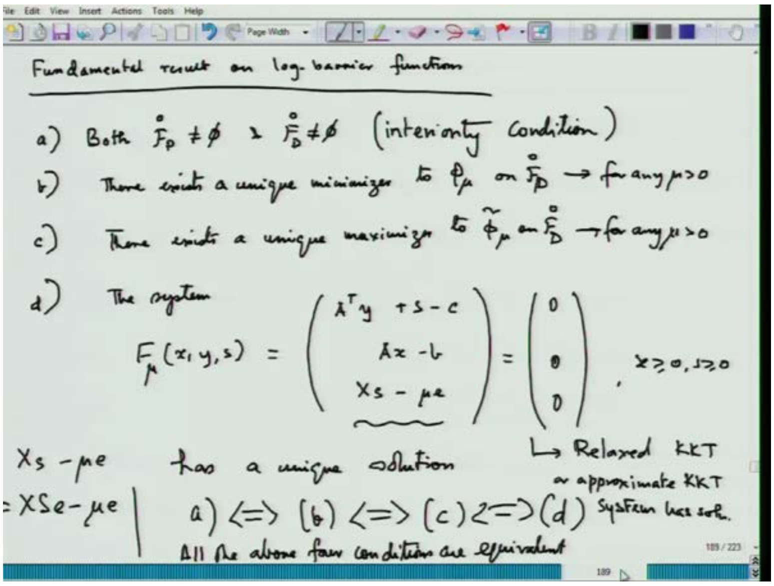
Convex Optimization

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

Lecture No. # 29

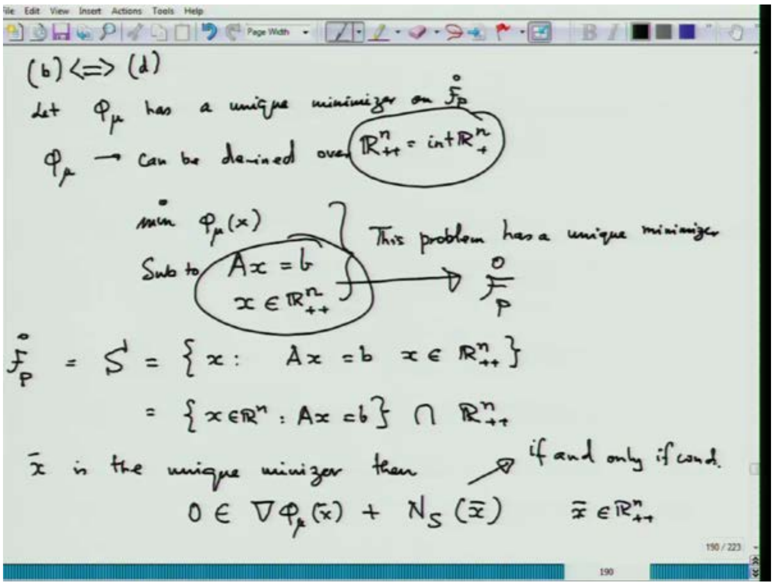
(Refer Slide Time: 00:31)



Yesterday night, we (( )) in the last class we ended by writing down a very, very, very, very important result regarding log barrier function, and this result is already in the screen as you can see. So, it says that, if I take the interiority condition that is the strict feasible set of the primal problem and the dual problem both on non empty. Then all the rest b, c and d will follow, that is for whatever mu you take, there would exist a unique minimizer to phi mu; whatever mu you take, they will be exist unique minimizer to phi tilde mu. And that system F mu (x, y, s) is equal to 0, we will have a solution that is the relaxed KKT or the approximate KKT system we will have a solution.

So, the unique fact that the interesting fact that this result tells us at this mu, it has a, it is not a very big problem, you can take any mu strictly bigger than 0, that is why we have written here for any mu and for any mu. Now, this is something very, very important, and we have to keep in mind, because this is central to the development of algorithms for interior algorithm interior point algorithm for linear programming problems.

(Refer Slide Time: 01:46)

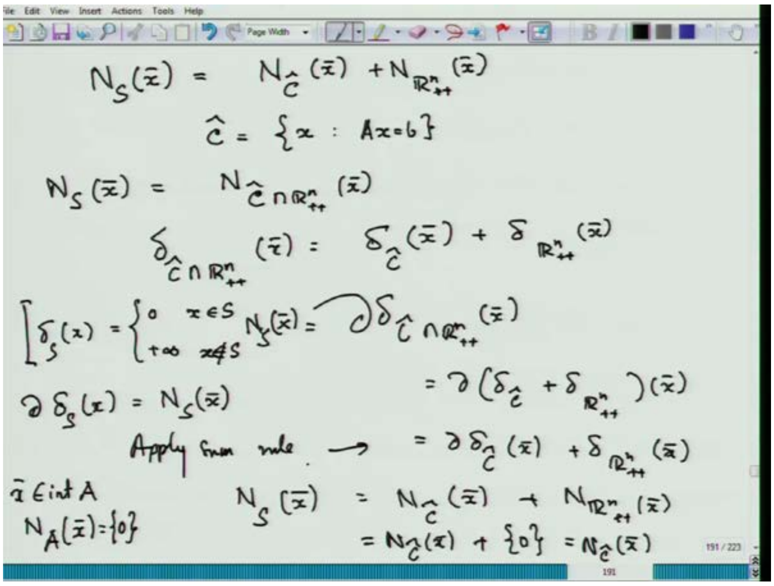


Now, we said that we will try to give a proof of at least a part of the result. So, we will just prove that, we will assume that, we know that a implies b. So, the interiority condition holds, and so we will try to now show that if I have b, I will have d and if I have d, I will have b. Now, proving b implies a and a implies b involves certain difficulties, which I do not want to put you in, it involves lot of technical issues and lot of the viewers here, who may or may not be doing any sort of may not be involved in rigorous mathematics, but just really want to know the basic ideas involved in this sort of algorithms, may not be thrilled with such a proof. So, it is very, very important that you also at the same time have some basic understanding that.

If I can find minimizer unique minimizer of the strictly convex function phi mu, that is the barrier function, then I am actually solving the in exact KKT condition, so here, I we symentling link between the solutions of barrier functions and the solution of the KKT system, the approximate KKT system, which will generate what we call the central path in interior point methods. So, phi mu here can be actually instead of being generated over only F naught p, it can be generated over int R n plus also and now, what we know, if we assume b to be true.

So, you know that, this problem has a unique minimize, because this is this thing constitutes the feasible set F naught p. So, I can write the feasible set, which is actually F naught p here. So, just for sake of it, I am writing it in a simple simpler way I am writing it f, if I am putting F naught equal to S for the time being and I am writing it like this, which can be written as intersection of these two sets. This x bar is a unique minimizer then the optimality condition, which is of course, this optimality condition is an if and only if condition.

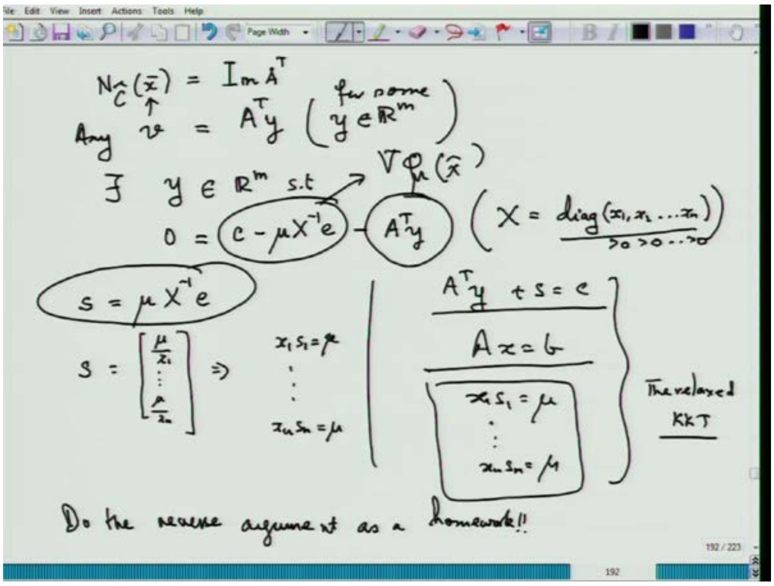
(Refer Slide Time: 04:27)



Then this optimality condition holds, then actually we can show that this is equal to this plus this where this is this, C hat is this. Now, how do we show that, we can write the indicator function of this as the sum of the indicator functions of this two sets, and this indicator function is nothing but as you see is 0 and x is in S and plus infinity, when x is not in S, and this is what we have discussed earlier and you know that the sub differential of the indicator function will give you a normal core and (( )) stuff. And this is applying the sub differential sum rule, so sorry this should be like this, delta of so sub differential of this…

So, the normal cone to S and delta of this, and then the sum rule is applied, if you do not remember how a sum rule is applied let me tell you that, the delta function on R n plus plus. The domain of the delta function the convex function is R n plus plus and it is continuous on R n plus plus and that is why you can actually do the sum rule of equality and then writing down this definition that the sub differential of a indicator function is a normal cone you have this, but this been open set, you can prove that if x bar is elemental interior of some set say A, then some convex set A, then normal cone to A at x bar is nothing but the 0 vector. This has… So, it is a cone which is a trivial cone, which consist just 0.

(Refer Slide Time: 06:20)

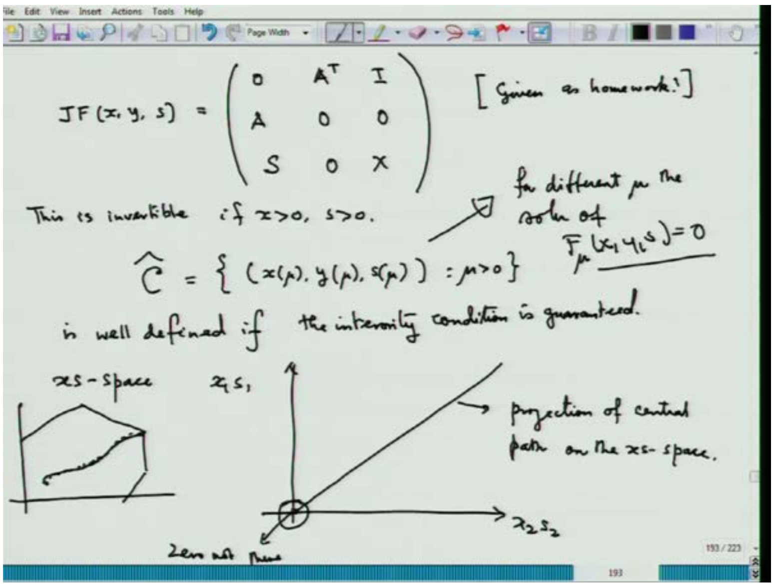


So, this N S (x bar) is nothing but N C at x bar; and N C at x bar, we have studied earlier is nothing but the image of A transpose that is why any v in N C at x bar therefore, there exists y in R m plus. So, instead of this, we should write, so for any v, it can be replaced for whatever v you choose here given give me a v, there would be a y in R m such that v is equal to A transpose y. So which means, there exist a y in R m such that, this holds because this is nothing but the grad of phi mu at X bar. And this is finally, so this is this is the normal cone element, which is the form A transpose mu. So, instead of taking y, I am taking just I am putting replacing y minus y does not does not make much of difference just for simplicity; instead of y, I have just put minus y does not matter, you can said that, minus y in R m such that this holds.

So, if I put S is equal to this, which is exactly if you look, if x where x is the diagonal matrix having these as diagonal elements, how is X, is in the interior because X is all this x 1, x 2, x 3 these are all bigger than 0. So, you can this definition mu of X inverse, this is completely clearly defined. So, this is nothing but the vector mu x 1 dot mu x n. So, you are getting x 1 s 1 equal to mu x n s n equal to mu. So, you are basically solving this system, but is required you are solving the system f mu x s f mu x y s equal to 0, and this is not the giving you this and; obviously, you have A x equal to b, because, your solution set has to be a feasible point and also it satisfies this. So, any solution of that actually solves this system it gives me a y, gives me a S which is this.

And In fact, the interesting part is that x is created from S is created from the x itself. So, if I know the solution of the barrier function over F naught p. Then from that I can create the S and then the y automatically is generated and hence I have also got this and also just by from the definition of S, I have got this. So, this is the relaxed KKT which is relaxed KKT, so this relaxed KKT is also solved by the same solution. So, b would imply d now, if d if I assume d that is such an x exist, then I can go back and do the same argument step step back and So, I can go and have have that this solution of the relaxed KKT system would imply that it is a unique minimizer of phi mu over F naught p phi mu been strongly convexed strictly convexed function that minimizer would be unique. Now, I would leave the reverse argument to you to do as the homework, but remember that this is the very very important notion.

(Refer Slide Time: 09:38)

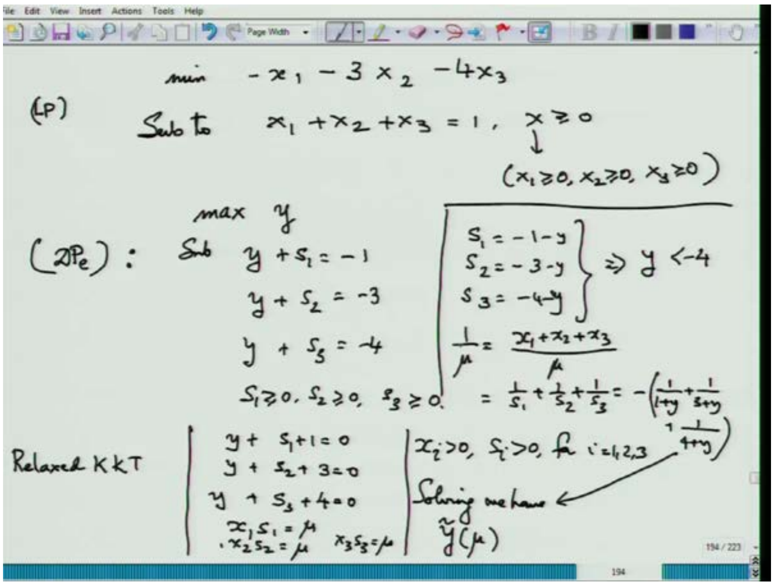


Now again I want to recall you, that if you look at the very fundamental Newton's equation then there is an for each x, y, s this is the Jacobian of that system. Now this Jacobian of course, F(x, y, s) is a A transpose y plus S minus C a x minus b and x s e. So, if you take this is this is exactly the Jacobian and this Jacobian is invertible if this and this is greater than 0 and this is very, very important for us to note. Now, what is the central path central path is, is nothing but for different mu's the solution of approximate

KKT condition or the relaxed KKT condition. So, for every you change the mu you have want a solution, you change another mu you have another solution.

Again I want to remind you though I have been reminding this for long time that keep on solving this system of equation and the all the point that you get creates create a path in the feasible space and that is known as a central path. So, the central path exactly is not really in just the primal feasible space, but of course, it will create a path in the primal feasible space, but C hat this central path, C hat is the subset of the primal dual space. Now, if I take the X S space that is X into S space, I should write more as then if you project the central path on the X S space then what you have x 1 s 1 equal to x 2 s 2 equal to x 3 s 3 equal to dot dot x n s n. So, this is exactly what you will have, you will have a straight line in the X S space. And this formalism of projecting the central path into the X S space has very, very important ramifications.

(Refer Slide Time: 12:19)



When will study for example, three different types of projection, three different types of method, three different types of two or three different types of path following method. Now, we will write down an example of computing the central path, you must remember that this central path is very, very well defined, because once you have defined the interiority condition that there is strict feasible solution. Then immediately, we know that whatever mu you take the f mu x y s equal to 0 has the solution and each of this solutions correspond to a path in the in the point in a central path.

So, you are taking the primal problem LP, so to minimize this function in three variables (No audio from 13.00 to 13.11) of course, X is bigger than 0. Basically this means, (No audio from 13:15 to 13:24) X 1 bigger than 0, X 2 bigger than 0, X 3 bigger than 0. So, for this problem I would like to find the central path, so I would now write the dual problem DP not really the dual, but the extended dual or sorry the equivalent dual. So, how do I do that, I would I would leave this again as homework for you to find the dual, but the dual in this case there is only one constraint. So, one equality constraints of number of dual variables equals the number of equality constraints.

So, we will have max of y subject to A transpose y plus S equal to C. So, will be three such questions, the three such C's here, what is my a ? a is just 1, 1, 1; and you have X 1, X 2, X 3. So, it is is basically the vector 1, 1, 1, so it is a one cross three matrix. So, A transpose is three cross one, it is the column matrix 1, 1, 1; so 1, 1, 1 into y, so it is y, y, y. So, A transpose y is just y, y, y, and you are adding S to it, which is S 1, S 2, S 3. So, this equal to C subject to S 1, S 2, S 3 bigger than 0. (No audio from 1**5**:13 to 15:21). Now, once you know thus, now my KKT condition what is my KKT condition, what let me write down that relaxed KKT.

Relaxed KKT is this system. (no audio from 15:38 to 15:49) y plus S 3 plus 4 equal to 0, x 1 s 1 equal to mu, x 2 s 2 equal to mu, x 3 s 3 equal to mu, this is my relaxed KKT. Of course, x i s i is strictly bigger than 0, because mu is strictly bigger than 0, x i strictly bigger than 0 and is strictly bigger than 0 for i equal to 1, 2, 3. Now, what do you have from here let us take the calculation here, so what do I have form this system, I would have S 1 is equal to minus 1 minus y 1, S 2 is equal to minus 3 minus y 2, S 3 is equal to minus 4 minus y 3 no sorry, I am making of very bad mistake y, y, y is same, there is only one y, this is what you have…So, all of these would imply that y is strictly less than minus 4.

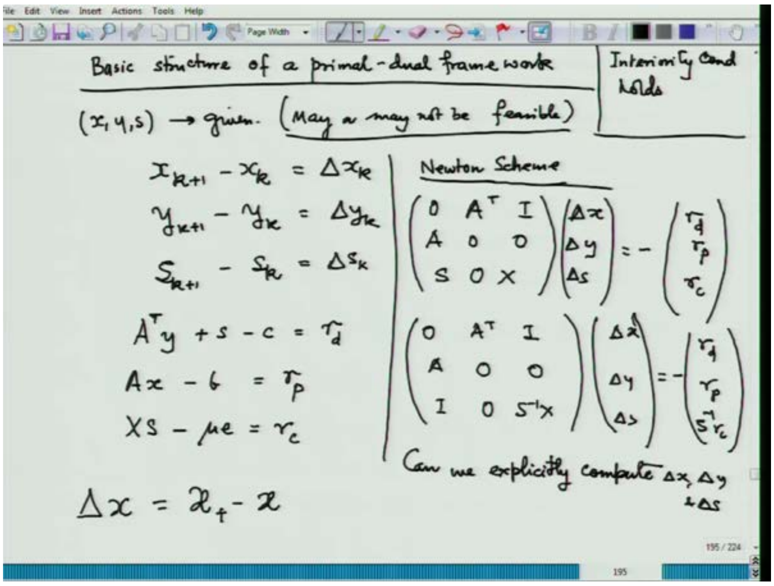
Now from this side of equation x 1 s 1 equal to mu, x 2 s 2 equal to mu and x 3 s 3 equal to mu. I will get the following expression. (No audio from 17:23 to 17:32). So, 1 by mu is one is x 1 plus x 2 plus x 3 by mu, now this is equal to 1 and then this will become 1 by S 1 plus 1 by S 2 using this set of equations. The last one this this set this set of equations and this now you can put, so mu is now there is a relation between mu and y. So, I can express y in terms of mu you see, once I express y in terms of mu, I can express S in terms of mu and hence I can express S in X in terms of mu, what is S mu is equal to

1 by X mu that is all mu by X mu, so mu is actually given to mu, so this is (No audio from 18:23 to 18:30) it will is a cubic equation that you will have.

So, you can solve the cubic cubic equation by the usual Cardano's method or some other standard algebraic method, you solve the cubic equation and solving, we have solution y tilde, which is the solution of this equation y is expressed in terms of mu. Now, once I have this no matter what I will I will know that once I have y mu y tilde mu, I will have S tilde mu, when I have S tilde S S mu basically. So, when I have S mu, I will immediately S mu, so that that will keep on generating the central path as I change mu, so this is one single example that even for a very simple case how difficult it could be to find the central path.

So, the question is, is it advisable to really make an attempt to do such things. The answer is of course, not; you do not attempt to do such things, because these things are so complicated for very such simple problems; it would be so complex, if the problem is very large. So, what we have supposed to do them that is the question. (No audio from 19:44 to 19:53)

(Refer Slide Time: 20:18)



So, we have to go to the basic structure of the primal dual framework, we will now go to the basic structure of a primal dual framework. (No audio from 20.05 to 20.25) Basic structure of a (No audio from 20.27 to 20.42) Now once I am talking about the basic structure of a primal dual framework, what we would have is the following is that we cannot now attempt to think about finding the central path. So, what we have to do we have to do certain slight modification in our Newton system, so that, we attempt to find points, which and we saw that new slightly modified system approximately and so we try to find certain points near the central path, and we have to show that it is enough to find such points, we do not need to find some extra something much more rigorous this sort of approximate points will finally, lead me to the solution, and that is that is exactly what you want. (No audio from 21:39 to 21:48)

So, in general given any x, y and s, I do not know whether they are actually primal feasible or dual feasible. So in general, x, y, s is given, which may or may not be feasible, so if it is feasible then it is fine, if it is not feasible. So, let us do the standard thing writing x k plus 1, y k plus 1 minus y k is delta y k this is the standard sort of notations. Some books of (( )) make a little change they write this as minus grad x does not matter just to adjust the minus in the Newtonian sign (No audio from 22.49 to 23:02) and let us write down this system (no audio from 23:05 to 23:15). Unless they are feasible we do not know either this is 0, so we can put this the dual residual that is, is not 0, but something else. So it define it basically, the primal residual and the complementary residual X S mu is r c.

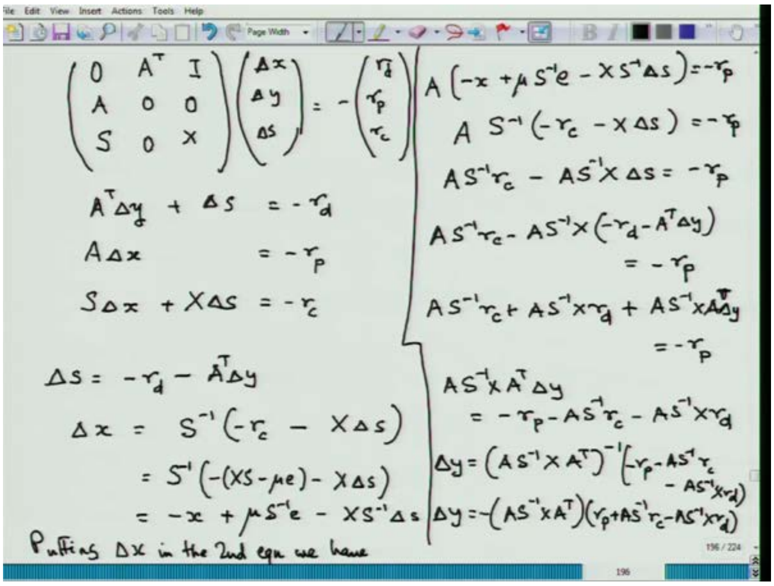
So, Newton's equation is of this form in general (No audio from 23:47 to 23:55) are the Newton's scheme rather I should say, if you are little bit of more where you write you would say Newton's scheme, this particular case it will look as (No audio from 24:08 to 24:27) this into delta instead of k, I am writing k here delta x k delta y k, I am just writing delta x, delta x means x, x k to x k plus 1 for the time. Of course, you can write delta x k and find for every k what what is it, but does not matter it will be almost the same, it will depend on the x at that moment, x k at that moment. So, I am what you can, I write delta x to denote, x plus and x in general this would mean k plus 1, this will mean k that is all x k plus 1 minus x k. So, do not bother you can put x k also, but I am just for simplification, I am writing it in this form (No audio from 25:13 to 25:26) this is exactly my Newton's scheme, but if all are 0 these two are 0 then 0 0 minus r c.

Now, can this equation system of equations have a solution, I can slightly modify this equation, now because X and S is strictly bigger than 0, because when we start our basic structuring and assumption, which is standard is that the interiority point point condition hold (No audio from 26:00 to 26:08) So, now what I will do, I can do a little bit of bit of restructuring. So, I will multiply this row block row by S inverse, so what you will have here is, so if I multiply s inverse it will remain the same it will new solution new system will give the same solution (No audio from 26:30 to 26:37) X by S basically (No audio from 26:39 to 27:02)

Now, can I actually solve them out explicitly can we (No audio from 27:09 to27:20) this is equivalent, I am not writing we can can we explicitly compute delta x, delta y and delta s. So, we take this and go back and and try to do do this one, so now, what we have going to show as we have said that, can we solve out this delta x, delta y, delta s. Now, the question is very, very simple that by solving this we are solving the Newton system. For this particular for the in exact or relaxed KKT system and that is what we need to solve to find the central path. So, this is very important that in our algorithms or the program that we will write will have to know the solutions of this.

So, if the first step is to know that can we solve the Newton method explicitly. So, it will be easy to program in because solving the Newton's step would become expensive, because of the inversion that would require of this matrix. If you cannot solve them explicitly, but though you will need some inversion, but it is very good to store the inverse at the very beginning, because these matrices are fixed and so then you can immediately write down the whole thing. So, instead of writing inverse of matrices doing matrices the inverse of this whole matrix or this matrix A whatever what we would like to do is without using the inverse of this whole matrix, so without using the inverse of this A or A transpose whatever, we can simply compute this delta x, delta y, delta s only inverting the diagonal matrices which are just simple things.

(Refer Slide Time: 29:01)



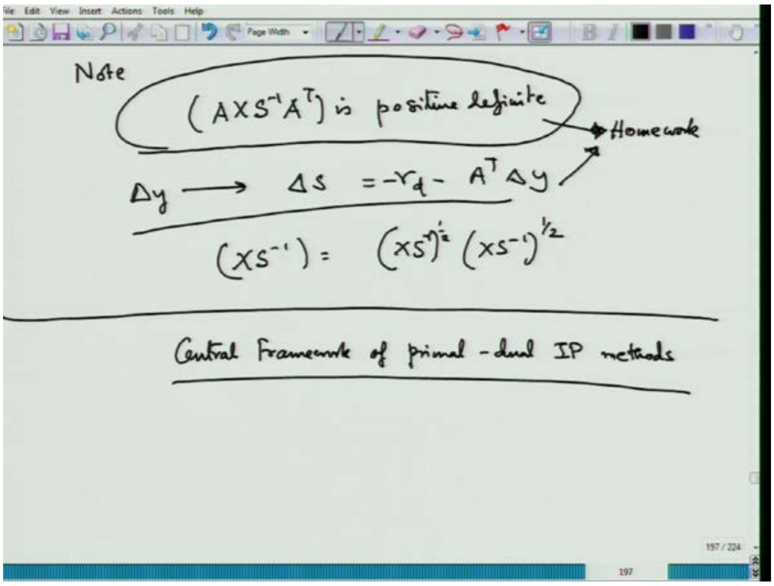
So, now we will write down the whole system once again which is (No audio from 29:00 to 29:07) to do the step by step computation. (No audio from 29:09 to 29:19) This is equal to minus of r c sorry r d, r p, r c. So, let me see, so what do I get from this, I get A transpose delta x sorry delta y or delta x, delta x is nothing to do is (( )) A transpose delta y plus delta s i into delta s is delta s is minus r d, A times delta x remaining at 0 is minus r p, S times delta x plus X time delta s, the last one minus r c. Look at the first one from the first one, I would have delta s is minus r d minus A transpose delta y. Now from the last one, what I will get forget about this one, now for the last one I will get, S inverse minus r c minus X delta s. So, what was r c, r c was just again reminding you r c was X S mu e (No audio from 31:09 to 31:24)

X S minus mu, so what would be done X and S will commute these are diagonal matrices, and so you will have minus x plus S inverse of mu mu would come out and S inverse of e, so it will become s 1, s 2, s 3, so mu 1 s 1, mu 1 s 2, mu. So, odd you can just write mu S inverse e minus X S inverse, S inverse x and s they will commute as diagonal matrices delta s, so this is what is delta x, I have got right. Now, I will substitute these delta x into this equation, so I will get delta s from there, so putting this, putting delta x in the second equation. (No audio from 32:37 to 32:49)

So, let us put delta x into the second equation, so A times minus x plus mu S inverse e minus X inverse delta s is equal to minus r p. Now once I know this fact, I can immediately write down, so x here means x k, so instead of writing this you can also write the second equation A S inverse minus r c minus x delta s is equal to minus of r p. So, you will have A S inverse r c minus A S inverse X delta s is equal to minus r p. Now delta s you know from here, there is something. So, you can write A S inverse r c minus A S inverse X into delta s is minus r d minus A transpose delta y, this will give you minus r p.

Now A S inverse r c minus A S inverse X minus minus plus r d minus minus again plus A S inverse X A transpose delta y is equal to minus r p. So, A S inverse X A transpose delta y is equal to minus r p minus A S inverse r c minus A S inverse X r d and delta y thus is equal to A S inverse X A transpose this into inverse of this part into minus r p minus A S inverse r c minus A S inverse r d. So, this is delta y, now you must ask me, why are you suddenly inverting this, let us see why we are certainly inverting this minus A S inverse X A transpose r p plus A S inverse r c minus sorry A S inverse X r d S inverse X r d.

(Refer Slide Time: 36:35)



Now why this inversion is done here, because this matrix is positive definite, now how do you prove that this is positive definite, so very important thing you note here, A X S inverse A transpose is positive definite. Let us see what happens with x, A X S inverse A transpose x. So, this would imply should be equal to (No audio from37:18 to 37:30) some some X X in as far the symbol. So, you take the any product, how would you say that this a is full rank, so how do you take take care of X S inverse that is the whole thing. So, this matrix is nothing but x 1 s 1, x 2 s 2, x n s n, this has to come with this one.

Now the positive definiteness of this part is an important issue here, and how would you prove this positive definiteness is something one has to think about a little bit, note that this is positive definite. So, it is not so simple to figure out that this positive definite it takes little bit of time. So, I ask you to go and do it as homework, so once you know delta y, you can figure out what is delta s, because once you know delta y, you know delta s is nothing but minus r d minus A transpose delta y, and once you know delta s you know delta x.

So, this can be also taken as homework, but it will be very important to know that, whether this matrix is having, now whether this matrix can be, how to prove that this is positive definite. It is slightly at important thing you have to know that there is something called a square root of matrix and you can express this as, for any positive definite matrix there exist another matrix b positive semi definite matrix, there is this another matrix b such that, b square is equal to a. So, if A is a positive semidefinite matrix then there is exist another positive semidefinite matrix b such that, b square is equal to a.

So try to write this as, so that is called the square root. So, this will be given by half, try to write this as this and then try to solve this problem. This this is anyway symmetric matrix, so do not worry about it. So with this we with this homework at the end, we figure we end this thing. And the next class we will start the central framework of primal dual methods, IP in dual point methods. See our standing assumption would be that, there is absolutely no question that the interior point condition is holding that is always there, thank you.