau has to go down that is exactly what I want, because I want to drive my tau towards 0 and that is the whole ball game, can I relate tau and alpha with tau? So that we can be sure, that the value of the duality measure is coming down. So, let us see how can we do this?

Look at the last row of the equation this row s of del x plus capital X of del x is this.

(Refer Slide Time: 26:54)

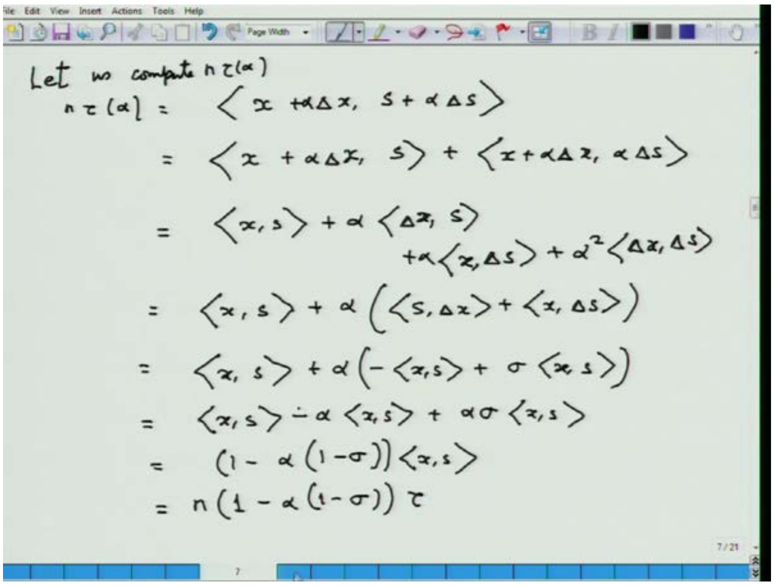


So, so I want to compute I want to compute tau alpha, let me look at the last equation. So, it will give me s of delta x plus x of this from the star equation from star. Now, I look at it component wise so, this is a vector x 1 s 1, x 2 s 2 these are vector sigma mu sigma mu sigma mu now I also want to look at this component wise this will become this s 1, s 2, dot dot dot and it is s 1 delta x, s 2 this delta x 1, delta x 2, delta x 3. So, let me look at it component wise what will happen right.

So, s is a diagonal matrix consisting of s 1, s 2, s n in the components. So, if I look at that let us see what actually happens, if I write this in a very broader way I would have. So, here if I write it down more clearly. So, first equation here would be s 1 delta x 1 plus x 1 delta s 1 to minus x 1 s 1 plus sigma mu. So, here there is a mistake this sigma tau epsilon. So, it should be sigma tau if you look at the equation its sigma tau not sigma mu, sigma tau here so, sigma tau and sigma tau and and so on. So, I can write s n, x n delta s n, sigma tau sigma tau.

Now, let the definition of sigma tau and then you you immediately see this will give me sum up sum up all sum up all all the all these equations, see if I sum up all these equations we have we have a following that is, s of delta x, x is a vector s 1, s 2, s n plus x of delta s is minus x s plus now you add so, it will be sigma n tau, n tau is again sigma x s by n. So, we will get sigma x s, that is what it is now I will do n into tau alpha is equal to x alpha, s alpha that is the definition. I will write down what is x alpha, x alpha actually for me means x plus alpha into delta x, s plus alpha into delta delta s. So, I will now do there do there detail computation

(Refer Slide Time: 31:59)

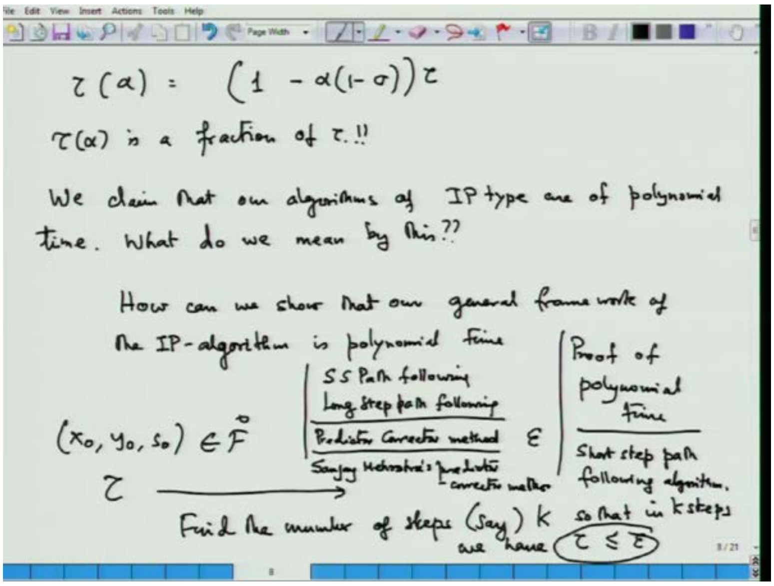


Let let us let us let us compute out let us compute tau n n tau alpha. So, what I have is n tau alpha is equal to I am just repeating what I wrote in the last thing, s plus you will see the estimate delta x into delta s will be useful here. So, now what I will have is x plus alpha delta x into s plus x plus alpha delta x into alpha delta s. So, I have x of s plus alpha times delta x sorry delta x into s, I just using the properties of inner products plus x alpha into x delta s plus alpha square times delta x, delta s.

Actually it is this it was important to estimate this, because we needed to find out tau alpha, that that was the reason to actually estimate this. In real research this thing came much later first wanted to compute this and then the other this thing was computed. So, that we know that we actually have to drive this part to 0. So, anyway we now know that this is 0. So, we have x s plus alpha times, s delta x plus x delta s and this is something we have just computed, we have just computed this to be x s plus alpha times this is nothing but minus x s plus sigma x s. So, it is so it is becoming x s minus alpha x s plus alpha sigma x s.

So, what I can write here is x s, I can take out that x s as a common thing. So, I will have 1 minus alpha into 1 minus sigma x s, but what is x says it is n into tau. So, it is n into 1 minus alpha times 1 minus sigma times tau. So, n into tau alpha is this.

(Refer Slide Time: 35:18)



So, what I would now have is tau alpha is 1 minus, again go back alpha into 1 minus sigma alpha into 1 minus sigma times tau. Now, alpha is a quantity between 0 and 1, sigma is also a quantity between 0 and 1. So, this is less than one, this is less than one, this product is less than one. So, one minus this something is bigger than 0.

So, into tau, so tau alpha is a fraction of tau. So, what we have concluded that tau alpha is a fraction of tau and that is what we require the value should come down. Now, how does the algorithm behave, if we are claiming we claim that that our algorithms our algorithms of the interior point type or the IP type are of polynomial time what do we mean by this this is something we really have to concentrate on now. So, what we have shown that in our general framework of IP algorithms we can reduce the tau alpha the duality measure which is of fundamental importance.

So, we have already on we have already built up something. Now what we want to show is that we can develop some framework, in this current framework we can put in some sort of little conditions which will immediately lead us to a polynomial time game, because in the sense that sorry it will show us that I can give you the number of steps in which your algorithm will terminate and that number of steps is not arbitrary, that number of steps is bounded on that above by a polynomial. So, if you know your number of variables in the division variables you can tell me in what number of steps you will lead to the solution.

Now, this is what we are going to discuss tomorrow, how can we show that our general framework of the IP algorithm which we have done which is actually this one this one is called the general IP algorithm. So, algorithm is polynomial time basically the idea is this you have a fixed x naught, y naught, s naught that you have epsilon F naught and now you start the algorithm, you start with tau naught and you are now going pushing the tau naught towards 0. Now, this is your epsilon is a precision parameter, now find the number of steps steps say k. So, that in k steps we have tau is less than sorry this is just tau.

So, I start with this and I have to show under what situations I can have or I can compute the number of step size k and we can show, that that number of steps that k that is required till we required come do this is actually bounded by a polynomial which can which can which tells us that this is the maximum number of steps, you would require to reach this. You can you would not need anything much more such a thing cannot be guaranteed for the simplex algorithm though it is amazingly effective in practice. So, tomorrow our first job would be to start with proof of the polynomial proof of polynomial time proof of polynomial time.

And then once, we have the proof of polynomial time we will start with something called short step path following algorithm. We will do its convergence so, the number or methods that we learn here are following: a short step path following, a long step path following of course, and there is something called the predictor corrector method and then once we know about the predictor corrector method, we will also give a brief outline of Sanjay Mehrotras predictor corrector method which is the most useful predictor corrector method that is used in algorithms.

We will try to if time permits to briefly touch on what is called the potential deduction algorithms, which is also a type of IP algorithms. So, with this we end the talk here, and tomorrow we will start with proving the polynomial time, and then get in to these algorithms. So, there we would need around three, four more lectures to have a brief idea about these things, and we will give references from which you can do your further studies.