Convex Optimization

Prof. Joydeep Dutta

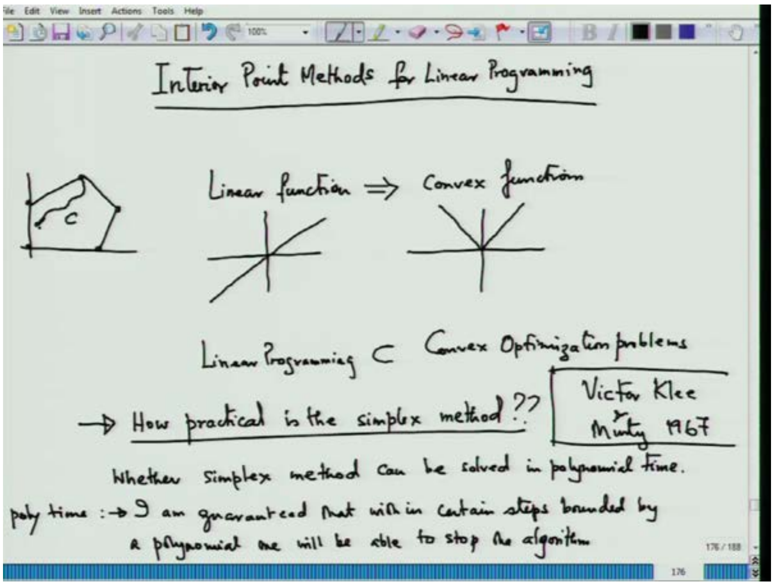
Department of Mathematics and Statistics

Indian Institute of Technology, Kanpur

Lecture No. # 27

Welcome, once again to this course on convex optimization. We had been discussing for the last few lectures on linear programming problems. Now one might know that in this group of NPTEL courses, that you are that you can see over the internet on through the you tube, there is a course on linear programming; largely based on a Minatorial approaches, stemming out from the simplex method. Then the question would be why I need to discuss linear programming here, separately. Let us understand that this course on convex optimization is about the theory of convex optimization, as well as the solution methodologies, and trying to really tell you about the most important class of convex optimization problems.

(Refer Slide Time: 01:18)



It is important to know at the very outset that linear a linear function is a convex function, which I am surely understand by now. Here I have two pictures; so every linear programming problem is a special class of convex programming problems. Now, question is, how practical is the simplex method that we have just learned. Now the simplex method that we have just learned, one is to remember that, it it is not the run of the mill approach that, textbooks take with tabulous.

I cannot say run of the mill, but the standard sort of approach one should say, that the textbooks take, but here what we did was approach due to Manfred Padberg which does not use the tabulous. But just does a updating based on very simple ideas, and that is exactly what the simplex method does an exactly what linear programming solving algorithms would program at their back. It is solving soft-wares have programs which is based on such an approach. Computers are not calculating tabulous; that is something we have to remember.

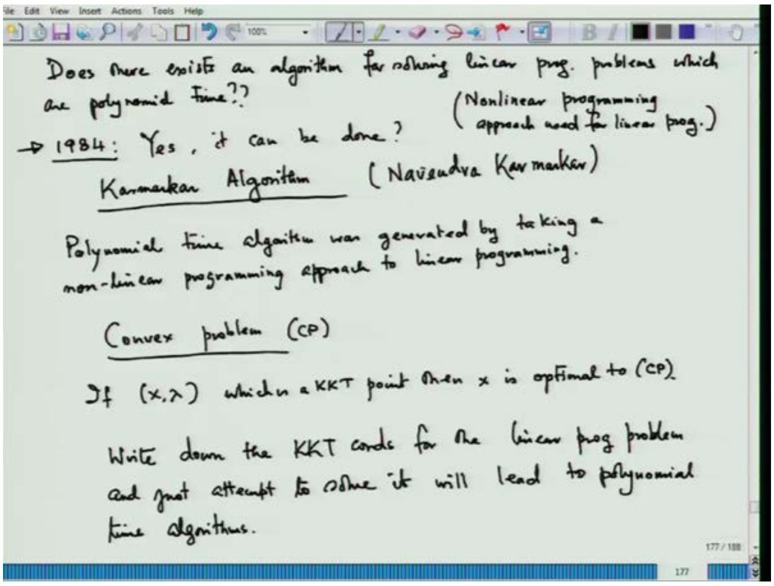
So, the approach or the linear program done here is different from the course that we will see, all possibly; they are already on the internet, and here we are now entering into one of the most exciting methodologies for linear programming problem which can be also extended to convex programming problem; we will see such an extension for semi-definite programming problem just after few lectures. Once we study a bit about interior point methods for linear programming problems. Now, interior point methods is slightly different from the simplex method, that you have studied; just to give you a little idea if you have a convex polyhedral like this. So, if, this is your c, what the simplex method does is to move from one vertex to the next; not just doing moving from one vertex to the next arbitrarily, but doing it in such a way. So that the new at the new vertex, the objective function value either remains same or it goes down.

In the interior point methods, we start from a point in the interior of the feasible set; that is all the components of x are strictly bigger than zero. And then we move along certain path, described by certain equations towards the solution. So, you move in interior. So, that is why it is called the interior point method. Now, the question we ask here you see, it is written, how practical is the simplex method. Now, one can say that, come on. Simplex method is very, very practical; there is the huge amount of industrial problems which have been solved by simplex method; real mathematics by the way, but here we are meaning not practicality from the practical point.

the From the point of view of practicality in the sense of applications, but we are talking about practicality in the sense that, from a algorithmic point of view, is it practical? That is whether, it is polynomial time; that is whether you know that or whether you can guaranteed that. In a certain steps, which you are going to finish standing this algorithm, that is the number of steps required to reach your desired solution, your level or accuracy. Can is bounded by some polynomial; that is, it cannot be arbitrary large

Victer Klee and Minty 1967 showed that, if you, there are certain linear programming problems which can become really very bad. So, the time complexity in the sense that, the number of iterations required to solve such problems as showed and be exponentially large. And that is not what is practical from the point of view running algorithms. Now, the interesting part is that, the interior point approaches; guarantee you a polynomial time algorithm for linear programming. Now about, what are interior point approaches? The interior point approaches, apart from the writings that you see there; the interior point approaches is to use a non-linear programming approach to linear programming.

(Refer Slide Time: 05:29)



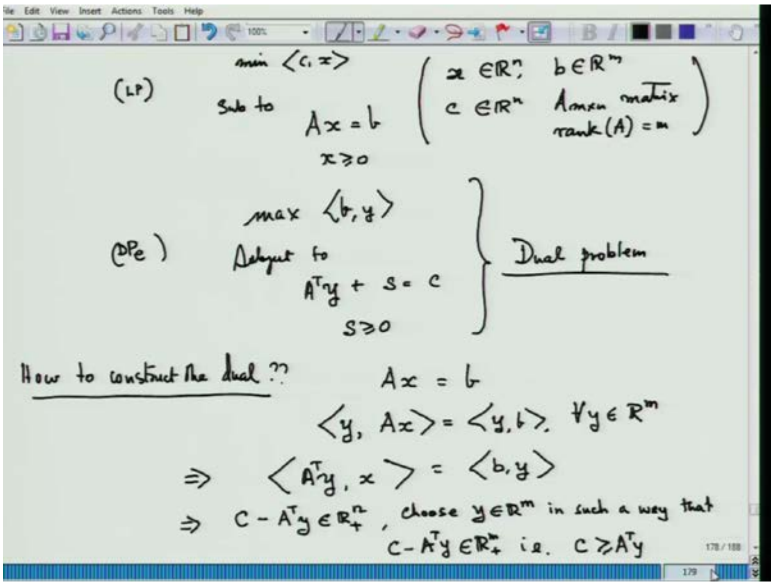
So, what does it mean by this? The first revolution, in this area was in 1984. Linear programming almost people thought that, it was just you know simplex method and its variations; and that is enough; and you can say bye bye. It is nothing much to do about linear programming, but to have built in soft-wares and apply them; but in 1984 there was a revolution which revitalized linear programming. And also bought back convex optimization to the center of all the action in optimization theory and applications, because people were so obsessed at the time in the 80s, about non convexity; they hardly bother about convexity; convexity was taken to be something very classical; more applications are in non convexity, and so we should bother about non convex case; not the convex case. But this revolution bought every convexity back to the center stage, and this was a famous algorithm called Karmarkars algorithm, due to a mathematician called Narendra Karmarkar, who now stays in Pune.

We should be proud that he is an Indian, and he when he was in bell labs, he had he did this interesting algorithm. But of course, I would not take my time to, but there's lot of keeping in views of limited number of lectures. I have on the subject, that I cannot really spend time on Karmarkars algorithm, because it is too detail; need a needs a lot of detailing; needs a lot of ideas, because it needs lot of geometric concepts are involved here, especially from projective geometry, projective mapping, etcetera which might deter the viewer, specially who are not so mathematical in client.

So, we will go and take a different approach. After 1984, what happened, Karmarkar made a comment which could be which may or may not be too much of over sitting. He said that, in many cases, his approach, his Karmarkars algorithm is much, much faster than simplex method. So, this thing inspired of sparked a revolution, and this thing lead to many more researches in the area of non-linear programming; many more developing the ideas of Karmarkar, and many more really looking at very different approaches.

The approach, that we are going to deal here in this particular lecture comes out of very simple fact; a fact which we have already studied. We know that the Karush Khun-Tucker condition for convex optimization problems is not only necessary, but also sufficient. So, if you know that there is a feasible x, and there is a Lagrangian multiplier lambda which is which gives you a KKT point, then x is an optimal to CP; and the interesting fact that once you write down the linear programming problems KKT condition; and try to solve those KKT condition, you are getting a interior point method which is polynomial time; and that is the strength and the beauty of this approach, because this approach can be understood by most viewer, because this is simply a question solving.

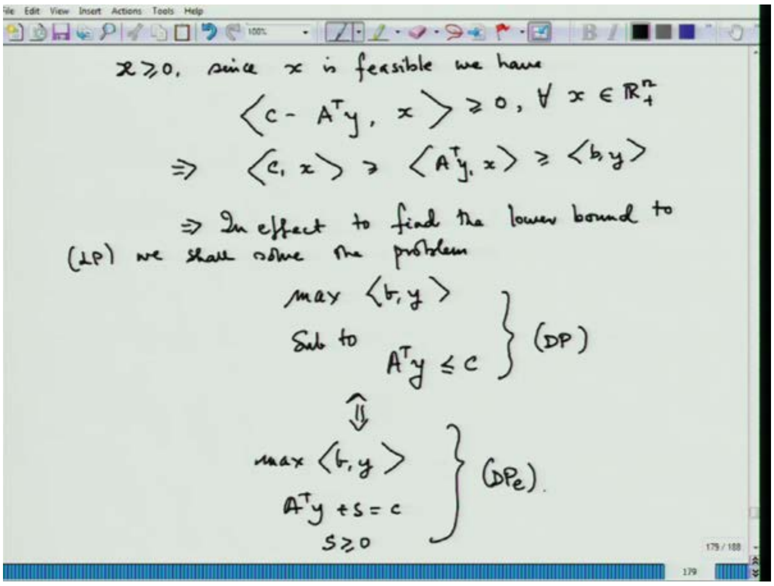
(Refer Slide Time: 09:33)



Now we go back and write down the linear programming problem LP and it is dual. Now, I just want you to remind that, given this linear programming problem in the standard form; this is the dual which we had already calculated, when we were studying the Lagrangian duality results. So, the lagrangian dual of this problem is this. Now, you might have forgotten it by now; what you tube is of course, you can go back something which in which you can go back and have a look, but now just to recall, you has to how to build this dual; instead of going to lagrangian duality, I would like to rather show you, a simple approach just directly taken off from the primal problem to show that the some problem going on at the back of it.

So, how to construct the dual? Now, you have Ax equal to b; it does not matter, if I multiply with a element, multiply by taking inner product. So, this is always true for every y. So, this implies I can write. Now, what I want is that, if I want to find a lower bound to this problem, the primal problem, how can I find it? If I know that, there is a lower bound, there is the famous theorem, and we have also proved it here that, it will show us that this has a solution so now observe a very simple thing. The simple thing is this; let me consider C minus A transpose y to be an element of R n plus; that is all the components of this vector is greater than equal to zero; that is choose y, element of R m in such a way that, such that C minus A transpose y is in R n plus; that is c is bigger than equal to A transpose y. Now, if you observe, C minus A transpose y is in r m plus, and because x is a feasible point.

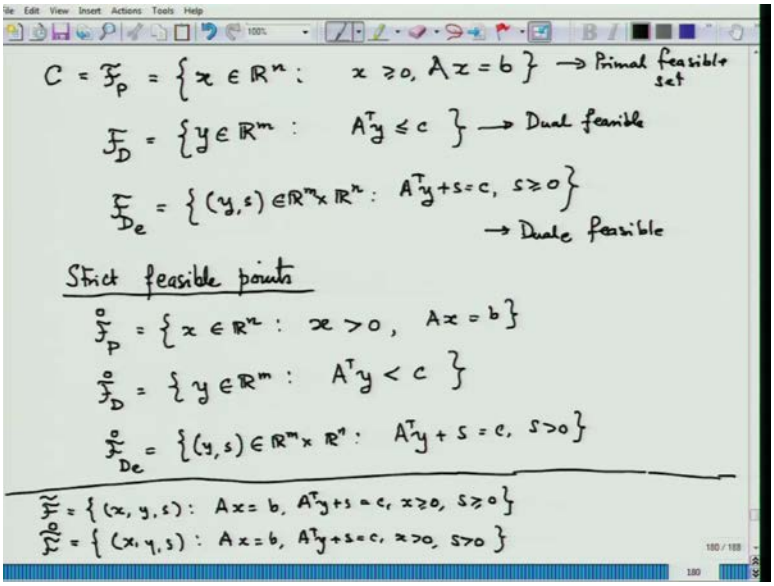
(Refer Slide Time: 12:22)



So, since x is greater than 0, since x is feasible, we have C minus A transpose y, because all the components are non negative. You know product with another vector was all components are non negative will give me greater than equal to zero. And this would immediately show me that C of x is bigger than equal to A transpose y x which is bigger than equal to b y. So, if I can fix up some y like this, then for this given feasible x or whatever feasible x, I choose, I will be true; I will be getting a lower bound. So, for all x in R n plus; this is true. So, for all feasible x, whenever x equal to b, and x is equal to greater than equal to zero; if I can find A y, such that c minus a transpose y is having components all greater than equal to zero, and b transpose y or b in a product y is giving me a lower bound of the original linear programming problem.

So, that is something fascinating; so, which means that, this is true for every such y, which satisfies this. In fact, if I maximize this now, this will that will also has maximum will also has supremum will also has a lower bound. So, in effect to find the lower bound, we shall solve the problem; maximize subject to A transpose y less than equal to C, but of course, I can add a slag variable to make an inequality equality, and this would lead to max of which is exactly is A 1, I had written earlier as a dual.

(Refer Slide Time: 15:35)



Now, we will call this problem as DPE with means, we have got a slag variable here or may be just name changed this name a bit, DP equivalent. right And the original DP which comes out naturally is this one, DP. So, DP and by adding the slag, I get a DPE. Now, this will give me. Now, once I write down these things, I need to put down some notations which would be useful, as I study interior point methods. The feasible set FP, instead of writing C; now, we will start writing the feasible set as FP, because that is why, because this is the standard notations that you will get when you look at the literature in interior point methods. So that, if anybody here is interested to go more into optimization, and really look look up the literature. Then this is the symbol you will get, symbol you will get to when you study the interior point methods.

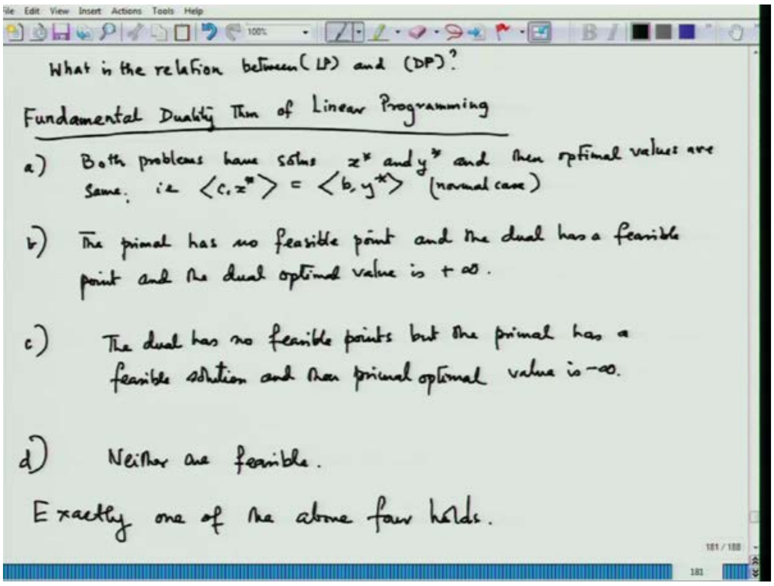
So, this is, now, of course, for the dual problem, the feasible set, the primal feasible set you can call this; sorry dual problem, you have y element of R m, R m; such that, A transpose y is less than c; now, when we write the dual problem in the equivalent form, that is I want to write something like this, FDE; you must observe that, the feasible dual feasible set is in a higher plane in the sense that, it is in a higher dimensional space. Because now, it will have y and s; so, it is R m cross R n with A transpose y plus s equal to c, and s is greater than equal to 0; apart from these two sets, we would also require its interiority; that is, we will require the what we will call the strict set of strict feasible points; we will have to put some sort of interiority conditions; So, this is called the dual feasible; dual equivalent, dually feasible, we can say; we are just inventing some symbols, but that is what it is.

One must remember, when we add a slag to a dual problem; though the problem at the end remains equivalent in the sense that, you have the same solution. But the feasible set is no longer equivalent in the sense that, they are in two different spaces; you cannot set this coincides with this. So, any y which will satisfy this, there would be a s for which y s would be in this; for any y s which satisfies this, that particular y will satisfy this; so, equivalency in this sense but not in the sense of exact matching of the sets; exactly equality of the sets. So, we will talk about strict feasible point; these are some thing; strict feasible point will basically you have you have been pushed into the interior of the feasible set.

Now, the same thing, well for primal, one would have x in Rn; now, x would be here strictly bigger than zero, means the interior of R m plus, that is every component of the vector x would be strictly bigger than zero; remember again, it is not greater than or equal to zero, but every component of the vector x has to be strictly greater than zero; this curtail difference is very, very important, and we will play a very fundamental role as we go along. And then, it remains the same; you do for the dual; just remember, wherever there is an inequality of greater than or less than equal to type, you convert it with the strict inequality; the inequalities replace within stricter version. So, this is for the equivalent problem.

Of course, there are some authors, who analyze the whole thing in a more combine form; that is they will choose the primal dual solution set. So, here instead of how do differentiate with this, you might say, you have to take in the same symbols. So, in order to differentiate, we will just put a round circle on the top, and that is exactly what is been done in the literature. So, some authors would refer to have a combine primal dual feasible set; that is they would like to write this as and of course, there is a stricter counterpart. So, we would continue to use any of any of the formalism, but we will largely focus on this sort of formalisms. What is the relation between the primal and the dual? That is the question.

(Refer Slide Time: 22:00)



What is the relation between LP and DP? Now, you know that, if status conditions hold in the primal problem with the primal which has the non empty feasible set; and primal has the solution; and the dual also has the solution; should have a solution; something like this, what you know about convex programming problem. That is the Slater condition hold, and the primal has the solution and the dual also has a solution and their solution values are same; their optimal values match, but the interesting part of linear programming is that, you do not require constant qualification linearity; itself is a qualification condition.

So, what we did in the case of convex programming problem was to use the slater condition, and make a proof of what we call strong duality result the which is the equality between the maximum of the dual and with the minimum of the primal. But for the linear programming problem, we do not need any conditions and for that we need a different sort of proof which depends on alternative theorem called Farkas lemma, which we will also state, but we will not do the proof now. For what we will write down is the Fundamental Duality Theorem for linear programming; we will write down the results and look at its consequences, but we will not go into the proof at this moment. We will do it tomorrow.

If we feel that tomorrow is also, it will get too much complicated, we will do it, may be after one or two days when we get more or less habituated with this stuff Fundamental

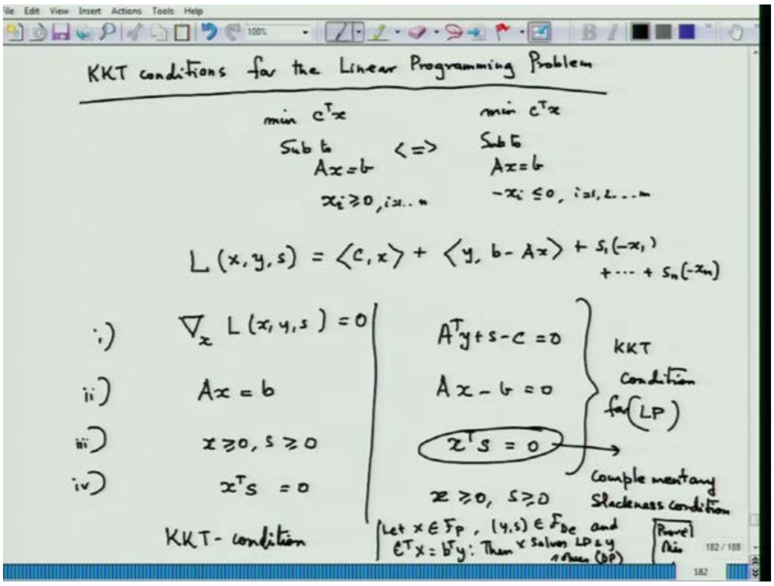
Duality Theorem. So, let me write down the fundamental results of duality. So, the four cases can arrives, number a. So, we will write down as a, b, c, d. So, number a, both problem have solutions x star and y star; that is they are they are non-feasible and then the optimal values are same; that is So, x star is a primal solution; y star is a dual solution. I understand that you know, by now that x star is a primal variable and y star is a dual variable; this is obvious, and I do not want to make a special remark about it.

So, this is equal to the minimum value of the primal equals to the maximum value of the dual. So, this is some authors would like to refer to as normal case. So, I the results that I am writing down here are from a book by Wilhelm Worst and Diether Hoffmann called Optimization Theory and Practice. This is the very, very good book for anybody which intensed to do a graduate work in optimization. Lot of things to learn from this book; can this book can be used even by expert optimizer.

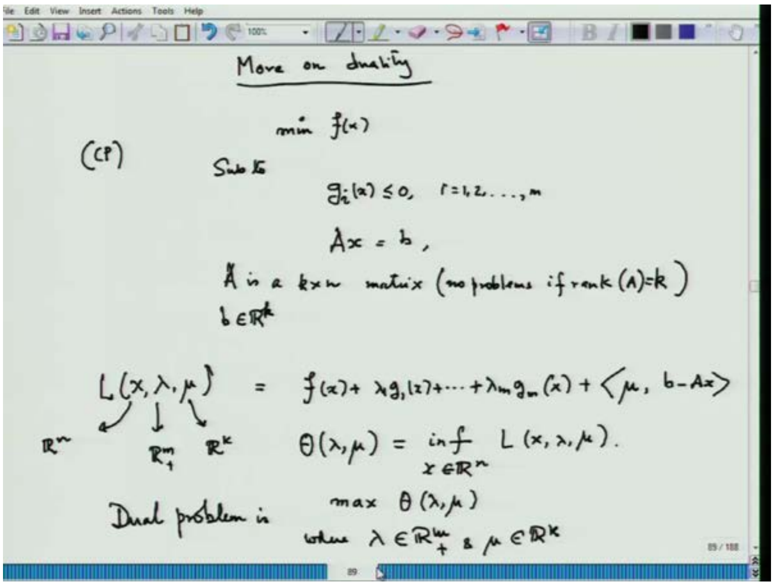
The second thing is that, the primal has no feasible point; and the dual has the feasible point; and the dual optimal value dual optimal value is plus infinity; the dual has no feasible solution, but the primal has a feasible solution; and then the primal objective value is minus infinity; cannot say it is minus infinity; this is the wake statement actually. Other primal optimum value is minus infinity. We cannot say that, what we say that the this is unbounded, basically. Primal primal problem is unbounded below; dual problem is unbounded above, which is short cut way of telling this is dual optimal value is this; primal optimal value is this; and the last one is both of them are feasible

So, if both of them are feasible, what is happening? If one is feasible other is not, what is happening? And the last is both of them are not feasible. So, neither are feasible. So, these. So, exactly one of the above four holds. Now, we have to say, this is what we know about the link between the primal and the dual; this will be very, very fundamental; this will give us some idea, when we write down and study a bit more. Now, the question would be, how do I device my optimality conditions. What is the optimality condition for the linear programming problem?

(Refer Slide Time: 29:22)

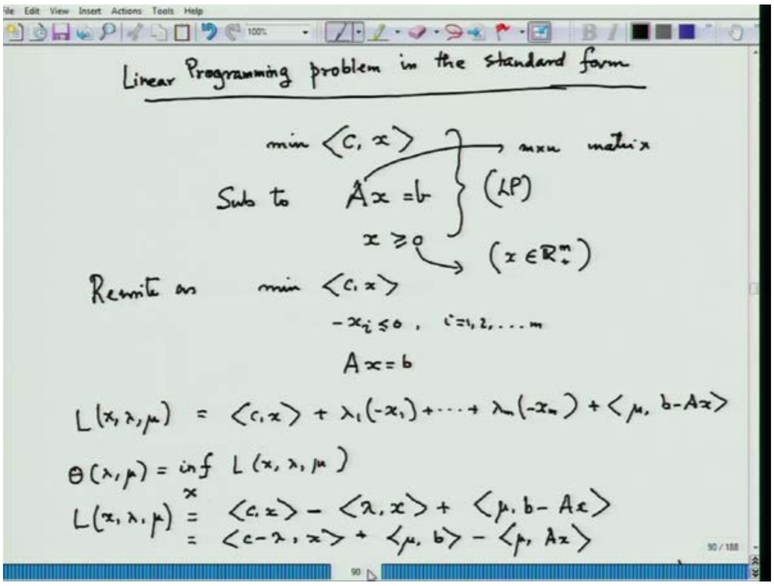


(Refer Slide Time: 30:00)



But the KKT conditions for the linear programming problem. If you go back at the saddle point theorems, I would just like to recollect; the saddle point results give me a short. You see, the saddle point results here g i x is less than equal to zero; if in for the standard l p problem, I have a x equal to b; l my x x less than equal to greater than equal to 0 can be posed as minus x; minus x less than equal to 0, and f x is nothing, but c transpose x, and slater condition in that case automatically holds, and since a is taken to be full rank, then we can automatically get the optimality conditions following this thing.

(Refer Slide Time: 30:37)



So, let us go back, and you see here. See, linear programming in the standard form we have written it in this way, and in this case also. We have write written down the dual, and what we are now going to say that, we can use the saddle point results, that we have here; the saddle point conditions to write down about the duality in linear the optimality in linear programming problems.

Now. So now, my condition is, my problem is, minimize A x equal to b, and each x i is equal to zero; each of them consisting of, this is g1, x1, x2, x n equal to x i, greater than equal to 0. So, minus xi is the g right. So, if i. So, I can write this equivalently as, here i is from one to n. So, you can see slater condition actually holds, and I have taken this to be full rank. So, all this are linearly independent and so, once I know this fact, so i can construct the lagrangian.

In this case, let us go back and see what sort of lagrangian we had constructed; see this is the sort of lagrangian get constructed; this would be the minus xi that take in this space, and this is the way we have constructed the lagrangians; and this is what we had. So, we go back come here to our study. So, we would keep some instead of lambda I am putting putting s here, because I the lagrangian multiplier associated with this is nothing, but the duals lack. So, let me just write it down C, x plus y into b minus x plus minus s1 into x1 minus s n into x n; this is exactly or s1 into minus x1 may be that; that is the better one. I think, that is the better way to write s1 minus x1, the way we had written earlier.

Now, what does the karush khun tucker condition say; I have not yet done, karush khun tucker conditions for this very this case, because this would need some more results. So, let us go by some sort of indusion; some sort of cut feeling, I guess. Sometimes in math, you have to do use your cut feelings; it is not that every time. Why I do not wanted to go into the proof of this, because the proof of this would require, some sort of theorem of the algorithm which is called the Motzkins Theorem of the alternative or you have to use make some complicated application of Separation Theorem.

So, I do not want to deter you from the course, because I would just keep on doing some proofs. But the idea is to make the main ideas proof, other than carrying work down too much in the proofs. I would give you the proof of the strong duality theorem, just as an example of proofs are done in optimization And the use of certain variable important the results called Farkas Lemma which are nothing, but very special applications or separation theorem. But going, but what we had in case of the lagrangian multiplier rule, what we would have is, this should be zero.

Number two: We should be having x to be satisfying this; we should also be having, because of feasibility and forth, you will have; this is the complimentary slackness condition for the inequality constants. So, x s1 into minus x1 is 0; plus s2 into minus x2 is 0; plus s n into minus x n is zero; each of them is zero, individually. So, if you sum up, it will get this thing this is equal to 0. Now, once I know that, this is what the KKT conditions are? This is what are my KKT conditions; I have not yet proved them, but I am just writing just down from cut feeling; you should know that the lagrangian, whatever we saw lagrangian when you differentiated it with respect to s, it should be giving you zero; that is, the solution x, given the solution x, there is exist y and s, such that x, y, s; x forms a critical point of the lagrangian function value; lagrangian function, when y and s are given.

So, here what we will do now is that, we will write down this slight, this explicitly; this will just give me look at the whole thing; it will give me because it is in terms of x. So, it will give me, and then this is, as it is, I can just change it a bit KKT condition for LP, but what is this KKT condition? What have you done? What we have first written the dual feasibility? Then we have written the primal feasibility. The dual feasibility first line, primal feasibility and just we have this additional condition called the complementary slackness condition; now, the very important thing is that, this is the task of everything;

the complementary slackness condition; this is the link between the primal and the dual; now, just if you look at it, this is the system of equation.

So, I can just put this on the other side; I will make it slightly more good looking, and I will put it like this; sorry it was C minus; the first part was C minus a transpose y equal to s; the first, I made a mistake. It was C minus; here is the minus A transpose y was equal to s. So, now I am making it slightly good looking by putting it as, A transpose y plus s minus C is equal to 0; A x minus b is equal to 0; and x transpose s is equal to 0. So, you see, this three forms a system of equation, and this three, because it forms a system of equations, I can solve them by certain methods. The method is the Newton method, that will employ, but while we solve them, we have to keep an additional the i. The additional watch, that x every for every solution, x has to be this and s has to be this; that is s has to be greater than equal to 0; s has to be greater than equal to 0;

If I have not done this, then if i do not check this, then I am actually not solving the KKT conditions; I am just solving this system. So, to solve the KKT condition and I have to solve this system as well as these two inequalities. So, how do we do that? To do that, we have to do Newton's Method. So, instead of getting into the details of Newton's method today, because it will take quite a bit of time to explain to you, what Newton's method is. I will rather concentrate on speaking about Newton's method tomorrow, but today I would just like to you to have a look at it. And rather think of how does one can solve this problem? And how does one get KKT condition. Is there any way you can do it?

I would like you to have a thought about this problem. Is there some other way, you can thought of proving this? Is there some way out? So, if there is some way out, it would be good to figure it out or at least try to make, try to see whether form here to here, it is making sense. And try to see whether you can have you really understood how the dual is constructed right. So, these are the very basic things that you you will require; you also have to get use to this sorry this kind of sets, that we are using. So, all this things are required.

So, we will go to the Newton's method tomorrow and so, tomorrow the first job is to see, how to apply the Newton's method to solve this? Once we know this, then we will go into the proper issues of the interior point results, and before I leave, I give you a home-work. Let x be feasible; so, I am giving. So, let x be feasible to the primal, and y s feasible to the dual equivalent, and C and C transpose x is equal to b transpose y; then x solves LP and y solves DP. Prove this. This will be a home-work for at least this evening, as we wind it up today.

Thank you very much. See you tomorrow and start from Newton's method.