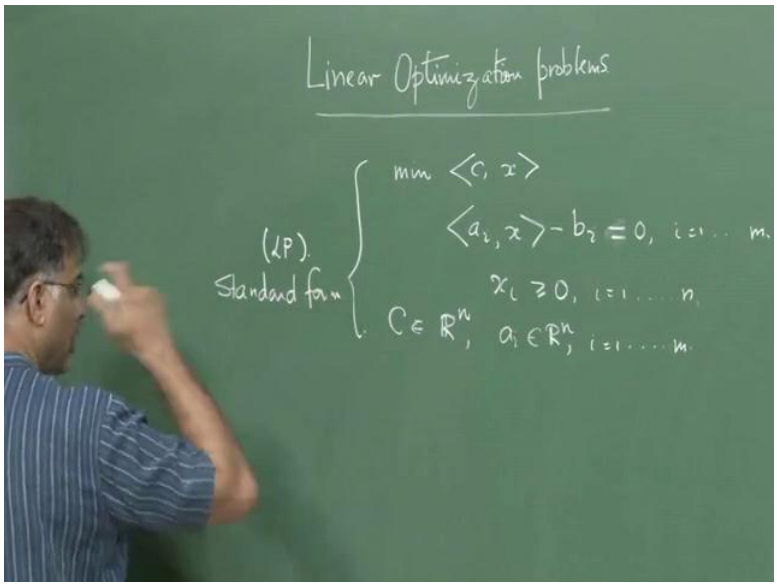
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 21

So, we are in the last part of our study of the karush Kuhn tucker conditions; it is not that or the fritz john conditions. It is not that this is the last part of the subject, but just because of the limited time we have in the course that we really have to end it here. Tomorrow we will give you a set of exercises on this issue, which I think you should try at home, they are very, very important. Now, some of these exercises would actually be given in a solution when f a q would be attached to the course at the end; so that those exercise, some of the exercises which are important would be solved.

And, today we are going to first show a very important thing, again using the motzkin's alternative theorem that we had studied in the last lecture which says that if you have a linear programming problem all your john multipliers would always be normal; there is no abnormal john multiplier. And, it means that the KKT condition always holds. Whenever all the john multipliers are normal we will call such john conditions as a karush kuhn tucker condition; after the seminar work of kuhn and tucker and later on found to be have also been done by karush. So, it is in this field of naming it is very very strange because somebody calls it karush john condition, somebody calls it Kara theory john conditions, somebody calls it you know john KKT conditions; there is lot of things. So, we will just have john conditions and the karush Kuhn tucker conditions.

(Refer Slide Time: 02:00)



So, we are here today considering linear optimization problems. When I am considering linear optimization problem, I am considering minimization of a linear function; any linear function of r n is given by an inner product; subject to so called affine constant, that is linear minus a translator real number. Maybe I should write equal to 0; there is a standard form. And, all the x i's are greater than equal to 0; the decision variables. So, here c is in is an element in R n and so is x then; of course, the decision variable. Each of the a i's are in R n.

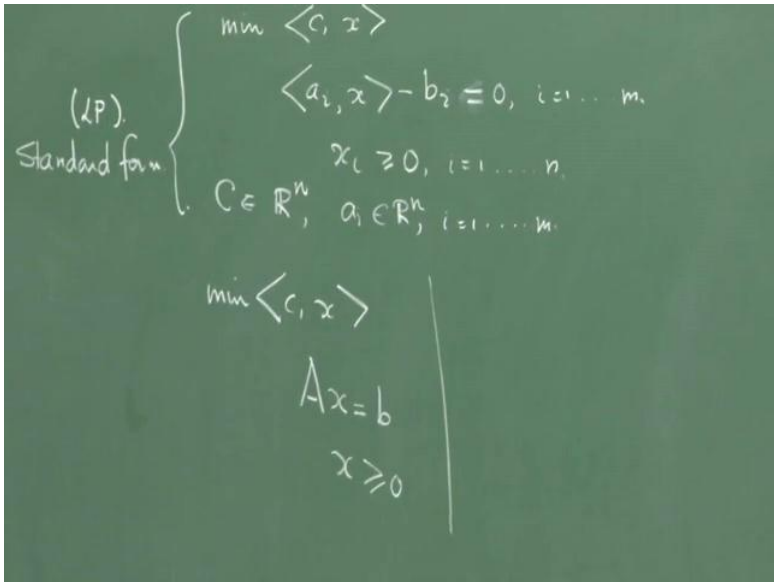
Now, if I consider now a matrix whose rows are these vectors a i, then I can write down this linear optimization problem in a more compact form. And, this is what is called the linear optimization problem in the standard form. So, I will call it LP; that is called and this is, this is what is a standard form. And, this is the form that is used to actually start solving it by using the so called simplex method or any other method. Of course, you could have a more standard form by putting this to be less than equal to 0. This requirement comes largely from practice; it is not really required for mathematicians to think about x i greater than equal to 0.

But in practice in most cases your decision variables are non negative integers. It could be say number of vehicles, it could be number of things to sell, it could be number of employees, number of students. These decision variables are or amount of food that you

buy the different type quantity and types of food; these are the types of food and the so you are making a diet plan and what is the optimal diet plan and all those things.

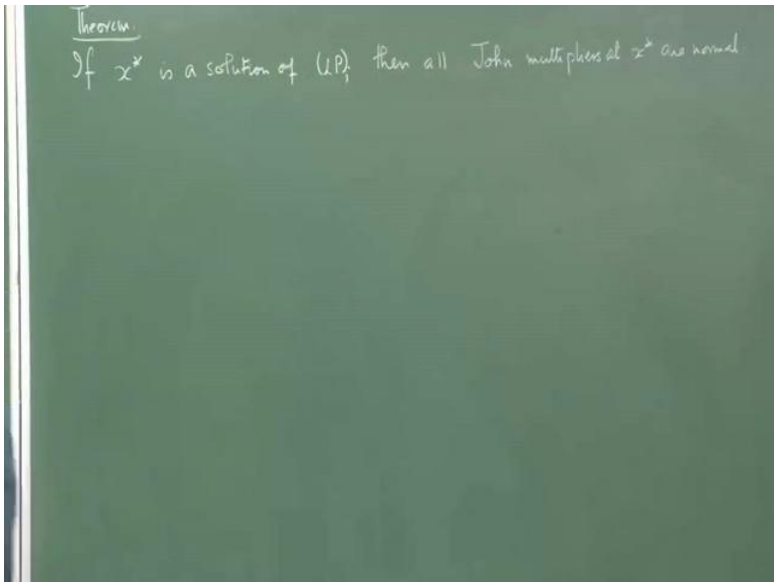
So, linear programming has a huge literature. A substantial amount of linear programming has been done in the course on convex optimization which I gave earlier. Here we are not going to spend so much time on linear programming, because this course is (( )) general foundation of optimization. So, we are going to mention this very, very important fact that, so you can write this problem in a slightly compact way.

(Refer Slide Time: 05:07)



So, if A is a matrix which is formed by taking all the rows as the vectors a i. So, it is the row matrix of m rows and n columns; giving you this equation basically A x – b, b is a row vector in a column vector in R n. And, x is also written as greater than equal to 0. And, this structure is very very important to consider because here you would observe that this x greater than equal to 0 actually means component wise greater than equal to 0. Now, once I have this I I am now inclined to write down the fritz john conditions. So, the conclusion that we will have is if x star solves LP, any solution of this problem any solution of this problem is a global solution. Because this is a convex programming problem, every local minimum is global that I have also proved in my other course in convex optimization. You see these two courses convex optimization and this foundation of optimization could be considered as a compact course and a two semester course in optimization.

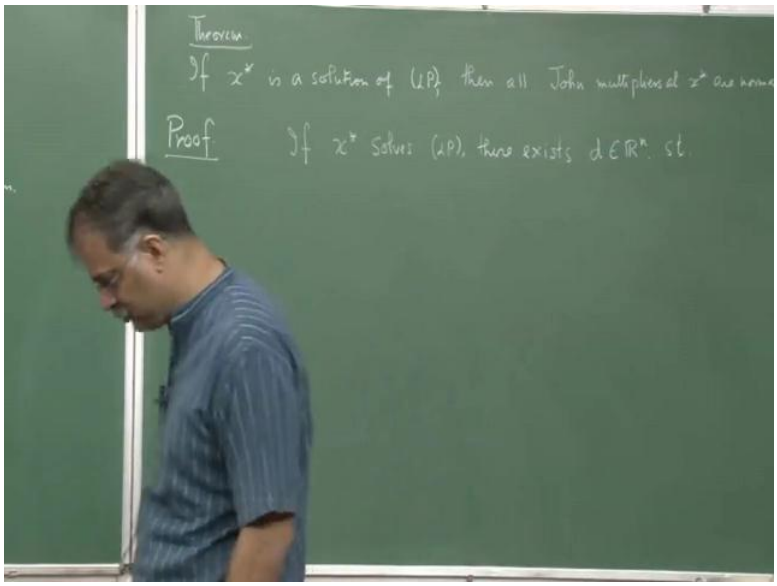
(Refer Slide Time: 06:38)



So, here if x star is a solution, so this is the major result. So, I would so this was known to Kuhn and tucker, one of these one of the celebrated results. So, these are known to Kuhn and tucker. So, I I would not put it there like that; I think historically that may not be a correct thing to do, but kuhn and tucker of course, knew this fact. So, I have written down the result which will say that if x star is a solution of l p then all john multipliers at x star are normal. As we, we will deduce this we you will soon see that there is nothing like a fact that the multipliers depend on x x star the solution.

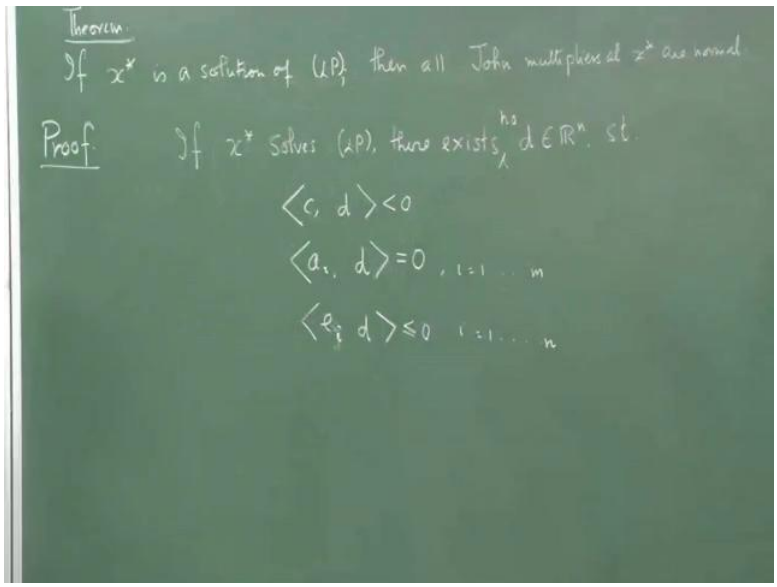
The multipliers that will appear or the john multipliers that will appear do not depend on x star. So, whatever be your x star the multiplier set is same. So, now let me just try to prove this fact. So, if you remember what we did when we tried to prove the fritz john conditions, we tried to prove that a certain system of inequalities are basically strictly less than they are strictly less than 0, then apply the Gordon's alternative theorem. Here we would go back trying to apply the Motzkin's theorem that we have learnt in the last class. I would like you to have a look at the earlier lecture; the Motzkin's theorem of the alternative, which I do not want to prove, put it here once again, because we have already done in just in a class before. So, what I would first prove is the following.

(Refer Slide Time: 08:43)



That if x star solves lp there exists, there exists d in R n; such that, let us see what is happening.

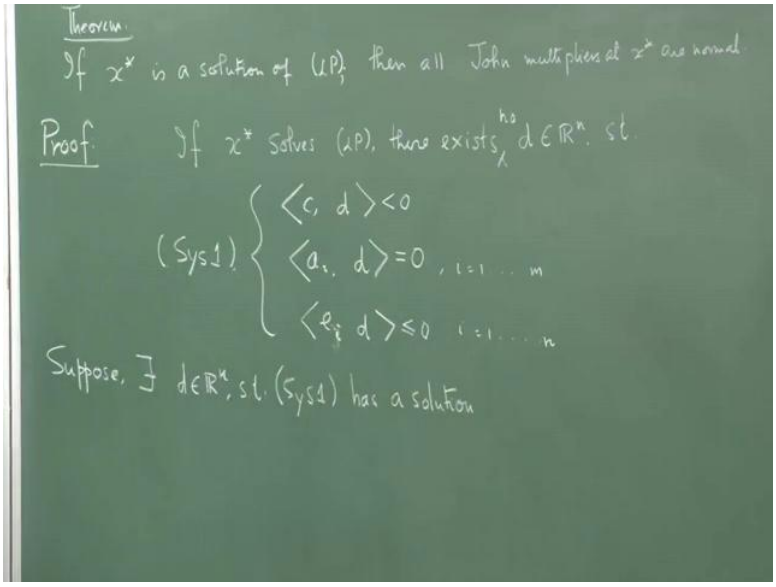
(Refer Slide Time: 09:28)



Such that c of d is strictly less than 0, a i d is equal to 0 and e i d is less than equal to 0. This is what I am going to prove, where e i is basically if you write all these x i's each x 1 as a function of x 1 x 2 x n then you can take the gradient of that and that would be 1 0 0 0 0, so that is e 1. So, where e i here I should write i equal to 1 to m equal to 1, you can put j also if you want e j, if you. No, let me put i that is exactly the thing that I have been putting; does

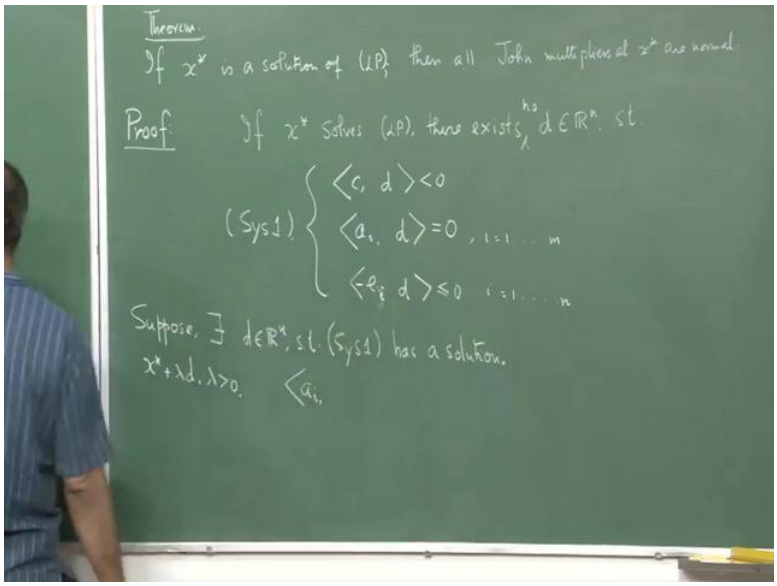
not matter. So, index is you know what this is the index over this; this is the index over that. So, i equal to 1 to i equal to 1 to what was that n; that is what we are going to first show. Sorry, I made a mistake there exists no d in R n. So, what I am going to show that there cannot be any d in R n, for which this is true. So, this would be a first system of the Gordon's theorem of alternative that we did in the last class. So, once you understand that you can immediately know, what is the final system that we are going to write; so, let us start by proving this fact. So, let us see how we prove this fact.

(Refer Slide Time: 11:41)



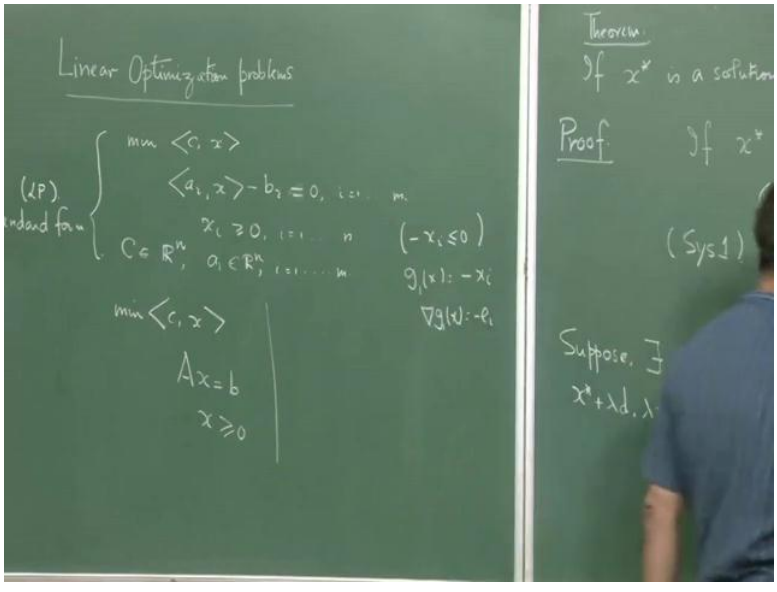
So, suppose on the contrary there exists I am just taking the mathematicians liberty to write like this. d in R n such that the (( )) system which I call sys 1; sys 1 has a solution. Now, you observe one fact x star is a solution. Now, I will prove that x star plus lambda d, x star plus lambda d will be a for some lambda strictly greater than 0 will be a solution, will be a feasible solution to the original linear programming problem. And, it would be a solution, whose where the objective value would be strictly lower than the current objective value c of x star which is the optimum objective value and that cannot happen. So, it is exactly the same way that we proved there we will show this. So let us see how we do it.

(Refer Slide Time: 13:23)



So, take a d so take you have this d construct x star plus lambda d with lambda strictly greater than 0. Then, what does this give you? This now, this shows you a i sorry this should be minus here; I made a mistake, because here it would be minus x i less than equal to 0.

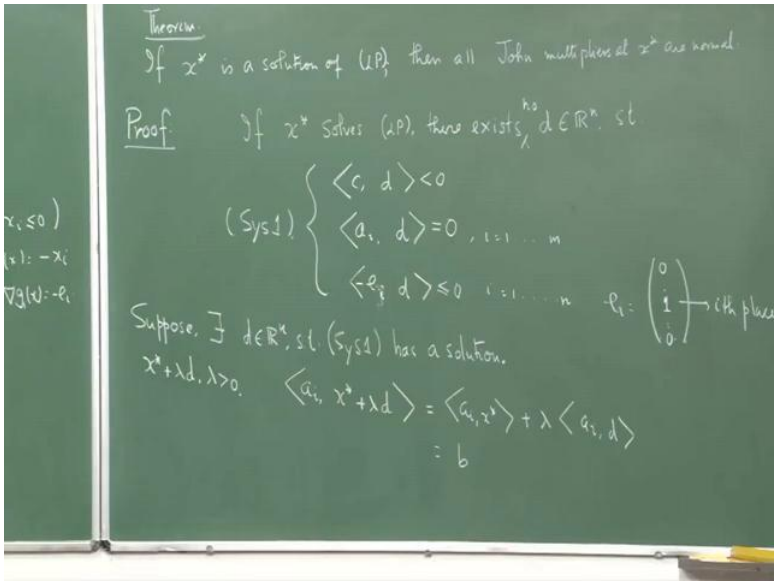
(Refer Slide Time: 13:55)



Because we have learnt it; when you write this one, you can also write it alternatively as minus x i less than equal to 0. So, you write it in the form of inequality constraint, so the

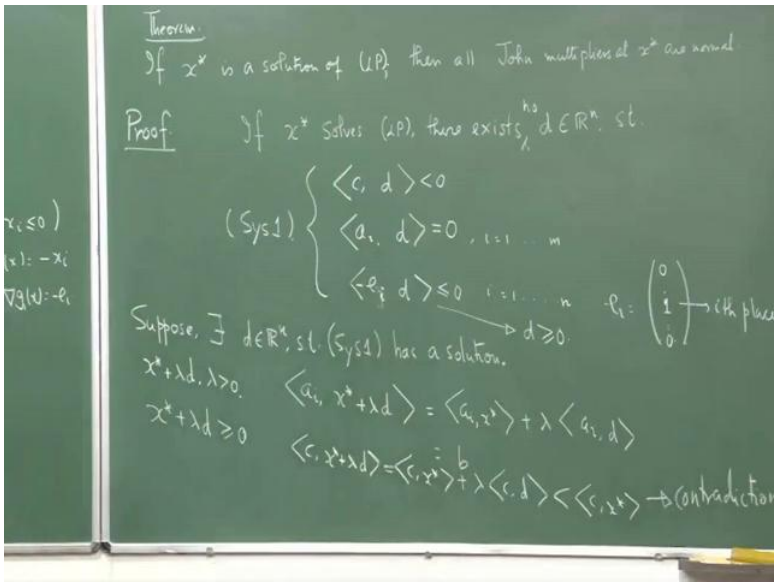
gradient of this function; so if I if you write this as g 1 (x) is equal to minus x i, then grad of g 1 (x) is actually minus e i.

(Refer Slide Time: 14:15)



So, e i's are actually vectors of this form one in the i th place and 0. Now, let us see what happens with this, so, lambda is strictly greater than 0; I will have a i x star plus lambda a i d. a i d is equal to 0, because d is a solution of this system of equations; while a i x star, x star being the solution must be a feasible point that is equal to b, so it is finally b. Now, how do I know that now x star plus lambda d is also component wise greater than equal to 0? To see this observe this equation; this would give you minus first one, it will give you minus d i d 1 greater than less than equal to 0, so d 1 greater than equal to 0; similarly, d 2 greater than equal to 0.

(Refer Slide Time: 15:32)



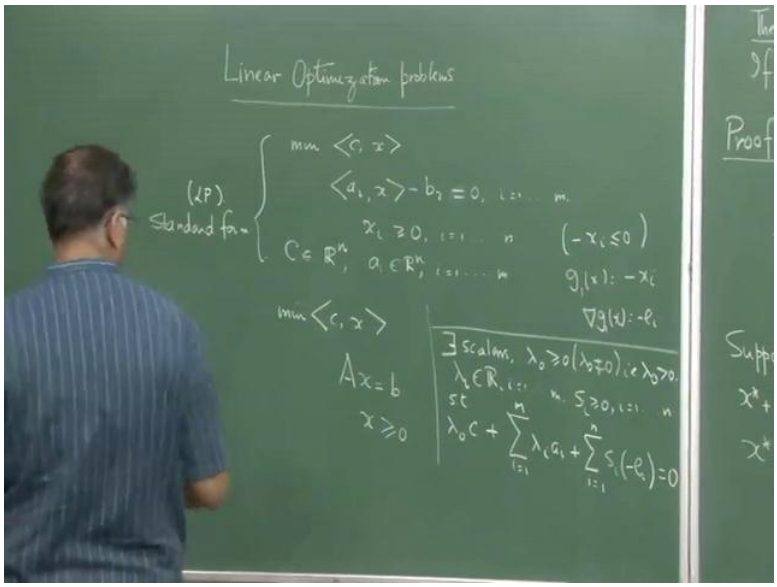
So, basically what this equation gives you is that d is a vector which is component wise greater than equal to 0. Once, you have that fact, you can immediately see that x star plus lambda d is also greater than equal to 0 component wise. Now, let us write down let us put this whole thing here. Now, let us observe what I what will happen if I compute the objective function at this point, x star plus lambda d. Now, the theorem would be changed to this point, so I will just basically use this now. So, this would be equal to c of x star plus lambda times c of d.

Now, lambda is strictly greater than 0 and c of d is any way strictly less than 0, because d is a solution. And, c of x star is c of x star; so this we have added a strictly negative quantity to c of x star. Which means what I would get is nothing but a quantity; this quantity which is strictly less than c of x star. So, what, what it means that x star plus lambda d is a feasible solution at which the function value is strictly less than the optimal function value and this is something which is impossible, this is an impossibility; contradiction. So, this is what mathematicians, one of the finest tools the mathematicians has is called proof by contradiction.

Now, once you have got this fact, so this fact is correct; that there if x star solves l p there exists no d in R n such that this holds. So, this system does not have a solution and this system looks like the first system in the motzkin's theorem of alternative. Once, that is done, we will see that we can now write down by applying the motzkin's theorem of the

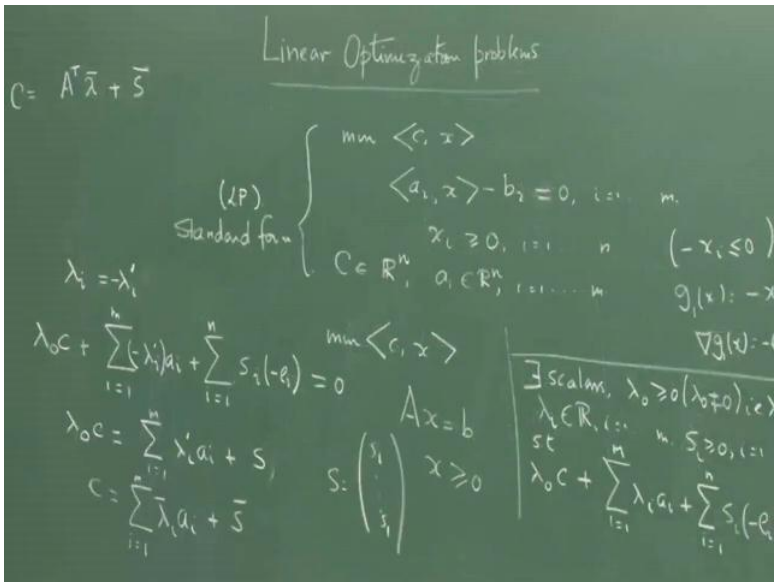
alternative, the LaGrange multiplier rule or the karush kuhn tucker condition whatever you want to call it or the fritz john condition. Now, I would again, now observe this fact that if this is the story that this system does not have a solution

(Refer Slide Time: 18:26)



Motzkin's alternative theorem will immediately tell me that there exists scalars, lambda naught greater than equal to 0 and lambda naught not equal to 0, which means, lambda naught is strictly that is lambda not strictly greater than 0; lambda i in this particular case, element of R i equal to 1 to m and S i element of because this is the greater than less than equal to 0. So, S i greater than equal to 0, i equal to 1 to n; such that lambda naught times c plus summation lambda i a i times lambda i a i times plus S i minus e i times, this is equal to 0.

(Refer Slide Time: 20:14)



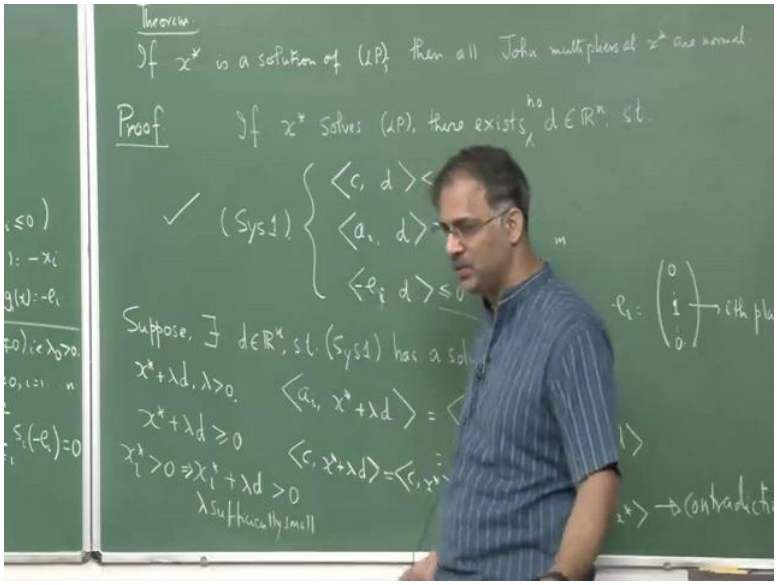
Now, what I will do in order to get something nicer and in the way people write in the actual literature, I will replace lambda i by lambda I will write this as lambda i dash because lambda i is just a real number; so I can just write it like this. So, I can basically add to lambda i, some number so that the thing is 0. So, I can write lambda naught c a i plus lambda i dash; so, it will become minus, minus lambda i dash. S I, e i will finally give me the vector s. So, lambda naught c is equal to summation lambda i dash a i plus S i e i that will combine to give me the vector s; where s is nothing but S 1, S 2, S n. Now, each e i is 1 0 0 0 1 all those things; so this is what I have.

So, now I will divide both sides by lambda naught because lambda naught is strictly greater than 0 to give me c is equal to, so I will write this lambda i dash by this as this. And I will write this as S bar; S divided by lambda naught. Now, what it is important to note is the following. Is that, this can be written in a matrix form because this is nothing but the matrix multiplication and you should be able to note that c can be written as a transpose lambda bar plus S bar.

Now, you know you observe it very carefully that this c is nothing but the gradient of this objective function, this is gradient of the constraints and these are associated with the gradients of this. But you must be thinking that there is some interesting feature of fritz john condition which is not here, I have not got that; complimentary slackness condition. Now, which how do you get that? I guess there are couple of ways to do that; but I will now

tell you one way and I will leave it to you to find some other way by looking at the affine version of the motzkin's theorem from (( )) to those who have access to that book. Now, what is important to know is the following. Here, what I really wanted was this greater than equal to 0. Suppose, I had x i strictly greater than 0, right, then I do not really, so I have x i strictly greater than 0 and what I really want to do is the following.

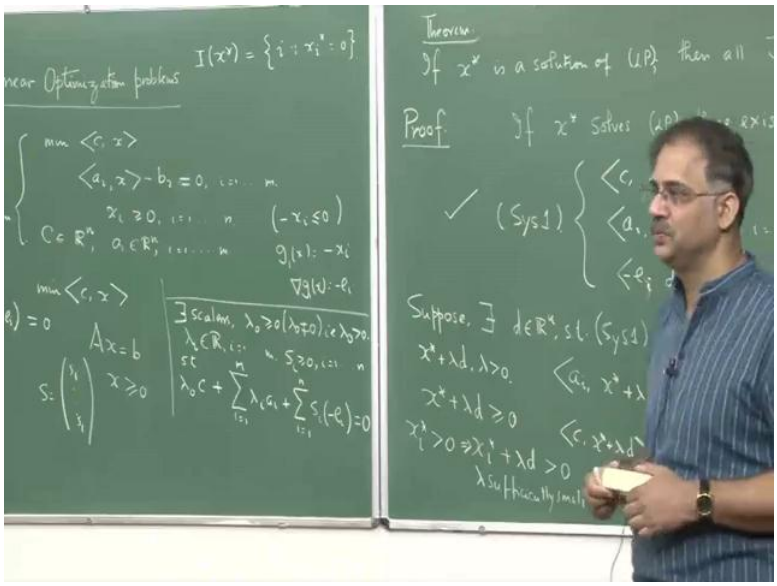
(Refer Slide Time: 24:45)



So, x x i star, so take any x i star which is strictly bigger than 0. Take d, d could be a negative number does not matter. Now, I now I have the controlling capacity on lambda. So, I choose my lambda in such a way make it so small that I can always have x i star plus lambda d, d could be negative to be strictly greater than 0, does not matter; does not matter at all. It will always have for lambda greater lambda sufficiently small.

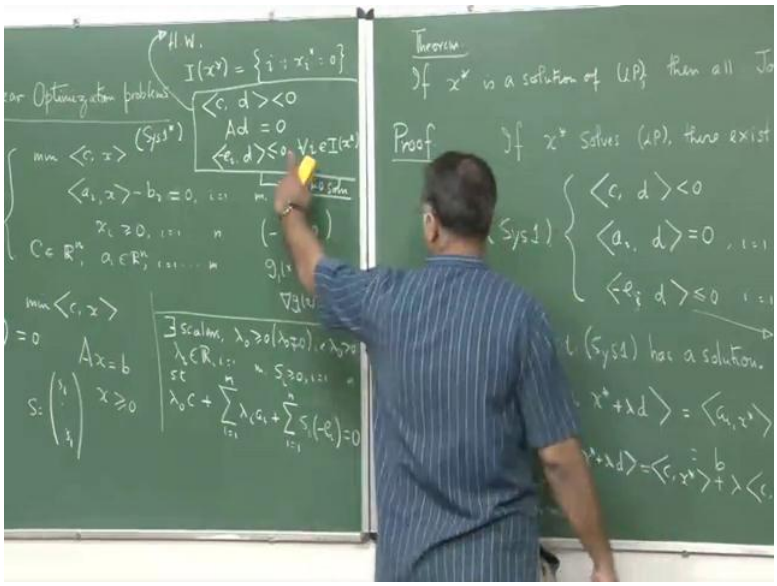
So, choosing my controlling my lambda I can always make x i star, whenever x i star is strictly greater than 0, it will always imply this by choosing lambda sufficiently small; lambda greater than 0. Now, observe that if I x i star is equal to 0, then I cannot do anything. Because, if delta is d is negative then this will be negative; d is less than equal to. So, d all the components of d is negative, then this will be strictly less than 0; so that will not be what we want. So, what we really need to look at are those points of x 1, x 2, x x n star, x star, x 1 star, x 2 star, x 3 star, x 4 star; that x star vector which are the components which are 0, that are the point where we really should concentrate.

(Refer Slide Time: 26:24)



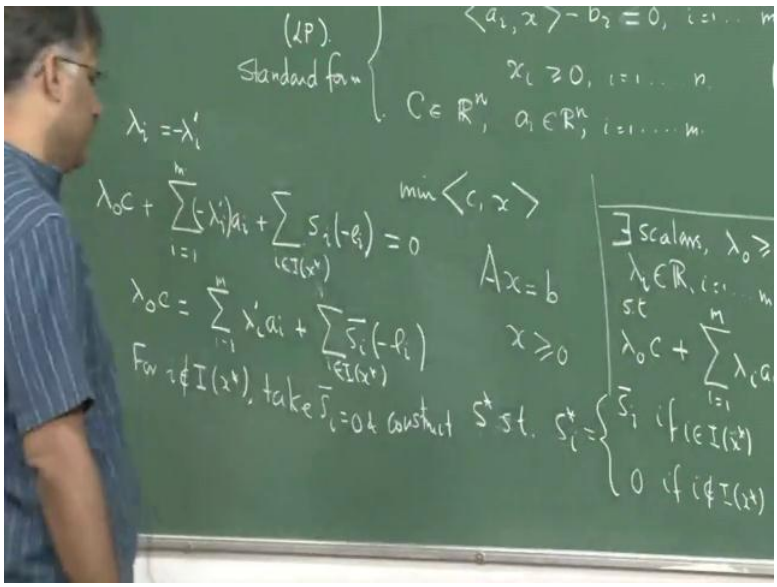
And, that needs us to consider what is called the (( )) index which we have already done in the last class also. So, consider all the i from 1 to n such that x i star is equal to 0. So, d, the d corresponding to those x i stars must be greater than equal to 0. So, this would lead us because we are missing the complimentary slackness condition and we are trying to push in, you know that those things there. Also, you could have taken here to be put one x i or something like that; that is a different issue. Now, what I want to now tell you is that this is, this is not exactly mimicking the way we proved the fritz john condition; this is not the, the exact one. So, what I am trying to tell you is that, now let us see we have concluded conclusively that d is greater than equal to 0 is required only when the case x i star is equal to 0. When x i star is strictly greater than 0, I have no problems.

(Refer Slide Time: 27:43)



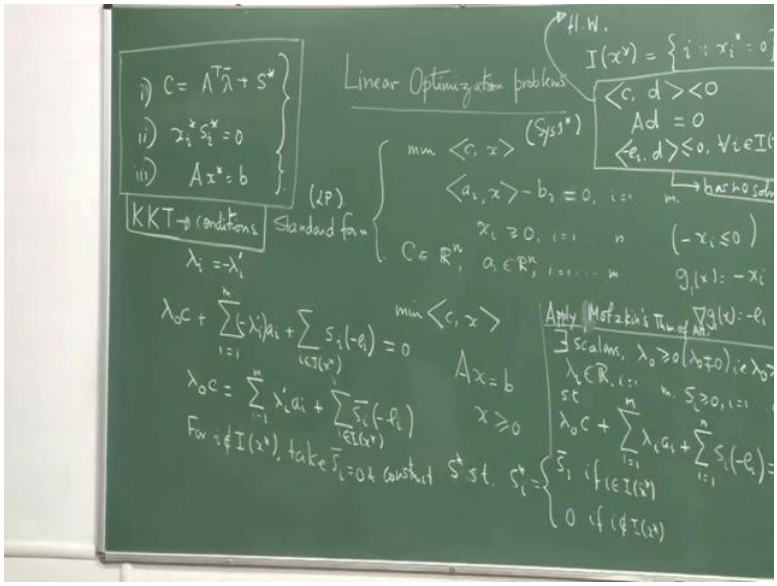
So, what I am now doing is I am now looking at this system. So, the same story I am now going to write it in a much more compact way, because you are now much more; I would only consider e i d to be less than equal to 0, sorry minus for all i in the active index set. Now, if I do that I can actually prove this fact. So, I can prove that this system which I can call as sys 1 star, this system has no solution; this has absolutely no solution. I would leave that as an exercise to you; please consider this as your homework, H W. Now, this system does not have a solution, so I can go back and make my changes here.

(Refer Slide Time: 29:00)



So, I will apply the same motzkin's alternative theorem that we learnt yesterday, but I will put this sys. So, basically what I will have now here is summation sorry i is belonging to i x star. Now, what I do S i bar. Now, we can write now, now for i naught in i x star take S i bar equal to 0. And, construct sorry S i bar minus e i, take that to be 0 and construct, and construct the vector S star which is S star, in such a way star such that S star i is equal to S i bar; if i is in i x bar, x star sorry and is 0 if i is not in i x star. So, in that way if I now compactify, what I have done here.

(Refer Slide Time: 31:22)



So, I can write this as lambda naught, lambda naught can be again in the same way. I can now write this whole thing as a vector S, S star. So, I can now compactly write this as A is equal c equal to A transposed lambda bar plus S star. Now, look at the nature of S i star. See, whenever i is element of i x bar, s i star x i star is equal to 0. So, x i star into S i star would be equal to 0. x i star into S i bar which is same as S i star would be 0. Now, when i is not equal to i x bar, x i is strictly bigger than 0 but S i star is equal to 0. So, which means that finally, I get what is what would you and I call the complimentary slackness condition; which says this is equal to 0. And, this is essentially the two major lines of the john conditions. And, you see that lambda naught, whatever way you get this is the only way to get the multiplier rule by applying the separation theorem which is the motzkin's theorem.

So, here I had forgotten to write, so I have told repeatedly what just for your memory I am just writing it at the last point; apply motzkin's theorem, motzkin's theorem of alternative

or a l t shortcut anyway. So, this is of course, people would like to extend it it like this, writing A x equal to b, A x star equal to b which is of course, any way has to be true; if x star is a solution. So, this is, this is essential system that you have to solve to find out x star. So, this is usually called this though this is the john multiplier, fritz john multiplier rule that we have got; you see the multiplier associated with the objective function of the gradient is always 1, a positive, which means, this always you cannot get it otherwise; this is what we call the karush kuhn tucker conditions.

So, we have learnt a very very important aspect of optimization today. The karush kuhn tucker conditons or K K T condition associated with a linear programming problem, which is essentially this fact that your x star is a solution of the l p problems and all john multipliers are normal. Now, as I told you that we had been studying algorithms for unconstraint optimization. And, while studying algorithms for unconstraint optimization we stopped at a certain point, we stopped after the Newton's method. We said that the Newton's method is some sometimes not very helpful because we do not know whether the hessian matrix is always positive definite at those i points of iteration. So, we cannot really get a descent direction at every x k. So, how can you remedy this situation?

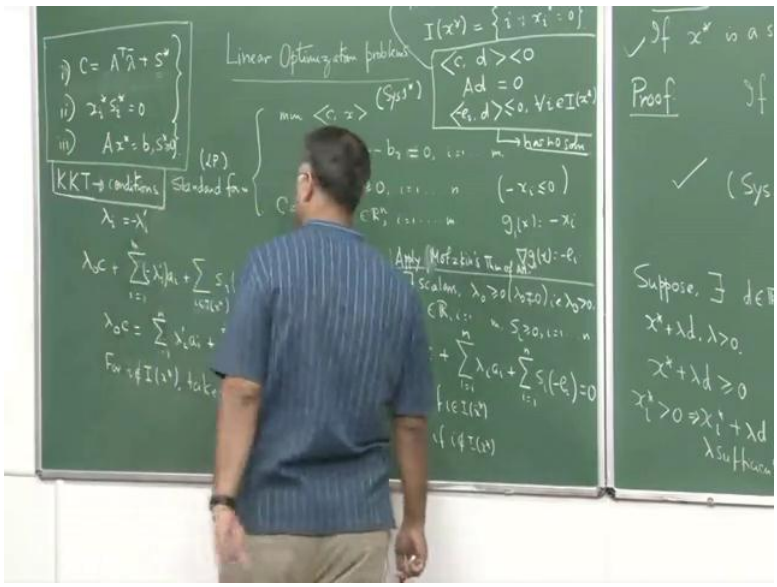
To remedy this situation we needed some modification, originally initiated by W C David and though there is an ironical story here which will tell you tomorrow. Before after I give the homework for this part on the KKT part, the ironical part is very fascinating W C David had actually invented the method but he is all, all papers following his actual work was published and his paper was published possibly the last among all the legendary works in that area. So, what I want to say is that there we remarked that if I want to study the improvements of the Newton methods called the quasi Newton methods or quasi Newton methods as some people would like to say, in that case we need to understand constraint optimization. All these methods have come after all these KKT conditions or fritz john conditions are known.

So, we need to use the optimality conditions there, the constant optimality condition. Hence, we shifted and reverted back our study and came to karush kuhn tucker conditions or the john conditions, whatever you want to call it. Now, once we have done that, we are now going to keep our promise and go back to the quasi Newton method. And, show that how these results can be applied to get a very interesting and conclusive theory. And, that would generate algorithms which are still very effective used in softwares are quite fast and

was considered as one of the revolutions in optimization in the 70s and early 80s. So, before all these other things like interior point method and conic programming and semi definite programming take over.

So, we what we are going to learn is essentially a revolution in optimization carried out by few great researchers in the subject. So, with this little fact that you have learnt today which is a I think a very very interesting fact to go through this step by step. Now, I I will leave you here with a question. Now, here I could go back and write such a system and do all these things because of the fact that I know the fritz john conditions at the very beginning. Since, I know the fritz john conditions I am writing this story; that is also true. Now, you can tell me one thing very clearly that I I will also leave you this part that to write the theorem in a nice way. What we have shown is that if x star is a solution, there would exist a multiplier lambda naught equal to 1, lambda bar in R n, S star in R n plus. Of course, here we have to note that this has to be in R n plus. We have written it here.

(Refer Slide Time: 38:25)



So, you can also add it here, wait S star greater than equal to 0; so such that these conditions are satisfied. Now, here because we knew about the fritz john condition we could you know gauge this change and make this change here. So, this is one way to go about it; this could be looking slightly artificial to you. Because suppose you are just given a linear optimization problem and you do not really know anything about the karush kuhn tucker conditions; how would you deduce an optimality condition? This is the question I

am keeping in front of you. You will get this solution in the FAQ but this is a question I am actually keeping in front of you for you to really ponder. So, with this I will end my talk today.

Thank you very much for listening.