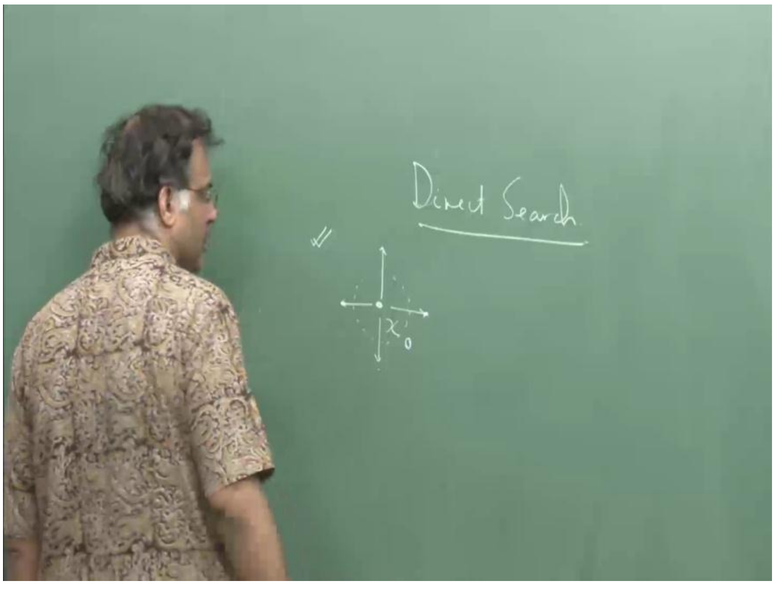
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 36
Direct search methods

Today lecture is a last lecture. In this lecture I am not supposed to tell you something new, I am not suppose to tell you something very difficult to advance. As a traditional NPTEL lectures, I am supposed to summarize what has been done give you an overview about is the subject in sense, and give an idea of what you can probably do with it. It is also important that I give you some possibly more home works, or little bit of some examples, if possible tell you about what has not been done. What I wanted to do because as you know there's number of lectures which is fixed.

It is also important to know that whatever I have done here may not be the most important things, needed by each and every one of you, who are listening to this lecture because there is a lot of things in optimization. And every practitioner needs something out of it. I really wanted to tell you something about direct search methods, in the sense that there is a heuristic method which takes some point.

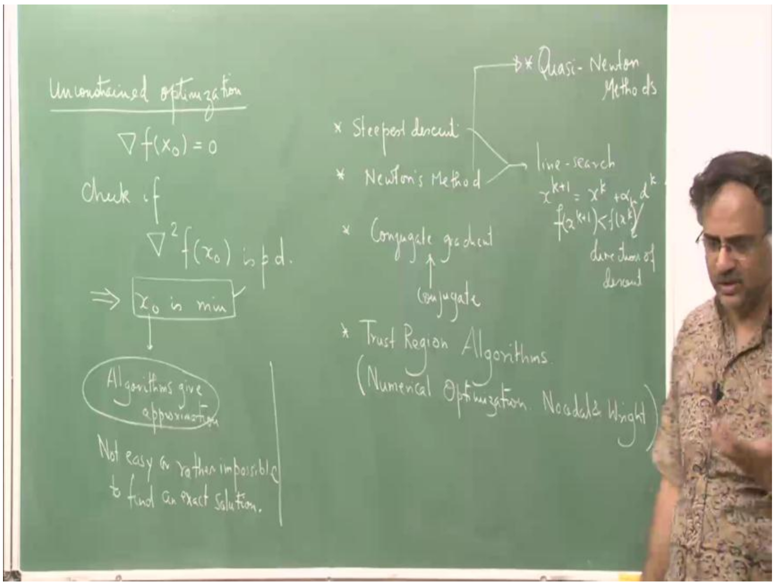
(Refer Slide Time: 02:02)



For example, I am giving a compass search it is got and look at a point x naught and see if the gradient is 0 at this point. So, this for different or does not matter even if you have a gradient information, you can do it without gradient information. So, about this point is not the optimum. So, you check for few points you move along this direction, move along this, move along this, move along this east, west, north, south. And see at which point function value is decreased take that point, if have not decrease a function value, right? Function values remains higher than this, then I have to reduce mine radius of the movement and then I again try.

So, these are called the one of the methods one of the direct search methods, these things appeared in the mid 60's, but they did not have a convergence proof. So, lost their clout with a mathematicians and they almost forgot and they later revived, that if you have differentialibility information for this sort of cases is going to actually, develop a convergence suit. Anyone to do something with that in this course, what this is not a time that you can actually do such things in the in the last lecture. You recollect in the course, we have started with unconstrained optimization.

(Refer Slide Time: 03:58)



Of course, unconstrained optimization is a mirror through which all over optimization can be viewed. So, the fundamental rule to find here is to find an x naught and then check if is positive divinity p d for if you can find an x naught which satisfy, this and this. And such an x naught is it will implies the x naught is minimum that is exactly, what is the thing that we learned. And of course, we learned through various algorithms how to find not exactly the minimum, but find some sort of a good approximation to it because it is very important to know except, for toy examples which we give in books like f x equal to x square.

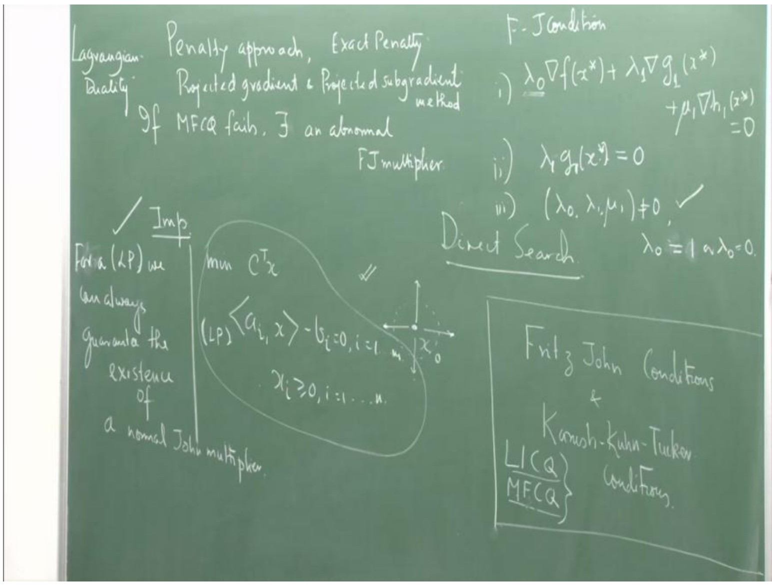
In actual problems, or even slightly complicated problems a fact which I am a want to stress, I am stressing in this course is that you cannot find the actual solution to an optimization problem. What you can find some sort of approximate solution, happy solution which you may like which you may not like. So, it is up to you. So, x naught is the minimum. So, x naught is a minimum here theoretically, but x naught to find to use algorithms. So, algorithms give only the approximations this is the fundamental thing you must remember, the algorithms give approximations not possible to, not easy or rather impossible in most cases an exact solution.

So, this is the fact that you need to remember, when you want to if you want to advance yourself in this subject optimization. Now, the algorithm that we did where is steepest descent, we spoke about the Newton's method, we spoke about conjugate gradient method. I want you to recollect these two line search methods. That means, these can be written as where d k is a direction of descent and you know what is a direction of descent, here it is slightly different here we are talking about conjugate directions. So, this is built upon conjugate directions.

So, there is another non line search algorithm called trust region algorithms, which we have not done. So, another important class at this moment a very important part of optimization research is trust region algorithms. So, I refer to the book by Stephen Wright which because it also an Indian edition, which published by Springer numerical optimization (( )), Stephen Wright. So, you can see a very, very good study of fundamental issue in trust region algorithms using this.

We also did another important the variation of the Newton method, which is much more useful to solve non convex problems is a quasi Newton method. Of course, you must remember that when we do the line search methods, we always have to keep in mind that we are expecting this, we want to do this. There is a interesting analogy between trust region algorithm, and quasi Newton algorithms because in trust region algorithms, need the use of constrained algorithms while, this also means use of constrained optimization to understand it to make its update, and this, and this, and this really does not need it.

(Refer Slide Time: 11:13)



So, now once unconstrained optimization was done, we came to the heart of the matter. We came to study the constrained optimization problems, where we studied the Fritz John conditions and the Karush Kuhn Tucker conditions. I would like to revise your memory, in the sense that a very important thing is to note at whenever, there is a normal multiplier for a Fritz John system that is same as studying, the at the Karush Kuhn Tucker conditions holds.

So, every normal multiplier, existence of normal multiplier is telling that Karush Kuhn Tucker condition holds. So, there were constrained qualifications like the linear independent constrained qualifications L I C Q at the Mangasariean Formuwich, Mangasariean Formuwich constrained qualification, which called the M F C Q these two always guarantee that if this is satisfied at the solution point, then there cannot be Fritz John multiplier, which is abnormal. All the Fritz John multipliers would have lambda (( )), I would just remind you the Karush Kuhn-Tucker conditions. So, let me take one constraint equality and one constraint inequality.

So, this is the F J condition, or Fritz John condition. So, what is done here if you observe the last seen condition is. A very important thing to note here is that if L I C Q and M F C Q is holding as we have said that there cannot be any lambda naught, which is strictly bigger than 0 see I have to write this also lambda sorry lambda 1 g 1 x, which is equal to 0, this called the complementary Satins conditions. Now, if these two any of two this happen then this can never be 0. Actually, in the Fritz John condition what we have is that lambda naught, lambda 1 and 0, 1 this is not equal to 0.

So, lambda naught is either 1, 1 means either greater than 0 or lambda not is equal to 0. So, if lambda naught is equal to 0 we can always rescale the multipliers with lambda naught that is divided by lambda naught and get this lambda naught to equal to 1. So, I can always write as this as. Now, you see this thing is guaranteed to be always strictly greater than this either this happens or this happens. This is the weakest condition, which guarantees that all the lambda naught are strictly bigger than 0, if this tells then it will be always confirmed that lambda naught is strictly bigger than 0, sorry, lambda naught I just want to repeat if this M F C Q fails there is always a set of multipliers satisfy this, with lambda naught equal to 0 that will there exists an abnormal multiplier.

So, a very, very important central thing in your learning is this if M F C Q fails, there exists an abnormal multiplier, abnormal F John multiplier. So, when M F C Q fails a abnormal multiplier is guaranteed. So, if all the conditions that will see in books like Arbitic constraint qualifications, INGOT constrained qualifications or the approach of routeneous in the recent times, they only do one thing they say that if my M F C Q is failed. My problem is an abnormal problem with an abnormal multiplier, but then does there exists a set of multipliers, which is normal that is lambda naught strictly bigger than 0. The answer surprising it to be is yes. So, this conditions guarantee that hold be at least one multiplier set with lambda naught strictly greater than 0.

So, the KKT condition would be satisfied that is why is or this things which are weaker than MFCQ are called constraint qualifications because is a constraint qualify, the condition then KKT condition is guaranteed. A very, very central issue is in linear programming. For a linear programming problem there is always exits a Karush Kuhn Tucker point or there without any constraint qualification. So, this a very, very important result.

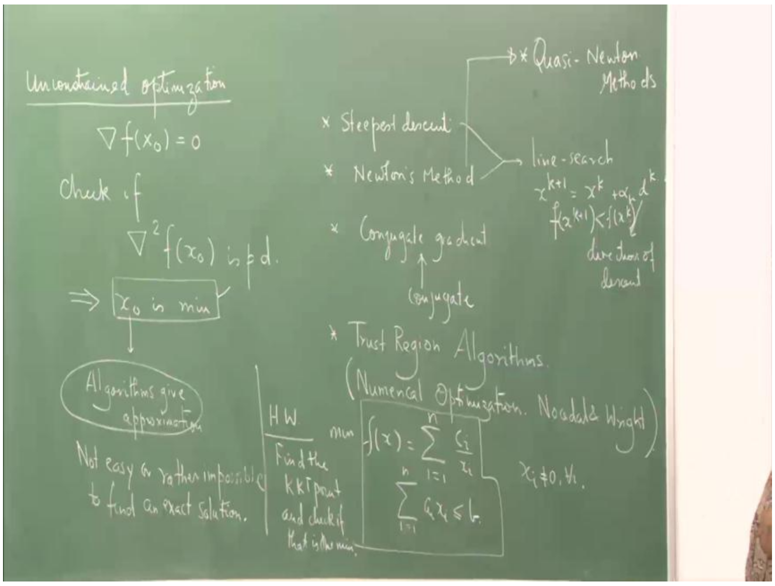
For an LP, this is called an LP an LP for an LP rather for a LP we can always guarantee, always guarantee the existence of a normal John multiplier, interesting part is if I take it this to be of any other function f x. So, differential function non convex also then if this constraints are linear, you can still write down the Karush Kuhn Tucker condition without any use of constraint qualifications that if actually, you have linear or a fine constraints then the existence of at least one normal Fritz John multiplier is guaranteed, this is a central thing a very, very important result.

The interior point methods which not the kermocost one, which is accord after that especially through the work of any got the use of Newton's method for solving, the use of Newton's method for solving this problem, this linear programming problem comes out of this very fact that the Karush Kuhn-Tucker condition always exists, there is always one set of multipliers which is normal. So, the KKT condition is always holding then we can solve the KKT system, very cleverly using the Newton's method, which we do not discussed here, but discussed in slight details in another course some convex optimization in NPTEL.

So, this a very, very important very, very central fact and this cannot be ignored. This is a very, very important fact possibly in the Karush Kuhn Tucker theory, this is the central fact I would say. Now, we also discussed something about and we have done some lot of examples, another book that I want to show you, which is a Springer text in statistic series book. This is a optimization make in available in Indian edition, but the usefulness of this is that not only the Indian edition because this book is written by statistician, who use an optimization work.

So, here you can see a lot of good examples from statistics and that is really a source of fun. Even for the optimizer to read this book it is really a big fun I would say. As a way a different way of looking Karush Kuhn Tucker theory. So, the thing I want to read at read it. So, let me just try to give you a type of problems that is useful in statistics and maybe you should be able to find the x in the solution.

(Refer Slide Time: 21:03)



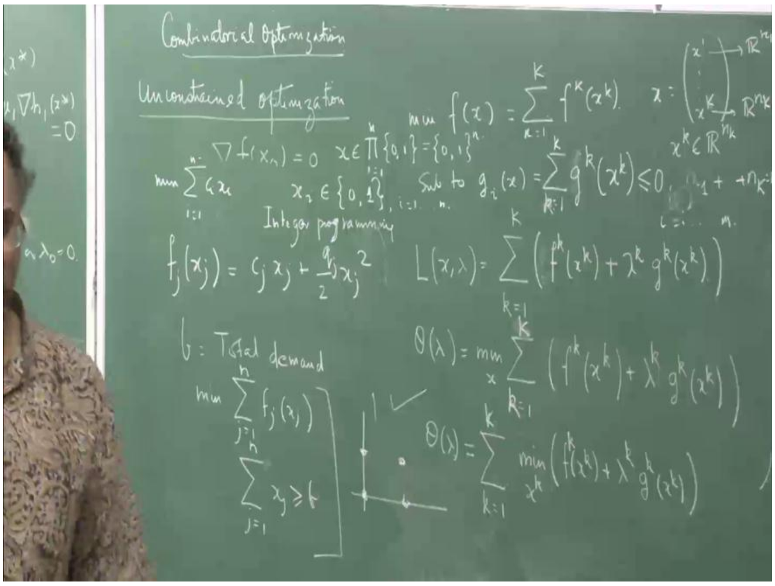
So, you are let me want to minimize f x. Now, here you observe the this is only the defined if x I is strictly bigger than 0 and x i not equal to 0 and other not bigger than 0, and I have and upper house space is constant. So, is gone a minimum of the linear reciprocal function, x i cannot be equal to 0 noise. See once x i is not equal to 0 this is a differentiable function. So, now find the KKT point and check if that is a minimum.

So, here is an example where you have to use the Lagrangian theory. Now, if you look at this problem I remind you again that once you have this KKT point, how do you check that is a minimum then you can always check the second order condition. First remember the second order condition that we have discussed, then you can check the second order condition here and see what you get out of it, or whether you can argue in some other way that is actually minimum that would be much more fun.

So, there many, many things that we have studied for example, we have studied penalty methods, penalty approach and spoken about exact penalty, we have spoken about the projected gradient method and projected sub gradient method. We have spoken about the projected gradient method and projected sub gradient method and we also studied in quite detail about Lagrangian duality.

Now, I will just write to you as a home work about ones problem, which you will see that Lagrangian dual, starting the Lagrangian dual of such a problem is actually very helpful. These are called problems, which can be decomposed or decomposable problems and. So, let me write down a problem let me write down its structure and you will see that Lagrangian, starting the Lagrangian dual of such a problem is more useful or simpler in for computation in the actual problem.

(Refer Slide Time: 25:01)



So, here you have to minimize the function f x and that f x is given as. So, it is x. So, there are. So, means this x variable is partitioned now into x 1 x 2 x k right and x k belongs to R of n k and n 1 n k is capital N. Basically, I am partitioning this vector into several blocks, each block here is a vector. So, this is a R n 1 and this R n capital K, this capital K.

So, once this is known let me now just write down the inequality constraints. ((No Audio 26:30 to 27:00)) So, this is a problem in a decomposed form. Now, once you have a this in a decomposed form then what is a Lagrangian, maybe I can put this be 0 here it is 0. So, this is where you will see a very good usefulness of Lagrangian duality. So, what would be a Lagrangian here, a Lagrangian would be ((No Audio 27:43 to 28:20)) this would be a Lagrangian. Now, you want to find the theta lambda ((No Audio 28:25 to 29:00)) this can be now written as minimum x h. Now, you see here I have now got to solve some smaller simpler optimization problems, but if I solved one of this optimization problems, the structure of the other is clear what would be the solution.

So, I have to solve a very simple optimization problem at a much lower dimension then n and if it is just a unconstrained problem now, again just for a fixed lambda I can actually solve this problem much easily. So, here you see there is a use of actually looking at the

Lagrangian dual, the computation of theta lambda is quite simple. I would give you the home work as to what would happen if I put b I here. So, if I put a b I here instead of 0. So, how what would be the writing.

So, here also you will have you will have the additional term minus b lambda, you do this and then whole thing you subtract minus b lambda, but let me just give you a problem. A problem in a decomposed form which is specially, used in the study power supply in electrical engineering, this problem actually comes out study of power demand in the sense that how can we satisfy the demand of power at minimal cost. So, if x j is the amount of power the j th power plant is. So, the n th power plants 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 if the j th power plant what should be the output of the j power plant. So, that the demand is met.

So, the demand is b say then the total amount of output of all the n power plants should be excess of b, that is the meaning that the demand is satisfied. So, suppose the cost of generating the electricity at the j th power plant, it does not matter if we put this at that just here. So, this is the cost it is the quadratic cost and b is the total demand. So, if x j is the output of the to have an output x j in the j th plant, this the total cost that you have to incur.

So, here it is strictly convex of course, suppose I would say that they can make x j is free any amount of j they can put theoretically, actually x j should have a bound and this problem is to minimize summation f j, x j and if b is the total demand then summation j equal to 1 to h that is what it should be. So, it is again a problem in the decomposed form, for which you can actually use a Lagrangian duality to get something.

It is very, very important to note that optimization is a very, very vast field, it is not that what we have done is abroad gamete optimization is just a mini scale gamete possibly. Optimization is not about what we have done is called continuous optimization, we have not spoken about variables about that could be discrete that is for example, you could minimize a objective function f x where x i, i at takes the value 0 or 1. So, for example, I just give you a simple problem like this I will just a linear programming problem.

So, x i takes up the value either 0 or 1. So, these are the problems called integer programming problems and we have not discuss them at all here, our x is always here R n rather than x b in the Cartesian product term. So, here x would b in the Cartesian product of n fold Cartesian product or 0, 1. So, 0, 1, 2 which is customarily written as. So, it is a sort of hyper que.

So, if you are if you are in 2 D then 0, 1, 0, 1. So, it is this point this point this point a grid basically, then these are the 4 points over which you need to compute the function values and list down the minimum. If it is only a 4 points like this right that is n equal to 2 then it is very simple just calculate the 4 points and get it, machine will give you the answer in blink of an eye. The problem is that what happens when n is very, very large and when n is very large things can be very difficult, you cannot make a total enumeration.

So, because you cannot make a total enumeration you must have clever ways to know which among this huge number of finite points is a actually, giving is the minimum that gives rise to a subject called interior programming or combinatorial optimization, here it uses methods from combinatoral and its approaches is completely different from the approach that we have taken in this course, and an approach of continuous optimization.

I want to assert a part of this department IIT, Mumbai the course that I have absolutely different (( )) of course, an algebra here we spoken about problems here you can sense that what sort of distribution that noise might fall and once, you may use techniques that are now which will you have listen in the two lectures, but can… the two lectures.