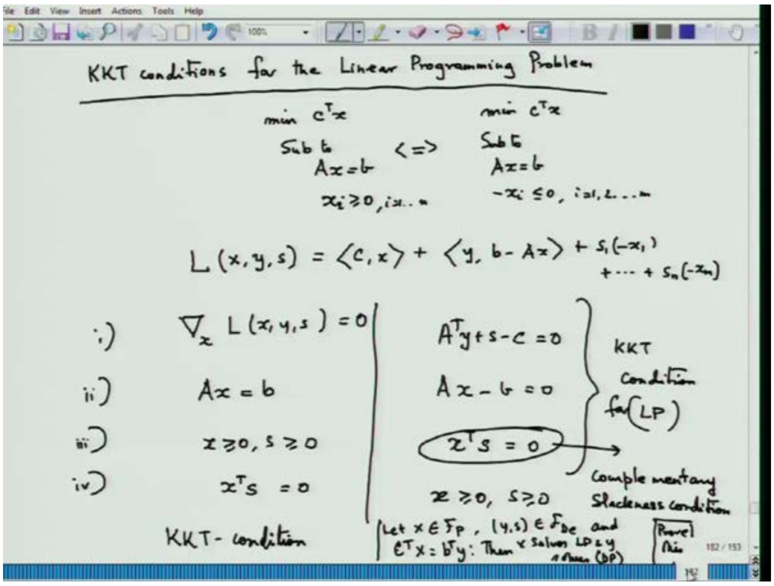
Convex Optimization

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

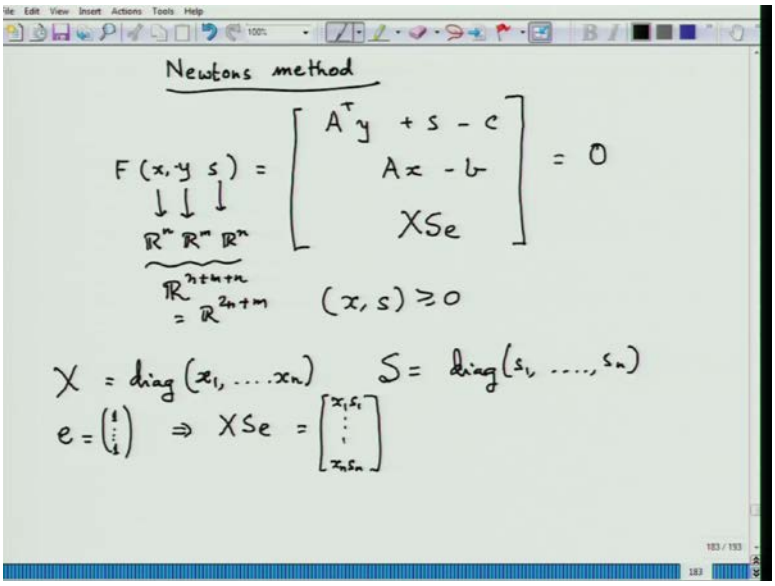
Lecture No. #28

(Refer Slide Time: 00:25)



Welcome once again, and we are continuing our discussion on interior point methods. Yesterday, we started our discussion gave a very basic idea, what we are suppose to do, and we wrote down the KKT conditions, and we are just attempting to know solve the KKT conditions, because solution of KKT conditions would give me the required minimum. In fact, the KKT condition solves to problem to solving the KKT condition, I not only solve the primal, I also solve the dual. So, these methods, that we will soon start doing they are call primal dual interior point methods, because they are jointly solving both the primal and dual. Interior point, because they remain inside the interior of the feasible set.

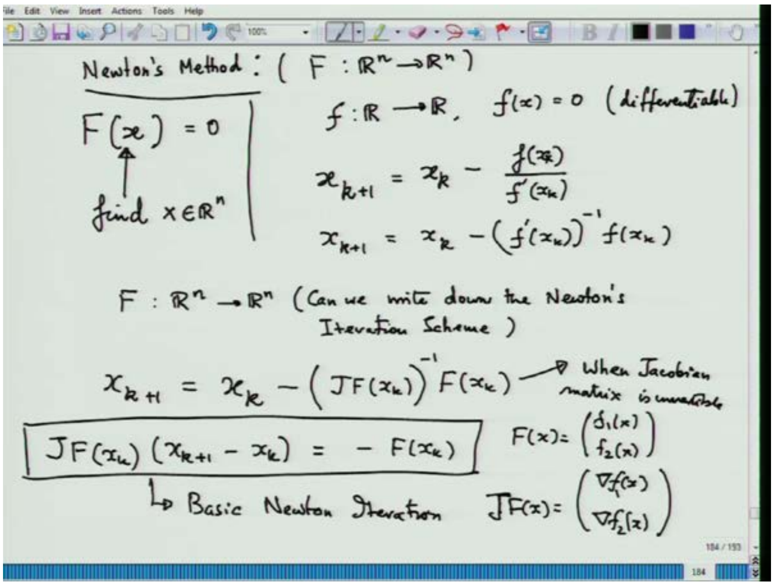
(Refer Slide Time: 01:02)



Now, let us observe that if I do look at the KKT conditions in the page before this one, I can write it down in a more standard equation form, that is I can write this KKT condition as a system in this form. (No audio from 01:21 to 01:27) A transpose this this, I will raise and tell you what I why I am writing this, where x and s are greater than equal to 0. X, when you are when you have vector x, and you are writing a corresponding capital X, this symbolizes are diagonal matrix, whose diagonal consist of all the elements of the vector x, all the components of the vector x. And this S is also a diagonal matrix consisting a wall.

Now, e is the vector 1 1 1, and of course you can see that, X S e is nothing but the vector is is nothing but the vector x 1 s 1 dot dot dot x n s n, all of them are equal to 0 right, that is a complementary slackness condition. So, this is in R n, this is in R m, this is in R n. So, basically my vector is in R n plus m plus n which is also R 2 n plus m. So, any element here any element of this form can be (( ))element here, of course, it is an element in this space R n plus m plus n; that is or R n cross R m cross R n. Now, I have to solve this equation and what do you mean by Newton's method for solving this equation. Newton's method we have learned also in calculus right.

(Refer Slide Time: 03:53)



Newton’s method comes into the very first course in calculus. Calculus one in most of over engineering colleges, most of over universities when you learn calculus one, you learned Newton’s method. Newton’s method here is, so, what we are going to look into it is a function F from R n to R n and we are going to try to find out find an x. So, if I have this equation, my question is to find an x, such that F of x is equal to 0; such that, this is occurring. Now, the Newton’s method, that you learned in the first year of your studies is, for a function from R to R, where n is equal to one.

And there you have f x equal to 0, if you take up your calculus books, you will see that, this Newton method works on a Iteration scheme given as follows, the k plus 1th Iteration is given in terms of the kth Iteration as… (No audio from 04:59 to 05:13). So, x plus x at the k plus 1th Iteration is the kth Iteration, x k is a k th Iteration x k minus f(x k) by f dash(x k). So, again write it bit more (( )) (No audio from 05:31 to 05:40), f dash(x k) inverse as to the power minus 1 into f(x k). Now, when I have the situation from F from R n to R n, can I write down the Newton’s Iteration scheme?

(No audio from 05:57 to 06:19)

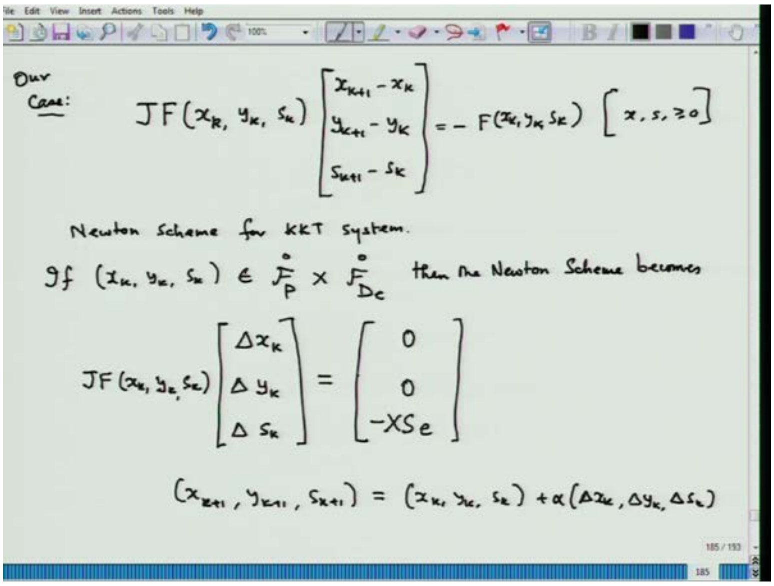
In this case, I have n and my derivative would be nothing but the Jacobian, see the, here, the derivative is not equal to 0 because here, if the derivative is 0, this will be meaningless right. This is of course, we are trying to find for differential function, these are all differentiable function, these are something, I should a mention. So, these are all differentiable functions, if you look at the function, that we have return on for the KKT scheme; that is also differential function. Now, how do I look into this thing and try to you know, how do I look into this Iteration scheme for R to R and try to get an idea for R n to R. (No audio from 07:02 to 07:09)

Now, here, we can user user intuition of course, I would leave you to humatically actually, I rigorously write down the scheme, but let us just use for fun using intuition. Now, for this, the derivative is nothing but a Jacobian mapping and we assume that, there is an inverse and then F of x k, but who told me that the Jacobian would have an inverse, that is a whole question. So, how does… So, if I write it much more, if the Jacobian was not having the inverse.

Then possibly, I could have written the whole thing down in this fashion. So, these essential is the Newton Iteration scheme. So, this is the basic Newton Iteration and this one is when the Jacobian has an inverse, Jacobian matrix does not inverse. So, when you function from R n to R n and R is bigger than one, a is n is bigger than one then you are derivative becomes the Jacobian matrix. So, I am sure that, you have about Jacobian matrix in you are basic calculus courses.

So, what you did take, if you look at this function F. So, let me just give a little bit idea about the Jacobian incase you forgotten it. So, you take this function F(x). So, if it is r 2 to r 2. So, you have a f 1(x), you have f 2(x). Now, you the Jacobian J f x is nothing but, you write the gradient vector is a row vectors grad of f 1 and grad of f 2, write them is row vectors and put them as rows. So, that is exactly a Jacobian matrix. Now, once I know this, how do I adopt this this thing to my own system to my own system of Kharus Kunt Tucker inequalities.

(Refer Slide Time: 09:45)



So, for our case our case it would be like this, the Jacobian of F at x our case, let me emphasize it is our case now. So, the applying the Newton's scheme in our system, now, x k plus 1 minus x k minus 1, I should have write in written x k, I will x k, if you if you are not convince, it is better to write x k and then latter on make it much more simpler. So, let me write this scheme J F(x k, y k, s k) this thing multipliding to this vector now, x k plus 1 minus x k, y k plus 1 minus y k and s k plus 1 minus s k and this must be equal to minus F of x, y, s. This is your Newton's scheme for KKT conditions of course, you have to have that or what is that call that x n has to be greater than 0, this is something, you have keep in mind.

So, this is the Newton's scheme for KKT system, but there is something interesting. (No audio from 11:03 to 11:11). Now, if sorry, this would be k, I had a mistake, this is x k, y k, s k right now. So, if x k, y k and s k is in strict primal feasible cross strict feasible strict F D e for the equivalent problem then what would happen, there is y k, s k is actually here, then what would happen. Then how do I write down the KKT system then the Newton's scheme becomes (No audio from 11:57 to 12:02) and that is what we will have in most cases because we will have a feasible point which is in the interior.

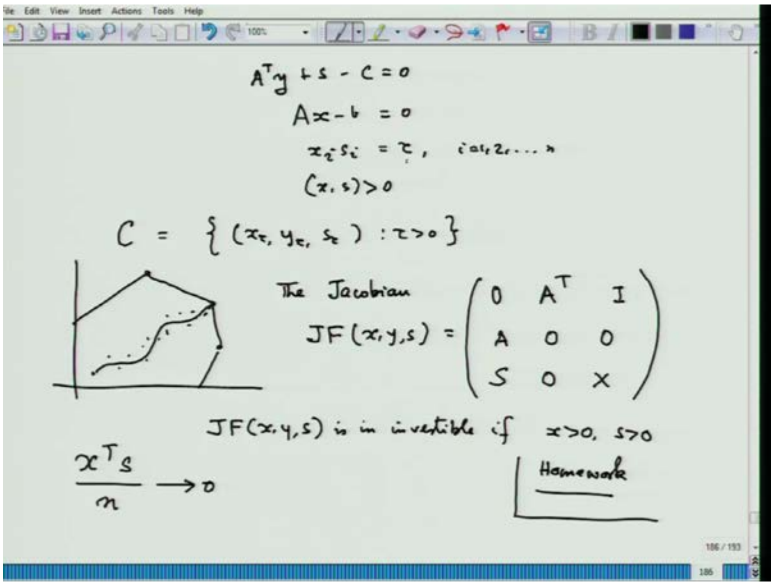
Now, our, that would allow us to simplify the Newton's scheme and the Newton scheme would now, look as. So, I will just put delta x k. So, this difference is, I will write as delta x k; these difference, I will write as delta y k; these difference, I will write as delta s k and this. So, once I have this, this would tell me, this is nothing but because it is, so, A x is equal to b and A transpose y plus s equal to c. So, 0 0, but only thing, you would have is that, this I do not know what will happen, because it is strict, they product is of course, something.

So, now this is this is the equation, I am suppose to solve, this is the… So, if I solve this equation. So, my new x k plus 1, once I solve del s k del y k del x k, if I find this values then x k plus 1, y k plus 1, s k plus 1 can be found by this following line such Iteration scheme which is x k, y k, s k. Sometimes, what is done, we just do not add this two del x k, though we have made that to be the sign, you can actually add this things, but sometimes from the point of your practice, it is useful to have some additional parameter.

So, this is giving you the Newton direction right. These, the solution of these are called Newton directions of the KKT system, but then... So, if I perform a lines (( )) the Newton direction for some sorts parameter, I will take a alpha between 0 and 1 sorry 1 is included then I have to see, what would happen, the if I just join this two x k plus del x k, I might get up point, which might not be strictly positive, that is a whole point, it can be in the boundary. The idea is to restrict it inside the feasible set and as the result of which, what we have to do is to make take this controlling parameter or line sets parameter (No audio from 14:47 to 14:54).

Now, once I know this, why I do this because what I want that the end is, x k plus 1 to be strictly bigger than 0 and s k plus 1 to be strictly bigger than 0. So, this alpha has to be chosen in such a way, such that this holds and the whole game is, how to solve this system (No audio from 15:16 to 15:27) how do I solve this system. So, the whole emphasis in interior point methods is to try to solve this system. So, this will lead us to the story of the central path. So, what we do is that, instead of the solving the exact KKT system, we would like to solve something called part of KKT system, which we just you would write down and let us see.

(Refer Slide Time: 16:14)

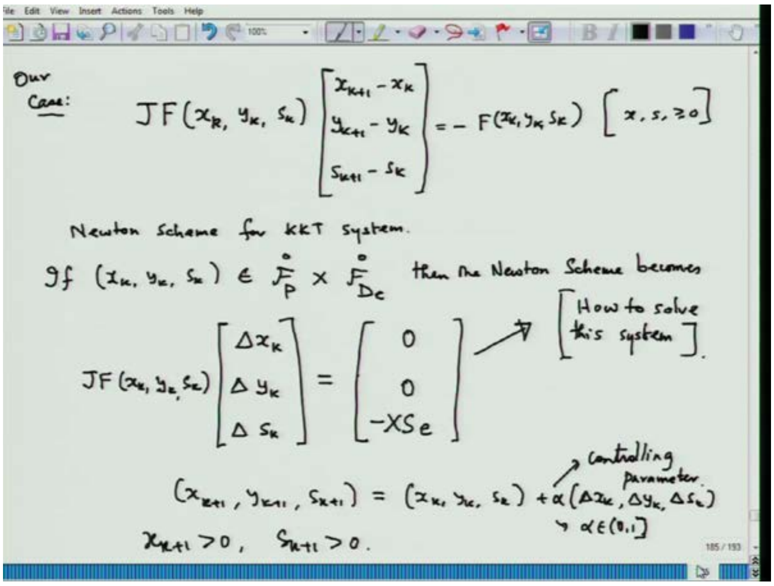


So, instead of solving the exact KKT system where I xi si equal to 0 for all i, I would try to solve the following system, (No audio from 16:13 to 16:22) xi si, I said this is equal to tau. Therefore, every xi si, it should be a number tau, though it is not easy to find something which will do it. So, in that case, there would give us, what is called as central path. So, if I solve this system equation, what I would get is something called as central path (No audio from 16:51 to 16:57) I keep on changing the tau, and what I do is I get a central path.

So, I have a I have this feasible polyhedron and I get a central path, which is all the points lying on this path is solution of this system, but when I am actually solve this system, I do not get the exact point, I get something nearby, but this nearby point then the, which are the approximation solution of this system cannot be to for off from this central path, it has to be some quite near to this central path.

And So, we will see that will define certain neighbour would which will push those points in a tube along tube or some sort of set, which will force keep keep the whole points the iterates within certain region of the central path, so, gradually will force them towards a solution. Now, instead of doing all this things, you might ask me a very important question.

(Refer Slide Time: 18:03)

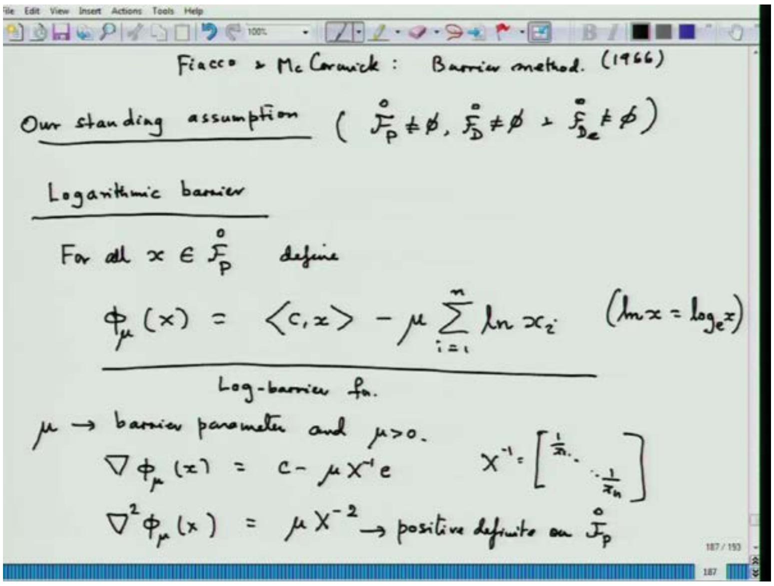


Now, you have this particular system for the KKT thing, that is, I have written here 0, if because I I have assume this because, if you are written here, this minus f x k thing here, this thing, this would become minus A transpose y k plus s k minus c k A x k minus b and of course, I am not writing it X S e. I am not writing the details, but now, so, what I am trying to say is that, you might ask me in that (( )), you said, this is the scheme, that you write, when you do not have differentiability, but if you are in the interior will this be differentiable sorry will be this invertible.

I am writing this scheme, when this Jacobian matrix is not invertible. Now, the question is will, this particular Jacobian matrix, when I take the Jacobian of this map will it be differentiable, if I take that fact that x k, y k, s k comes from this set and that surprisingly turns out to be s and that is one of the major results of this whole area. So, this is the very brief idea of this central path which we have given. Now, we will get into a more important issue, we will first write down the following following important result may be though we have not done the central path yet in complete detail, we are not we are not on the central path yet in complete detail, but we are going to write the following important results.

The Jacobian (No audio from 19:51 to 20:01) this is given by this matrix. It is homework for you, to figure out that, this is the correct answer, this is the Jacobian. Now, J F (x, y, s) is invertible, if x is strictly greater than 0, s is strictly greater than 0, this is very, very simple this is very, very simple, these full proof. So, I will again leave this as homework to you. Now, before we deal with central path in so much details, this important that, we look into this problem slightly more historically, see the interiors doing writing an algorithm been in the interior our feasible set is not a very new concept, it is a very old concept and you will find in the very old book by Fiacco and Mc Cormick.

(Refer Slide Time: 21:07)



And what happened is that, they device, what is called the barrier barrier method. And this barrier method was a nice interesting idea, where the points which are in the interior are not allowed to move towards the boundary. That is, if you are moving towards the boundary, the function values they create a function call the barrier function was value would start going towards plus infinity and that is, so you you would lose in minimizing such a function because your idea just like the penalty function, you create the barrier function by combining the constraints and the objective, what do you would that, you do it in a such a way that, you you iterates that every stage as would solve the sequence of barrier problems are inside the interior of the feasible set in contrast to the penalty method, where you have it in the outside.

So, now, I am talking about penalty method etcetera which have not spoken above, which would be in a separate course, but you need not get too much worried about it, what I am telling is, that there was a historical method, it was a wrong 1960s, I guess 1966. So, in the, if I not to very wrong. So, this people was interested in this idea completely theoretical idea with a nice convergence proof everything, but suddenly when interior point methods where been look into after (( )) revolution and we will started looking more deeply into the interior points method.

Many researches like Michael todd, Mike Jim Renegar they figured out, there is deep connection between the barrier method and the Newton scheme. So, the barrier method will play an extremely fundamental role in interior point method for linear programming. So, we will define, what is called a logarithmic barrier function and we would like to show that this barrier function is intimately linked with the solution of this Newton system or equation or the solution of the KKT system. So, it is intimately linked.

So, barrier method is linked with so, if you solve, if you optimize or minimize a barrier function in this particular case with certain barrier barrier properties which certain type of barrier functions, you are actually solving the KKT system. So, it is very important that, this linked be established because this is the deep link and this actually illuminates the whole mechanism of this interior point techniques because the barrier method, so, the first class of interior point techniques. So, our standing assumption,

(No audio from 24:09 to 24:36)

Now, you might say (( )), what are you writing the this could be pi, this could be empty, you can figure out example that this could be empty, I will not give you the example right way, I will give you the example after I do a bit of the subject. So, for for this current moment, let us assume, let you have this things given to you and now, you trying to build up this link with the barrier function. So, we have something called a logarithmic barrier function, you will understand, why it is logarithm logarithmic barrier function because you have a… So, for all x in the feasible set, all x in F naught P define the function.

So, this is a constraint and then the x, greater than equal to strictly greater than equal to 0, that constraint is pushed into the l n means log natural. So, if you are getting confused l n x is log to base e. So, x i is are all positive because I have taken them from the strictly feasible set and this is my logarithmic barrier function because I have used the log. So, this is, what is called the log barrier function. So, usually it is known in the literature as log barrier function, mu here is called the barrier parameter (No audio from 26:25 to 26:30) and mu is strictly bigger than 0.

So, it is important that, if I want to optimizes problem, I would like to know it is a gradient and hessian because that is, what is done in standard optimization problem that you because you will do unconstraint optimization not really unconstrained optimization, but you are basically looking at these problems because, you are you want to solve this problem over this particulars set, which is an open set because you are in the interior of the feasible set, you not you convert be in the boundary. So, this is an open set.

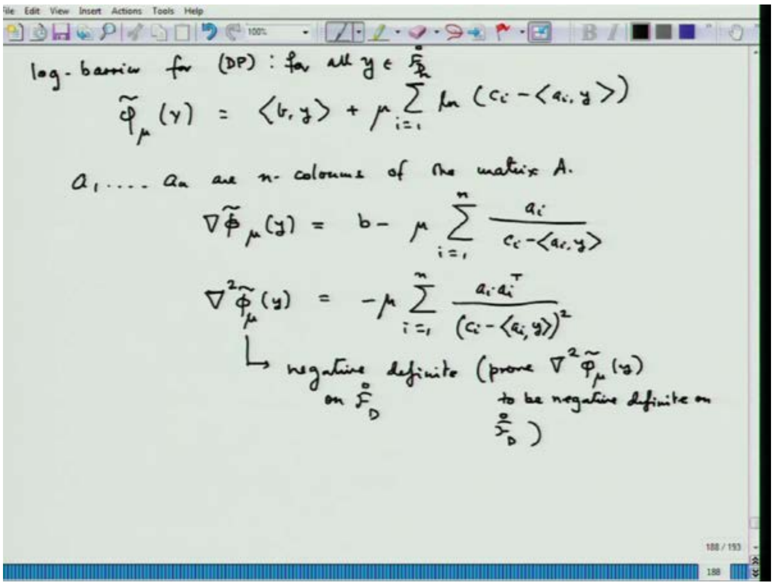
So, it will be just same as unconstraint minimization. Now, if you optimize over whole r n and optimizes over open set in r n, you are optimalitic conditions are same of course, it is not so, simple to say that, optimality conditions are same, how they are same, that would again involve in one to deep more deeper discussions about the geometry of the sets the associated tangent normal and… So, we will not get in to all this things, but just listen to the following thumb rule, if I have an open set in r n, my optimalitic condition just like a unconstraint once.

Now, I also want to determine the nature of this problem, whether it is convex or not this is convex of course, this is also convex, this problem is convex. So, what sort of convex is, is it strictly convex or strongly convex, some better property then just convexity. So, you see here, I would have, (No audio from 28:09 to 28:19) you know, what is x inverse because here, all the capital a in capital x is the diagonal matrix consisting of x 1, x 2, x n all all this this elements are positive.

So, X inverse is nothing but a matrix of this form (No audio from 28:32 to 28:37)that solve. An of course, you can see, figure out this one, you can figure out this at home at your lesser time. So, this actually means nothing but x 1 square, x 2 square. So, this is p d positive definite. So, this hessian matrix is positive definite on F naught p. So, this shows that, this function again go back to your old old notes, old or very beginning study that because of this is a positive hessian is a positive definite, this function is strictly convex.

So, phi is strictly convex on F naught p. So, our conclusion here is a log barrier function phi mu is strictly convex on F naught p. (No audio from 29:33 to 29:40). Now, what about the dual function? So, now, now the next natural question is, what about this dual problem to can we construct a barrier function for the dual problem, answer is easiest.

(Refer Slide Time: 29:57)



So, let us construct the log barrier function, not for D P e, but for D P just because they that is where the strictness comes in. So, you so you will have something like this, why is because, we are talking about dual variable, this is the dual objective, here, we add and not subtract because it is write it in that form, because I have written l n of c i minus a i y, a i is one row vector, a i is the… Here, a i y is the vector denoting the first row of the matrix A sorry the first column of the matrix A. So, it is A transpose y is, if you compute out A transpose y. So, A transpose means, if you multiply.

So, the first column, the i th column of a, the i th column of the matrix A is denoted by a i. So, if you do A transpose y, if you compute out this matrix, this matrix consist of, if you compute out this matrix computation, this is a vector A transpose y and this A transpose y is a vector which whose every, whose i th position consists of this particular l number. So, but this is strictly less than c i, because y is dual variable. So, for all y in F naught D, I have define it like this, i is from 1 to n, they are n column. So, a i, a 1, a 2, a n are n column of the matrix A (No audio from 31:59 to 32:06) that is it.

Now, what about (()) does it have anything to do, it does not have a nature property like convexity and all those things, let us compute out this. (No audio from 32:19 to 32:23) So, this will give me something this form, which you can compute out.

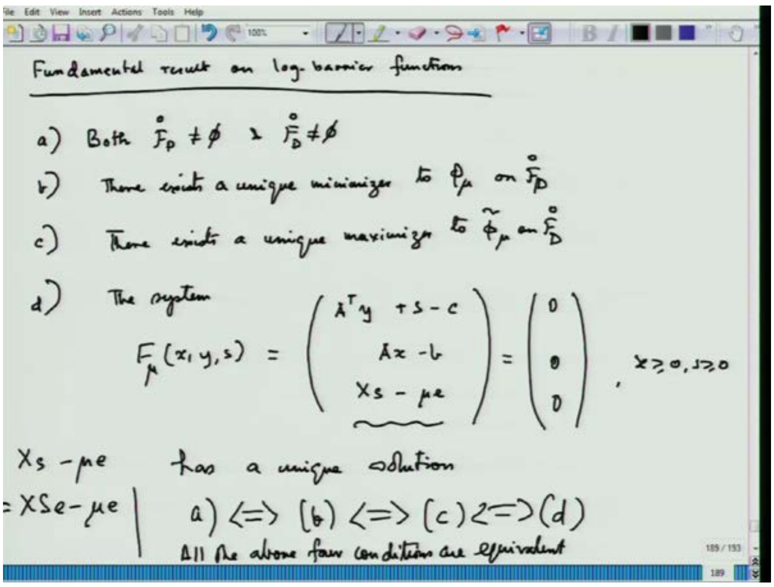
(No audio from 32:31 to 33:08)

Now, this is negative definite (No audio from 33:13 to 33:18) because, if you look at this one, this matrixes is mu is positive and if you look at this matrixes and because y is positive effect, this is positive. Now, a ai ai transpose, these are because of course, I am assuming that, a i is they all of them cannot be 0, a i(s) are some of the i(s) are non zero. So, the rank assumption tells us that, the using the rank assumption, that rank (( )) m, you can actually prove that, the product that this matrixes negative definite. So, prove this in homework, prove grad square phi tilde mu y to be negative definite on F naught D. (No audio from 34:08 to 34:16)

So, this is, these are some facts which let us know. Now, what is the relation between these barrier functions on the KKT conditions? Let us just have a look and we will try to end our discussion today with this very, very fundamental theorem and and will start up discussion about the central path tomorrow and also about, that also will discuss a bit about the primal, do all frame work and all this things. So, and how on, how to exactly solve the Newton's system, will give you the exact solution on the Newton's system.

So, here, we are not (( )) prove each and every thing, which is not feasible because of the time constraint in this course. So, you are not going to prove each and every step, it is not possible, but give you the major ideas because here, we are now putting in a some sort of small capsule, because interior point is a whole course, you can, is the proves will take a quite quite bit of time.

(Refer Slide Time: 35:38)



So, some proof, you would given not every proof and, but here, will state or the fundamental theorem, fundamental result fundamental result on barrier function, (No audio from 35:39 to 35:45) but I will not prove it, (No audio from 35:48 to 35:53) see, what would happen is that, proves of this would require certain sophistications, which might not be available with all the (( )) mathematical sophistication, which (( )) available with all the audience seeing this course.

So, keeping in view, the mind in, keeping in mind, the view point of, keeping in mind, the view point of all the audiences, I would like to emphasize all results now, own be given a proof like, we will giving in the very big fun first part, the very basic part, that we are giving proof. So, everything, but here, he want be giving proves everything will emphasize the result certain smaller results etcetera of course, we would reproving, but not each and everything. Our idea here is not really to show you that, come on how do you how do you solve the or how do you numerically write the algorithm, but give you a very brief idea.

So, will not write down each and every algorithm, we will write down one basic algorithm, through which we can test the theory that we have, we would develop now, but then will also give certain more algorithm will still that references and you can go and have a look at them. So, let us write down this fundamental theorem to end this course today. So, let us write down the following facts. (No audio from 37:33 to 37:37) So, they these are the non empty, this is known to you, (No audio from 37:40 to 37:45) there exists of course, will be unique minimizer, because (No audio from 37:50 to 37:55) because it is strictly strict minimum, (No audio from 37:58 to 38:07) if there is the minimizer, there should be unique (No audio from 38:10 to 38:18) maximizer.

So, you know, I just want to remind you that, this one what you have got here, this one, this shows that, this function is actually concave not convex, but concave on F naught D. So, you have to maximize the concave functions. So, dual problem is always the maximization (No audio from 38:47 to 38:53) sorry F naught P, this is F naught P, this is F naught D, the system, (No audio from 39:01 to 39:17) the instead of tau, let me put some little be, I will write a F mean see, what would happen is that here, we will show that, if I can minimize a barrier functions, I am actually minimizing this problem with this equal to mu.

So, if I if I fix upon mu and, if I minimize a barrier function then I am actually getting, finding a unique solution of this system or equation and that is the important connection because this is, what we want to solve. We are not hell bent in solving the KKT condition, but we have to solve this approximate condition and that is very important for us, from the practical point of view, this is, what we have to solve because we have to be in the central path. So, the as solution of the barrier method, the barrier function leads to the generation of the central path. So, that is the view to with theoretical link. (No audio from 40:07 to 40:13)

We will try to give a little bit of prove for this one, because this is very very fundamental. (No audio from 40:18 to 40:22) So, X s, X capital S sorry or you can write X S e. So, this actually means x 1 s 1 equal to mu x 2 s 2 equal to mu, here, I could also write this this part is a last part, see X s mu e can be return as X S e minus mu e same same thing does not matter. So, this… (No audio from 40:48 to 40:58) So, let us have this four condition sit and down, four as a things, four statements, we are made four statements. And now, we are showing that a implies b and b implies a, b implies c and c implies b and hence c implies a and a implies c.

So, what we says that, all the above four conditions (No audio from 41:29 to 41:33) are equivalent that is, if I can solve the log barrier function then then I am also solving this system, give me the mu. So, once you know that, if I, if there is the logs solution to the log barrier function, if there is the solution to the unique dual. So, what happens is that, the solution to the log barrier function. And the solution, if you have a solution to the log barrier function, when you have a solution to the minimi minimizer of the log barrier and maximizer of this dual log barrier for the dual one then you have solving this so or if you just know that, I you you have a found a minimizer on this one, you know that, there is a solution to this.

So, solution or the existences of the central path is intimately linked to the solution of this to problem, that is minimizing phi mu over this and maximizing phi tilde mu or F naught D. So, this is a very, very important thing, may be you will give a scheme of the prove of this tomorrow and then we will move on to the central path and other related issues. So, the idea as I told you again is to move along this central path, starts from point in inside and start moving along the central path. So, if I each point on the central path is the unique solution of this system, but the question is that, is very difficult to solve this system.

So, then our idea would be to solved a approximately, but still solve it in such a way. So, that I remind near the central path and I can make once step, one step, one step ahead. So, what how do I know? My idea is that, I will have to force this tau to go to 0, I have to I have to force this tau to go to 0. So, at a every step, I have to whatever new, whatever x i s i, I compute the approximate 1, 1 has to compute, this sort of sort of control parameter like this, which is call the duality measure.

Basically, I would like to force this duality measure to go to 0 in a basically from from just very simple point of view, I am forcing this tau to go to 0, if I as I make tau smaller and smaller, I keep on, if suppose, these the solution, I am move then I keep on moving along this central path and go and hit the solution, that is that is the basic idea and so, tomorrow start with some brief idea of about proof of this fundamental result on log barrier functions and then the KKT then the approximate KKT system and then we will go into the other issues.

So, we will have the, I will have some four or five more lectures on this whole thing. So, we can take it slightly easier, and then will go on to semi definite programming which is almost the end of the course. So, we will basically have learned two very important class of convex optimization problems, they are algorithms, they are deep properties, and you will see the semi definite programming so important, because lot of things come under semi definite programming problems. Lot of problems can be model as s d p which cannot be model as linear programming problems. So thank you very much, see you tomorrow.