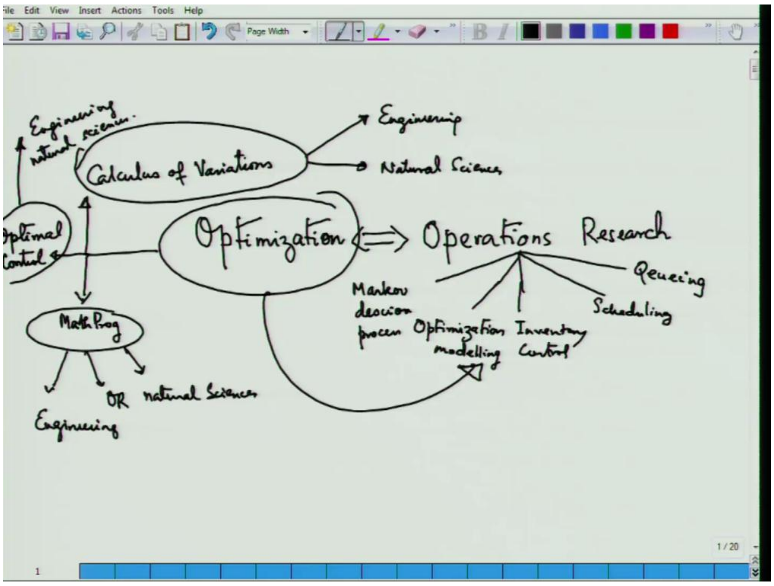
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 15

Before I begin today's lecture, I would like to remind you of certain things. In this country, there is a very big misunderstanding or rather a misconception I would say.

(Refer Slide Time: 00:32)



Misconception is that optimization as a subject, which is actually a part of mathematics; or if you want it to be slightly more politically correct or incorrect, whichever it is, it is part of applied mathematics. We equate it to or we rather say optimization is same as operations research. But this is far from truth, because usually in many, many places, operations research courses teach largely linear programming, which is minimization of a linear function over a polyhedral convex set. Of course, I have not told you anything about polyhedrality or anything like that. Yet… And there is a course on convex optimization, which is already live on youtube given by me; you can go and look at that course to have an idea what are convex sets and functions and linear optimization and linear programming,

Now, the question is this. So, people equate operations research with optimization; and there is a growing tendency among mathematicians in our country to say that, this is a part of operations research, which is largely done at business schools. So, optimization cannot be a part of mathematics. What the business school guys forget that, operations research is not just optimization, optimization tools are applied in operations research. So, optimization problems appear in operations research. And, many of the areas, many problems can be solved by optimization techniques in OR. For example, there is optimization modeling here in OR; there is for example, inventory control; there is scheduling; there is queuing – queuing theory.

So, there are many aspects of operations research, for example, Markov decision processes can be also taken as a part of operations research. So, these are the tools, which people in business school need; it is not the tools, which every other thing needs. For example, a problem in physics would need bother about inventory control. So, optimization is a subject, which is feeding these problems in business or operations research. It is pretty good to teach operations research at business school while all mathematicians forget especially in this country that, optimization is a very ancient subject. There are many many geometrical problems of maxima and minima, which we had studied in the very beginning.

And, optimization for example, in its fold has mathematical programming, which is finite dimensional optimization, which is actually applied to this place. Also, has in its fold a very important ongoing chapter in mathematics, the calculus of variations, which is infinite dimensional optimization. And, its modern version for more complicated problems is optimal control; not to say our other issues. So, these are highly mathematical subjects; this, this and this. So, optimization is a subject in mathematics, which is helping other subjects. So, calculus of variations for example, is used in engineering and other natural sciences. This is used in engineering, operations research, natural sciences; engineering and natural sciences.

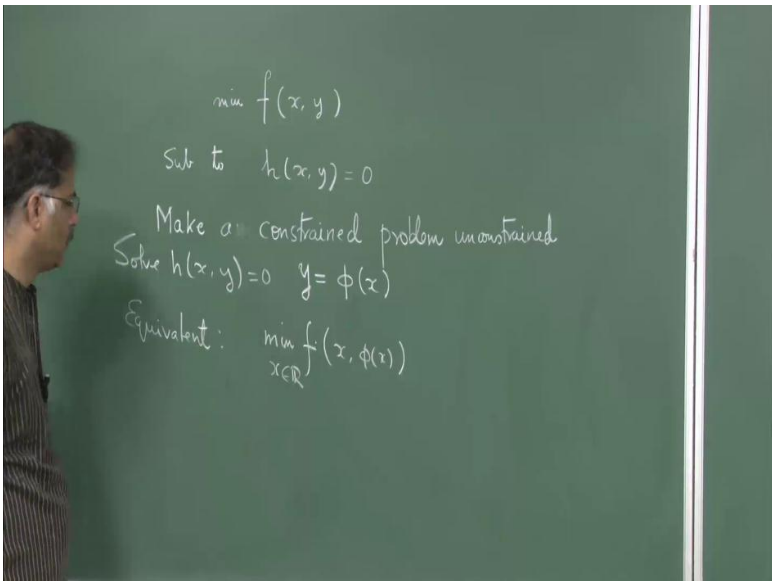
Optimal control problems are even used in biology. But, they are mathematically very involved topic; not only a beautiful part of mathematics, but a very very lovely and interesting part of mathematics. So, optimization is actually a very vast in one hand helping operations research and other hand helping other sciences, and been centrally a mathematical issue – centrally a subject with mathematics of breathtaking beauty. So, it is my request to those who do mathematics and want to divide as pure and applied mathematics; please note that, subjects like calculus of variation for optimal control is a part of optimization. If that if you consider – I do not know at all whether you would, is a part of mathematics, because one has to understand a lot of function analysis that come in because one may need it to answer or solve problems in the calculus of variations. So, lot of types of functions, spaces, etcetera have come up because of the requirement of calculus of variations. So, calculus of variations just like optimization in those days, had said lot of things into mathematics and let to lot of mathematical development.

See if you at all consider calculus of variations and optimum control as subjects of mathematics, optimization, which is the mother subject, which contains all these subjects cannot be a subject, which is not a part of mathematics as mostly said. And, the number of people teaching optimization in maths department in India is dwindling, and possibly across the world, in many many places. And, optimization is now largely is done in business schools unlike the US universities; the people who teach from the business school here largely are not maintaining the regard that you need to teach optimization (( )) mathematical (( )) because they are just told that, this is a tool that you have to apply; it has got nothing more to it. And hence, the mathematical beauty of the subject is lost.

Now, giving this a very basic and important fact, I also now, want to make sure and tell very clearly that, what or whom I am expecting to be a listener to this course. I do not expect the first year engineering students to listen to this course. Thus, engineering students should be at least in his third or fourth year to listen to this course; and a mathematics student should be at least in his masters degree first year to listen to this course, because it will be mathematically involved; and as we go along, the mathematical involvement would be more.

Of course, we would not at all go to infinite dimensional spaces; we will try to avoid the issues on non-differentiability as much as possible. But, still keeping into that level of elementary in a suite say, still I would request people, who are not mathematically inclined not to get into the details of this course; of course, you can just skim through and just see what is been told. But, lot of proofs, etcetera would be done. For example, now, we will start talking about the Fritz John conditions or the John conditions; and from here we will go along.

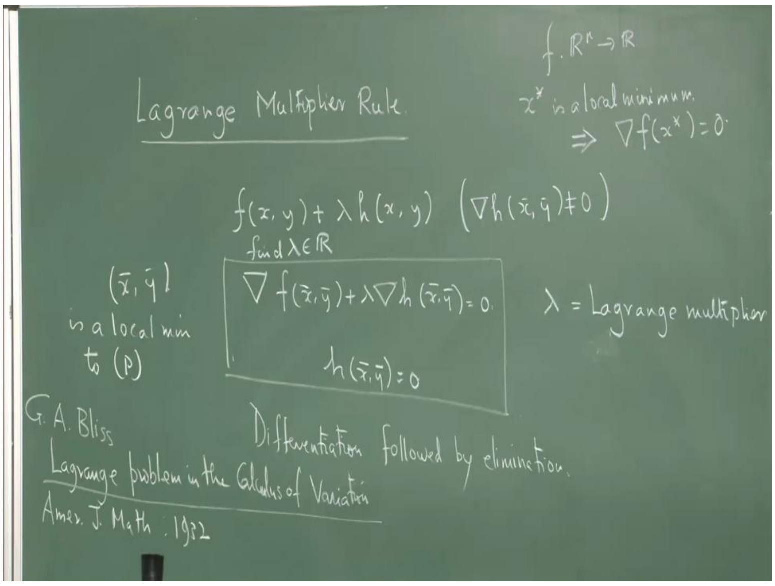
(Refer Slide Time: 09:08)



Now, let me start with a problem, which you study in calculus. In calculus, when you study function of two variables, you are usually asked to look at a problem like this, where you minimize a function of two variables – two real variables: x and y subject to another function in satisfying h x, y equal to 0. So, basically, take all those x, y's, which satisfy this. And, among those x, y's find the one which minimizes this function. Now, the problem is that, how do I go about it. If I am just given this problem, how do I start to visualize an approach to the solution? Now, the whole idea is that, we have some information about unconstraint minimization; which is a much more simpler thing.

The question is here I have constraints; so how would I go about dealing with the constraints? Now, in order to do so what I would do is to replace the constraints somehow and make the problem unconstraint. So, idea is to make a constraint problem unconstraint. So, how do I make it? Suppose I am enable to solve this equation h x, y is equal to 0; and I now can write to y as a function of x. So, from here, I solve h x, y equal to 0; and I can write explicitly… h x, y equal to 0 – I write explicitly and I get y as a function of x. Basically, I should write y equal to y x, but just… So, there is an explicit relation between y and x. Once I know the explicit relation between y and x, then of course, I can put in… I can replace y here with… And, now, I can just minimize over R; it becomes an unconstraint minimization problem in R. So, this is an equivalent problem. So, the equivalent minimization problem is this. So, if I can always do that, then why I need to bother about so much certain things, which you have heard called the Lagrange multiplier rule?

(Refer Slide Time: 12:16)



Now, if you look at this; what I will do next is to if the function is differentiable, I will take a derivative and equate it to 0. But, you have also heard about this thing called Lagrange multiplier rule, where what you did in the calculus course was to add the – just take some lambda as if it is coming from the air; and you added the constraint to the objective and found a new constraint – a new objective, which is now unconstraint. And then your job was to take the derivative first with respect to x and with respect to alpha; that is, find an h x, y. So, here you will get 3 equations – 1, 2 or 2 partial derivates near 1 equation. So, here you have 1, 2 real numbers and third. So, there are 3 variables and you have 3 equations and you can solve it. So, you… And, it is ok; hence, I solve it. But, nobody asks the question, why you do this? This lambda is called a Lagrange multiplier and this is called the Lagrange multiplier rule.

Now, can you make arbitrary rules in mathematics? That is, just say if I want to do this, just add something to something and just take gradient equal to 0, and you will get the solution. How do I know that the thing that I will get from here can be considered as a candidate for solution? But, unfortunately, in our calculus classes, we never question this. At least in India, I have never seen people questioning this fact that, how do I know if I do that, I will actually get a candidate, which can be even suspected of being a solution to this problem, to that problem. This has also got something to do with our mathematical education and the way we take science, because here we are really discouraged to ask question; it is from my own childhood experience also; my own student experience. You are discouraged to ask question; you are just a listener, who listens to whatever is been told and you have to be producing in the exam whatever is been told to you.

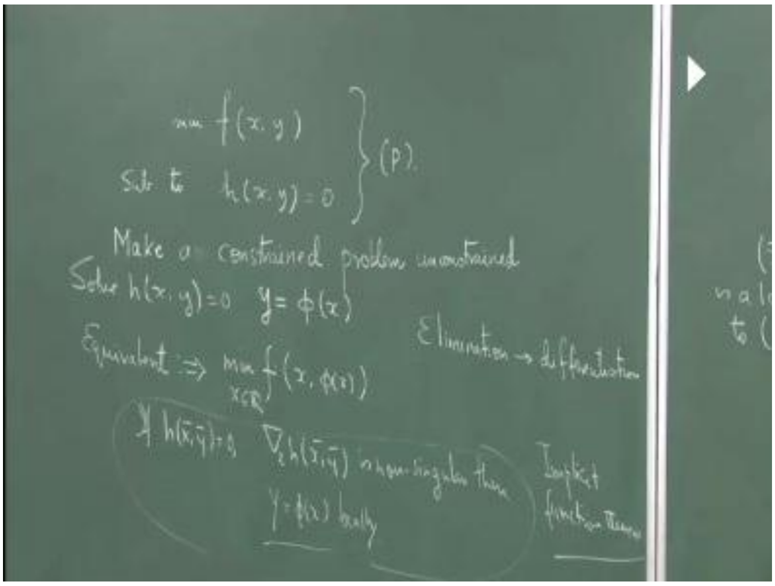
But, any sensible person with a sense in mathematics would ask that, I know that if I have a function from R to R or even say R n to R, does not matter; and if you say that, x star is a local minimum; (( )) minimum, local minimum is our… and it gives you the local minimum; then it implies that, so any local minimum must satisfy this equation. So, you know that, whenever I… If I have a point x star, which satisfies this, this could possibly be a candidate for local minimum. So, whatever Lagrange has done like this – in this way and we just mug up these things, we possibly do not ask ourselves a question. So, if x bar and y bar is a local minimum to this problem P say; what is the necessary condition? What sort of condition does a local minimum to P satisfy? Lagrange did not provide a proof that, if you do by this, that any local minimum would actual satisfy a thing like this. If you are using to find a local minimum, at least a candidate point whom you can suspect as a local minimum; then what you should do is that, you should be able to find a lambda in R such that this is true. Find a lambda in R. So, your job would be to find lambda in R such that this is true. Now, Lagrange did not prove this; it was proved much later in the twentieth century.

Now, that is the whole point people do not ask; people do not realize that, this is the necessary condition. If x bar, y bar is a local minimum, then under certain condition… For example, here I would have grad of h x, y not equal to 0; if that happens, then what I have written here is true. Then x bar, y bar would necessarily follow this set of equations; and hence, if you calculate x bar, y bar from here, you can suspect it to be a candidate for solution. But, here what we do in our calculus course; we said whatever you get, if x bar, y bar you get; basically, you will get the lambda also as a y product, which you think is just somehow it has come unnecessary a will. But, then if you just get the x bar, y bar; you say this is the solution.

We still do not realize that, optimization is not such a simple thing that (( )) you take to get something from the necessary condition and say that this is the solution; it is only a candidate for solution; you can suspect it to be a solution. But, you really have to guarantee that it is a minimizer – local or global, whatever.

So, if you want to guarantee, you need some other conditions, whose satisfaction would lead to the fact that, that x bar, y bar is indeed that critical point; which satisfies this is indeed a minimum. And, that is the second order condition on which we will touch a little bit at the end, but not at the end as we go along for the study of the Fritz John and Karush-Kuhn-Tucker conditions. Now, this rule is called the Lagrange multiplier rule. But, this actually a necessary condition; under certain condition that, if you take this fact that, grad of h x bar, y bar is not a zero vector, then this condition will hold true. So, here suppose I can write y is equal to phi x, then what I am able to do, I am able to do this; I am eliminating y and replacing it by x as function of x; and then I am minimizing over x. So, I am eliminating and differentiating.

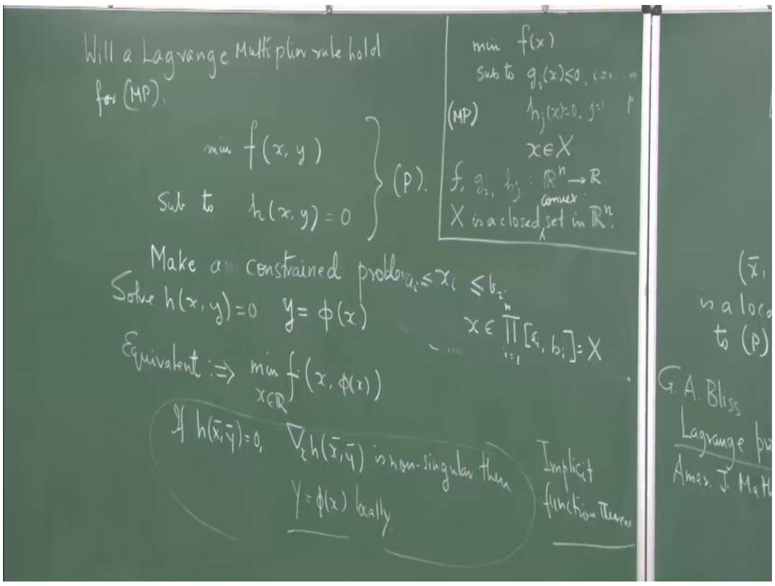
(Refer Slide Time: 19:55)



The standard procedure is elimination followed by differentiation. For what Lagrange said was something different. He said first differentiation and then elimination; differentiation followed by elimination. So, it is a reverse procedure. But, if I can always eliminate and then differentiate, I will prefer this. I do not want to get into all these. But, the problem is that, you cannot always get this explicit expression. So, you cannot always have the elimination. But, you can get an explicit expression locally; and that comes by using the implicit function theorem. So, what you can get is, if h x, y equal to 0 and the Jacobian matrix is nonsingular, then y is phi of x locally. Locally means there are certain points near x bar, y bar. So, if x bar, y bar is a solution to this; so around there will be a certain points around x bar, y bar suppose this (( )) So, if this is equal to 0, then there will be certain points around x bar, y bar – a solution. If this is satisfied; so in a small neighborhood of x bar, y bar, for all pair of points x, y, there will be this explicit relation. Of course, y bar would also be phi x bar. So, this idea is called the implicit function theorem. We will not get into the details of implicit function thermo nor we will apply it to the proof in a different way.

Lagrange applied this method successfully to huge amount of problems in the calculus of variations. So, there is a standard set of problems in the calculus of variations called the Lagrange problem in calculus of variations. And, one of the most interesting papers in this direction is due to GA bliss. It is called the Lagrange problem in the calculus of variations. And, it appeared in the American journal of mathematics in 1930 – American journal of math, 1932 I guess, if I have not mistaken. It is a very old paper, but it shows to what extent and how successfully in the calculus of variations, Lagrange has applied this method; and that that is how it became famous as a Lagrange's multiplier rule.

(Refer Slide Time: 24:22)

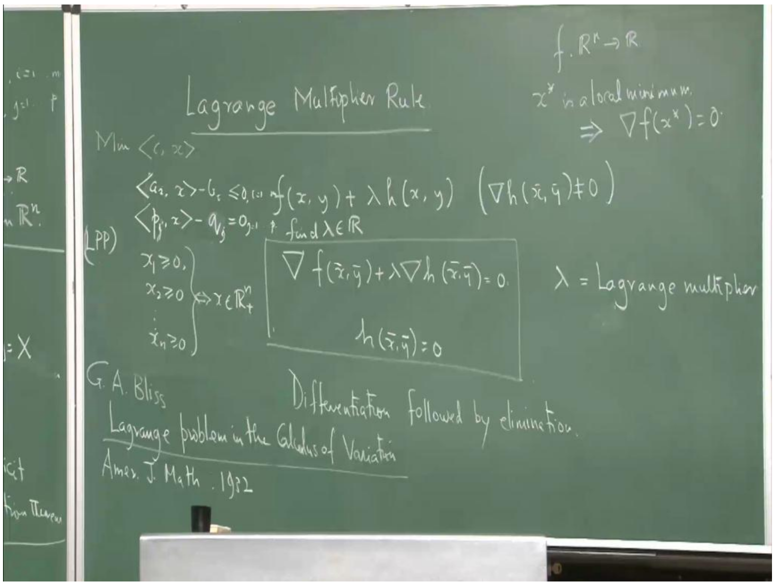


Now, let us come back to the twentieth century. And, in the twentieth century, inequality started entering the picture of constraints. They were not equalities only. And, that inequalities became the hallmark of many engineering applications. This sort of entrance of inequality started with the linear programming problem, which many of you might even have studied, but not to extend that. We would like to, but… So, a general form of optimization problem became of this form – minimize f x subject to g i x less than equal to 0. Please remember, when I have spoken about this implicit function theorem, we are all expecting very nice behavior from these guys; means I am accepting this to be continuously differentiable and all those things – 1 to p and x element of capital X. So, f, g i, and h j – each of them are functions from R n to R. And, X is a closed set in R n.

Now, this is of course, the objective function, which you have already known. This is the inequality constraint; there are m of them. This is the equality constraint; and this is the additional called the abstract constraint. For example, in many cases, you will be given the scenario, where x would be lying between… Say x i would be lying between a i and b i. So, x i would be lying between b i and a i – the closed interval for every i. So, the whole x is now lying in the cartesian product of these intervals. And, this set is nothing but the x in most cases – many many cases, but not in every cases. So, this is the general form of the mathematical programming problem or the finite dimension optimization problem, which we will henceforth refer as MP. The question is – will a Lagrange multiplier type rule or approach hold for MP? So, that is a question, which was faced by the people interested in optimization in the early second half of the twentieth century, because a utility of linear programming was becoming very clear during the OR; and hence, linear programming means these all functions are linear or affine. We will discuss it in detail later on as what sort of problems can be modeled into this general form.

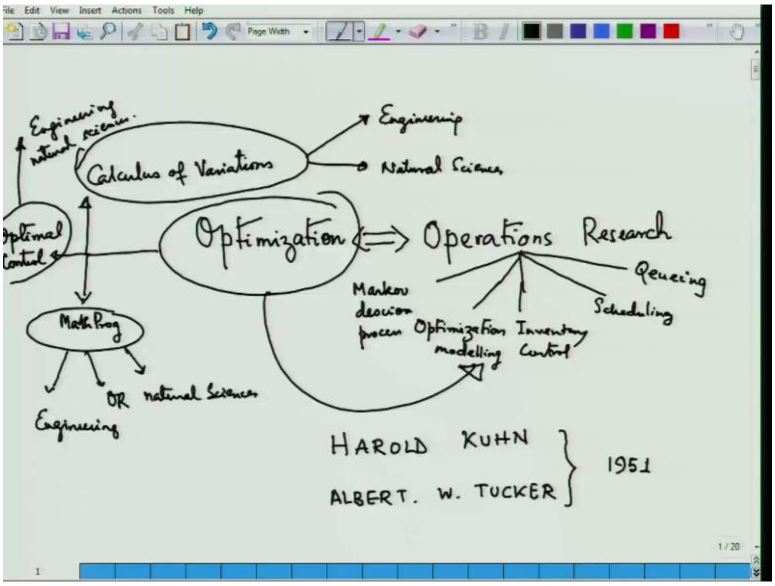
Now, a question was this – will a Lagrange multiplier rule hold for MP? And if, under what conditions and what are the issues? If I assume that all of them are continuous, it is differentiable. Will such a rule hold? And how can I use the multiplier? How can I handle this and all those things? Now, just having a closed set is not always a feasible one. What we will have is a closed and convex set. We have just spoken about convex set, but we have not gone into much detail. So, we will have a closed and convex set, because we can observe here we have taken a convex set. Now, we have simplified the problem slightly and said, if this is the scenario… Of course, you can take x equal to R n also. What I am going to get? What sort of Lagrange multiplier rule I am going to get? And, here there is a interesting piece of history of non-linear programming or just optimization theory, which we need to tell you that, that would be interesting to you. Going back to the history, these all records in the 1950s and late 40s linear programming, where there is a same problem, but with x of course, of this sort of form and maybe x is just greater than equal to 0; X is R n plus; capital X is R n plus. And, if f is a linear function, g i's are affine functions, then h i's are affine functions and that is all; and X is in R n, because that was a sort of thing that was running in linear programming, because…

(Refer Slide Time: 30:11)



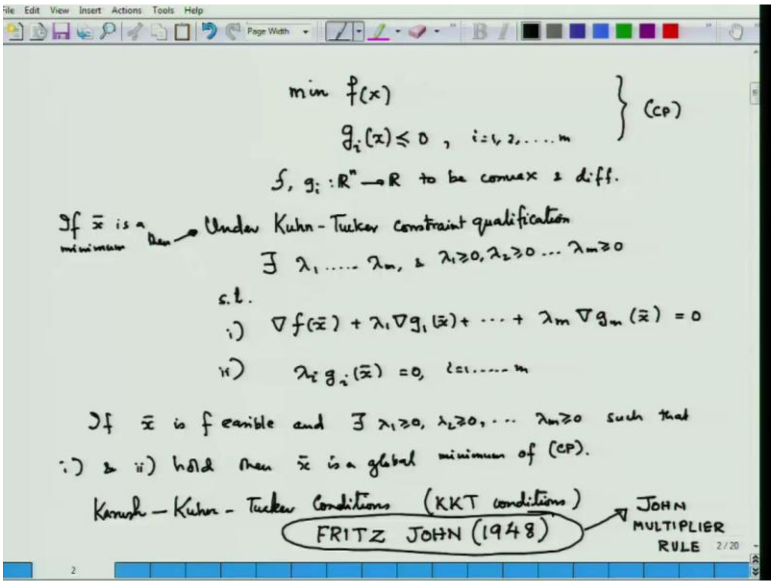
Let me just write down a standard linear programming problem; that is, say you want to minimize a linear function; any linear function is given as an inner product in a finite dimensional space. That is a (( )) theorem and you can just prove it out of (( )) Very very simple to prove this. And, once you know this, now, you have to have constraints. Once this is a linear function subject to certain inequality, equality constraint; so I can have a i x minus b I, which are affine functions less than equal to 0; i is from 1 to n. I can have say vectors d i x minus c i is equal to 0; not c i. I will just change the notation – c i's are already here – p j minus q j; j is equal to 1 to p. And, you have all the x 1, x 2, x n. So, all of these – whole thing can be collectively written as x belongs to R n plus. So, the problem is to minimize c x over these two constraints and x is in R n plus. And, that is called the linear programming problem – LPP. (( )) problem of that form. In 1951, there was a seminal paper published in the proceedings of the symposium of mathematical statistics that which was held at the University of California in Berkeley.

(Refer Slide Time: 32:30)



And, there are the authors, who are Harold Kuhn and Albert W. Tucker. So, both were from Princeton; Harold Kuhn been the student; and Albert W. Tucker been the Ph.D. supervisor; and Albert W. Tucker was the Ph.D. supervisor also of the now famous John Forbes Nash. And of course, those who have seen Beautiful Mind possibly would not realize that they have shown both Tucker and also Kuhn, because a very close friend of Nash was been shown in the movie and he was Kuhn; and Albert W. Tucker was the teacher, the person who started talking in the very first scene. He represented Albert W. Tucker. So, they (( )) read this (( )) who published this paper in 1951, where they published some sort of a Lagrangian multiplier rule for this class of problems, where everything was continuously differentiable. But, what they did not do let us see.

(Refer Slide Time: 34:08)



So, the problem they considered that, was a problem, which was devoid of equality constraints. So, they said, we will not bother about the equality constraint, Lagrange has already bothered about it; we will just bother about inequality constraints. And, took all these functions f and g i from R n to R to be convex. And, they proved that, under certain condition, which is called the Kuhn-Tucker constraint qualification. Under certain conditions on these constraints… So, under Kuhn-Tucker constraint qualification, they showed that, there exists… These are all differentiable functions: convex and differentiable; that there exists lambda 1, lambda m with lambda 1 greater than equal to 0, lambda 2 greater than equal to 0, and lambda m greater than equal to 0 such that… So, just relook at it. Of course, it is characterizing a minimum. So, if x bar is a minimum… Because it is a convex problem, there will be no local minimum; which we will talk about later on.

If x bar is a minimum, then under Kuhn-Tucker constraint qualification, there exists this such that… You see there is a sign on these constraints; it does not free like the lambda here just in lambda in R here (( )) only one constraint just for sake of explanation. So, the number of multipliers is same as the number constraints. And, you see here the multipliers are assigned. Number 2 – a very important notion called the complementary slackness condition. They said this will happen if the functions are convex. But, they also said, if x bar is feasible… So, if it is convex programming problem, we will call it CP. So, if x bar is feasible and there exists lambda 1 greater than 0, lambda 2… such that 1 and 2 hold; then x bar is a global minimum of CP. So, x bar is a global… So, that is what they proved. So, this became famous. This whole story, this whole result became famous as the Kuhn-Tucker (( )) Now, the important thing is that, this result, which is true for convex case.

Now, you see this Kuhn-Tucker constraint qualification is useful; we are not telling what is that. When you are proving that, if x bar is a solution of this problem, then this will hold; it is only for the necessary part, it is required. But, for the sufficient part, then you do not need to bother about this; you need to know that, whether this is just having a solution and whether you can solve this system of equations. This result became famous as Kuhn-Tucker conditions. And, the optimizers were very happy, because somebody finally, gave them analog of the Lagrangian multiplier rule when equality would come, inequality would come. And, for a convex case, it was both necessary and sufficient; they actually were going from linear to convex, convex differentiable and so on.

Now, they had proved for the linear case, without any additional constraint qualification mean addition assumption on the constraints, you could prove a rule like this; of course, only with the inequality case, not the equality case. Now, it was later on found that, Karush, who was in a group of the Chicago group of optimization lead by Magnus R. Hestenes, he had in his masters… This is W. Karush had written about this sort of multiplier rules, but in a very different context for a very special problem. So, nowadays it is known as… From 80s, I think it is known as the Karush-Kuhn-Tucker conditions or KKT conditions. But, the real story of the Lagrangian multiplier rule for this problem, where we both have inequalities and equalities was not told by Kuhn-Tucker.

Unfortunately, most books in optimization do not try unless written by a very good optimizers to highlight the role of one of the most unsung heroes of optimization theory; and his name is Fritz John. He was a very specialist in partial differential equation and did this single contribution to optimization theory, which really tells you about the general nature of a Lagrange multiplier type rule for this problem for having equality and inequality. And, he published his findings, his result way back in 1948 and nobody knew it when this Kuhn-Tucker conditions came. Kuhn-Tucker condition is just an easy corollary of what Fritz John had done; it is much more general. He first sent this paper to Dube journal of mathematics and they rejected it. Later on, he published in some obscure proceedings. But he for example, is unfortunately the unsung hero of mathematical programming and optimization theory.

People just do not talk about the conditions, which John has given. But, here in this talk, we want to give this person is due and we would start discussing, what is the John multiplier rule. So, the KKT condition is as we will show an easy corollary of this – the John multiplier rule. Essentially, the John multiplier rule is what we need and nothing much more. So… And, that is actually is the Lagrange multiplier rule for this sort of problems consisting of both equalities and inequalities. So, with this, let me stay. And, in the next class, we will first take just this problem without convexity thing and take the John's approach and see what happens. And then we will add the equality and then take John's approach and see what happens.

Thank you very much for your attention.