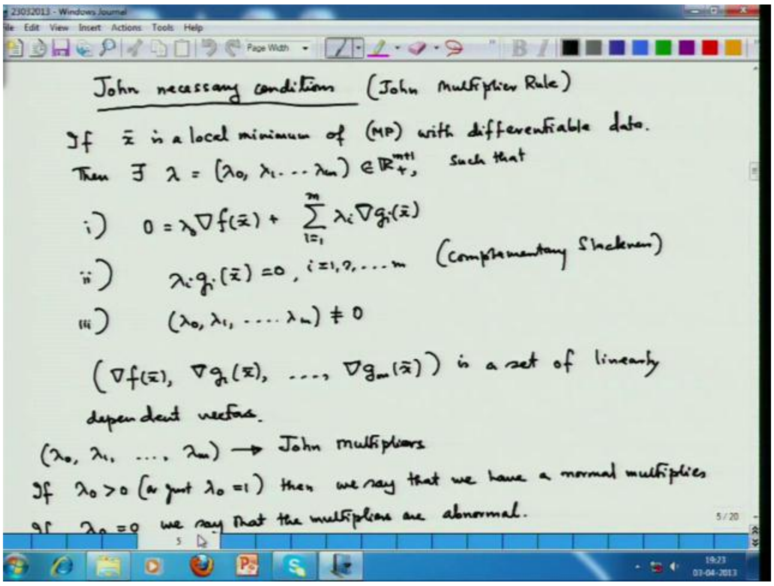
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

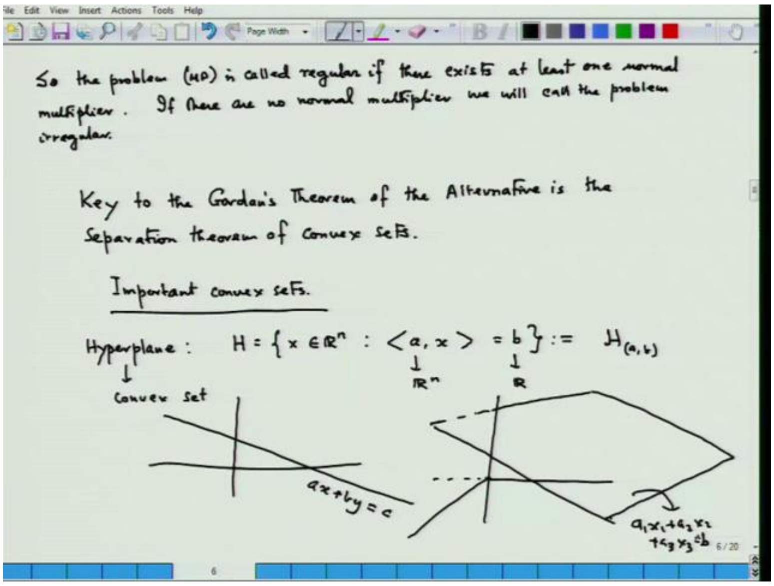
Lecture - 18

(Refer Slide Time: 00:18)



In the last lecture, we deduced essentially largely on the board fritz john or the john multiplier conditions. Here you can see the application of the Gordon's Theorem of the alternative, which leads to the john multiplier rule or the fritz john's multiplier rule whatever you want to say. Of course, there are lot of things has to be said about this normal multipliers, abnormal multipliers, examples must come. But we are going to first give you a brief outline about how this proof came about in the sense that the crux of the proof is the application of the Gordon's Theorem of the alternative. And it is very important at this stage to know how do you actually prove the Gordon's Theorem of the alternative? So, here I guess we did not use the Gordon's Theorem, write down the Gordon's Theorem of the alternative.

(Refer Slide Time: 01:14)



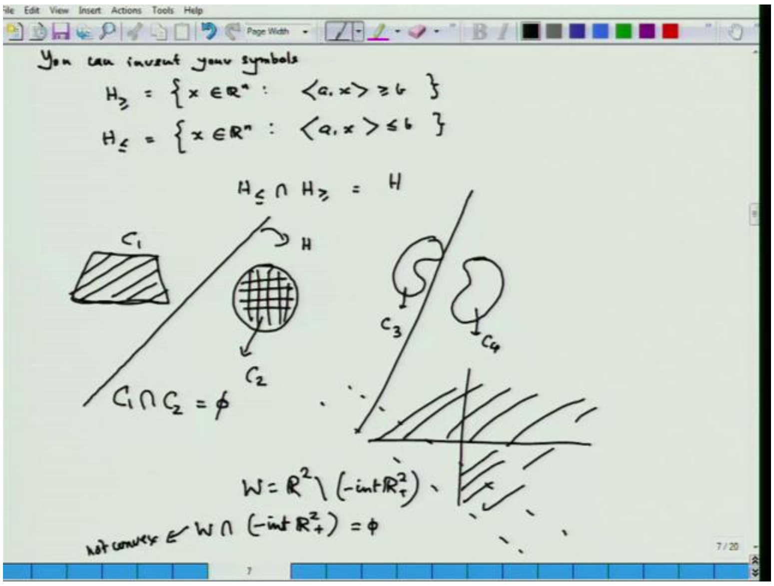
So, let me tell you, key to the Gordon's Theorem of the alternative is the separation theorem of convex sets. I have a series of lectures in the same imperial set of series of lectures series of 40 lectures on convex optimization, and I think the 7th or 8th lecture would contain this separation theorem for convex sets where the thing has been in done absolute details.

But here, because we have lot of other pressing details lot of other things to do, and so, here instead of getting too much into the issues of convex analysis, we would rather give a brief outline of what is the separation theorem and that would be enough for you to get an idea about how these things are used. And let us look into the issue of separation theorems; that brings us to some important convex sets. So, these are already there in the other lecture, but, I am just giving recalling the other. So, the first important convex set is the hyper plane. So, Hyper plane is the set of all H, set of all x in R power n, it satisfies an equality of the form this so, this a and b, a which is R power n, and b which is in R so, these two determine the hyper plane. You can in fact write hyper H (a, b) many authors do so; so this is a definition of a hyper plane, this hyper plane is a convex set. Typical examples of hyper plane in two dimensional case is a straight line, this is a hyper plane of course, you can understand these straight lines are written as ax plus by equal to c.

So, it is inner product of ab with xy equal to c, and another in R3 typical example is the plane where you usually the equation is written as a1 x 1 plus a2x 2 plus a3x3 equal to b.

This is the equation of the plane so, that is the idea now this hyper plane or straight line you can see is dividing the plane R 2 into two parts, this is one half space and this is another half space, this is the upper part, this is the lower part. So, this leads to two more convex sets associated with the hyper plane.

(Refer Slide Time: 05:08)

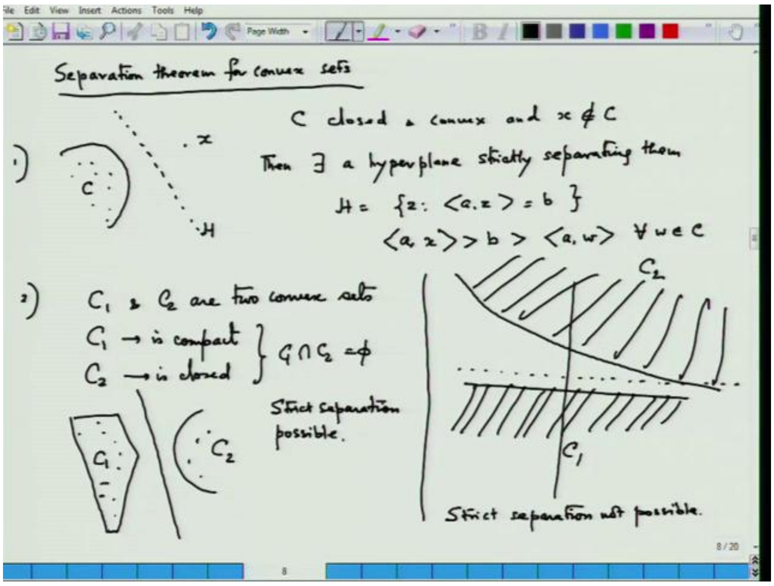


One is called upper half space which is you can write like this, you can invent your symbols. So, once you know this it is a set of all x element of R power n such that, this is usually referred to as the upper half plane, then there is a lower half plane so, these are all half planes. Find all the set of all x which this so, this which satisfies these two. So of course, the only intersection point between upper and lower half space is the hyper plane.

So now, of course, you could put strict inequality here to get the interior of the hyper planes, they are called the strict upper half plane, and the strict upper half space and the strict lower half space. The idea behind the separation is very interesting, in the sense it says if you take two convex sets which are not intersecting, one is C1 and another is C2 and what is given to you is C1 intersection C2 is phi. Then you can always find an hyper plane which can be drawn in such a manner such that, the set C1 is in one half space and the set C2 is in another half space. This is always true when you have two convex sets which do not intersect, but, this fact cannot be said about none a convex sets which is of course, you can have a non convex set like this you can have another non convex set like this which are not intersecting C3, C4. Of course, you can say oh I can draw a straight line like this of course, that is all right you can do it, but, you cannot do it for every such situation where you can do It for every such situation in a convex case.

Here for example, if you take this set, a set which is very important in multi-objective optimization. If there are more than one constraints, so this is a set W which is R power 2 set minus int R power 2 plus. So, leaving this part if the interior of this third quadrant everything else is considered in W, now W intersection minus int R2 plus is phi. Now you cannot draw any hyper plane where minus interior R2 is in one side and this whole thing is in the other side. So, here the things are broken, but, this is a convex set while this is a non-convex set this is non convex not convex and this is convex.

(Refer Slide Time: 09:06)



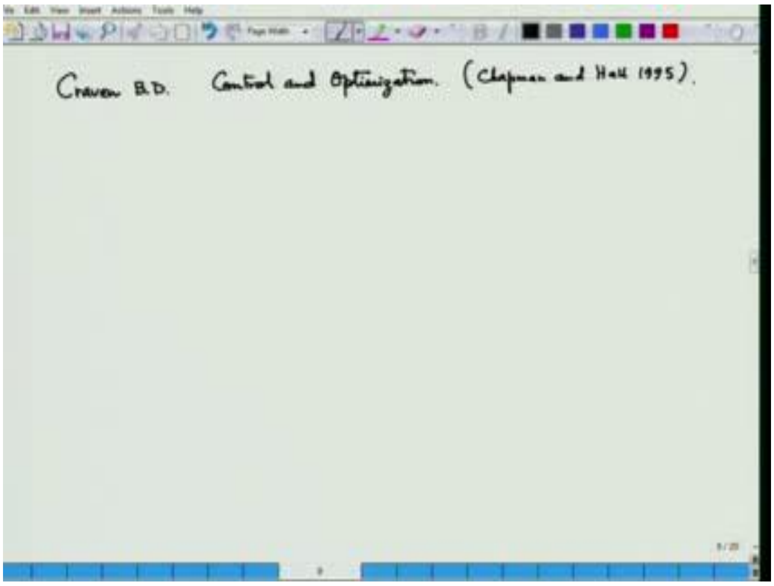
So, what are the essential separation theorems? So, essential separation theorems are as follows. First case is, you take a close convex set C which need not be bounded and a point x outside it C closed and convex and x not in C, then the conclusion is then there exists a hyper plane strictly separating them. Let me tell you what does it mean; means you can draw hyper plane whose in whose strict half spaces, C lies in and whose other strict half space x lies. That is none of them are on the boundary this hyper plane does not contain any point from any of these two sets that is what is the meaning of strict separation.

So this is one thing now this is the first and the basic result. And the second result is C 1 and C 2 are two convex sets. C 1 is compact, C 2 is closed. This is C 1 and this is C 2 with the fact that C 1 and C 2 do not intersect. Now, compactness here becomes a very essential thing because, now we say that if this happens we can do strict separations sections strict separation possible you can say that strict separation is possible. What does it mean so let there be an hyper plane which is strictly separating so h is hyper plane set of all x such that a x of course, it is not a is not zero. It means that whole so if it is 0 and a is 0 then b is equal to 0 b must be 0, then basically any x will do but, that is that is not the thing a is not 0. Now, if we look at this, what does it mean maybe I should write it more I should not write x because I have taken x here so maybe I will write z.

So, let this define this is this hyper plane so what does it mean it is strictly separating. So x is in the upper half space here so a of x strictly bigger than b while a of w is strictly less than b for all w element of c that is the meaning of strict separation. You see here we have taken c one to be compact if you take just c one to be a single point then it is also a compact. So, whatever we know about this result can be applied to this actually one can prove this and then apply it to prove this we will not do any proof here we are just giving an outline.

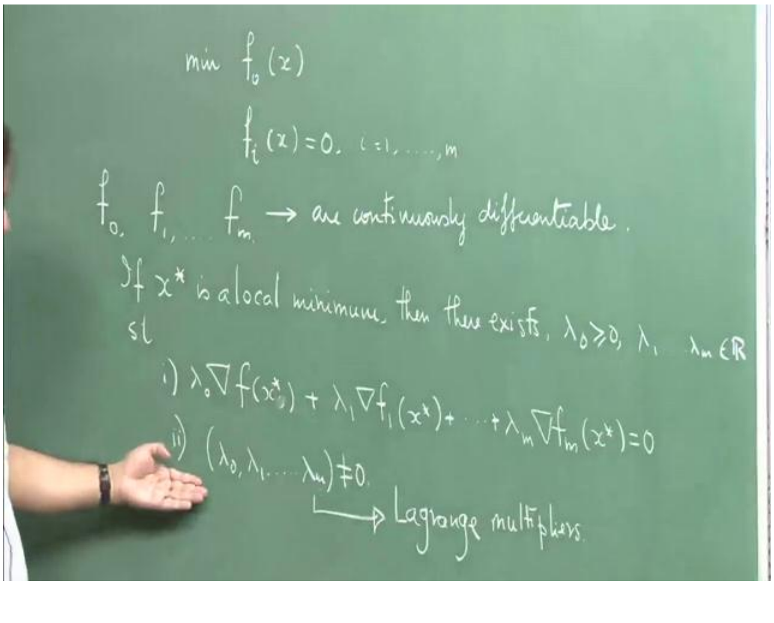
See here the compactness of the set c one is very, very important because if you do not have compactness then we cannot guarantee a strict separation. So, consider these two one is the lower half plane which I call c one e to the power minus x right and then take the upper part. Now, both sets are unbounded both are closed now you cannot find a strictly separating hyper plane if I draw this line for example, there will be a time when it will come and cross this. So, you cannot find a strict separation separating hyper planes strict separation not possible. Now, once this is known you might ask me then prove the Gordon's alternative Theorem, but will not prove the Gordon's alternative Theorem.

(Refer Slide Time: 15:09)



But refer you to this small and nice book by B. D. Craven is called control and optimization. It is published by Chapman and hall, and it is way back in 1995. It is a very nice proof of the Gordon's Theorem of the alternative, which they call the basic alternative theorem. Now the question would be going back to the john conditions again. So, we will now go back to a simpler mode a more traditional form of an optimization problem where I will just talk about inequality constraints.

(Refer Slide Time: 16:18)

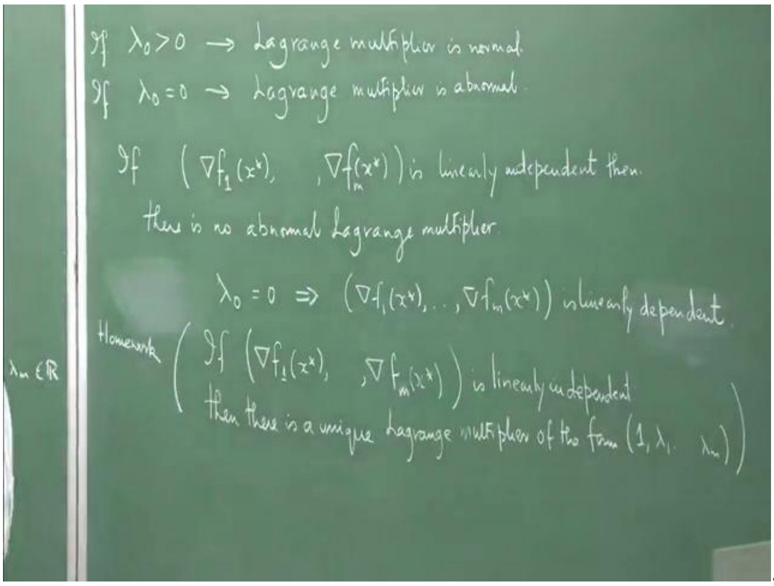


In this case, we will get what is called the famous Lagrange multiplier rule and now let us assume in this case when because we need certain thing called implicit function theorem to derive a multiplier rule f naught, f 1, f 2, fm are continuously differentiable. So, we are now looking at a much more different sort of problem instead of in equalities, let us go back and look at the traditional issue of in equality constraints.

First study some examples from in equality constraints, and then try to get it with in equalities. You have one major point is that, they have the issue of complimentary slackness condition, and to satisfy in equality, and satisfying all these things are not so, easy actually. Note now that if x star is a local minimum then there exists lambda naught greater than equal to 0, lambda 1, lambda 2, lambda m, element of R means they are all free such that so, here it is quite a simple rule. The most crucial point as I again mention is, this fact, this cannot be 0, whole vector cannot be 0. So, this instead of calling it as a john multiplier, we will call it as a Lagrange multiplier. So, when you just have equality constraints these are usually referred to as Lagrange multipliers. You have read about them in your calculus course, but of course, you did not know that you really had to show the existence of such lambdas now.

Now, what Lagrange multipliers are doing is that it converts the whole constraint problem into an unconstraint problem and then you are just checking the differentiation of the unconstraint problem. This is the basic philosophy of solving any constraint optimization problem, you really have to convert the constraint problem into an unconstraint problem; and then solve it just like an unconstraint one that is without constraints. Now it is enough to know that this it is enough to know just this fact.

(Refer Slide Time: 20:51)

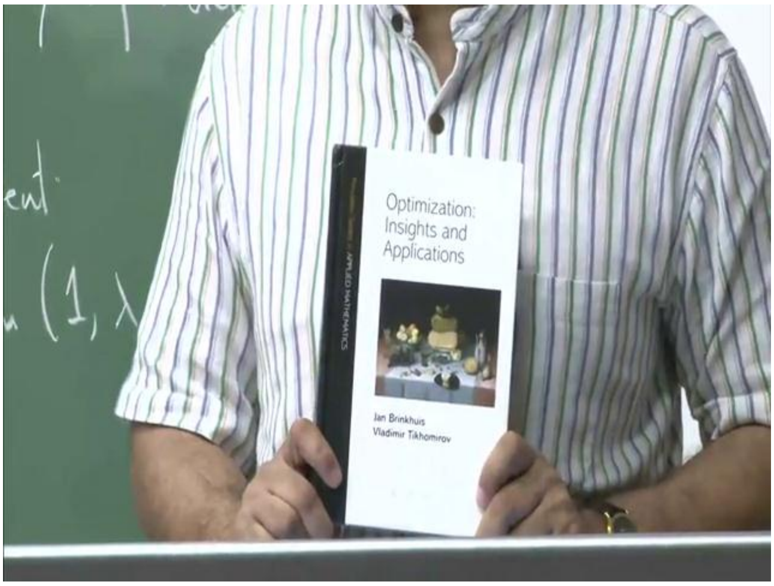


.

Similar way, if lambda naught is greater than zero than Lagrange multiplier is normal, if not we will call the Lagrange multiplier abnormal. Now, it is also important to remember when we can have a situation when there will be no abnormal multipliers. So, if this vector is linearly independent then there is no abnormal multiplier. It is very important to note here the following fact, that what is the definition of an abnormal multiplier, but lambda naught is 0, if lambda naught is 0, then from here it means one among lambda, one lambda m 1 among these at least has to be non-zero so, these vectors are linearly dependent. So, lambda naught equal to 0 would immediately imply linear dependence so, lambda naught equal to 0 implies in linearly dependent.

Now, there is a interesting question so it means that if it is linearly independent and lambda would is not equal to 0. But see, if it is linearly independent, I will never get a abnormal multiplier the interesting question is suppose, I have linear dependence that what does it mean? Does it mean that it will have a abnormal multiplier? The answer is actually no. It need not have an abnormal multiplier even if you have linear dependence. we will use examples to demonstrate that both a normal, and an abnormal multiplier can be present we know that given an x star these multipliers are not unique unless you guarantee linear independence if there is linear independence then it is unique right. If there is no so, if I leave that as a home work to you of course, if this is linearly independent then m must be less than n which is a very basic fact about linear independence that maximum number of linear independent vectors in R power n is of course n. Of course these functions f naught, fi these are all functions from R power n to R. If this m is linearly independent then there is a unique Lagrange multiplier. This is a home work for you unique Lagrange, unique Lagrange multiplier of this form. Because, once lambda is not zero I can divide it by this and I can get a new set of multipliers.

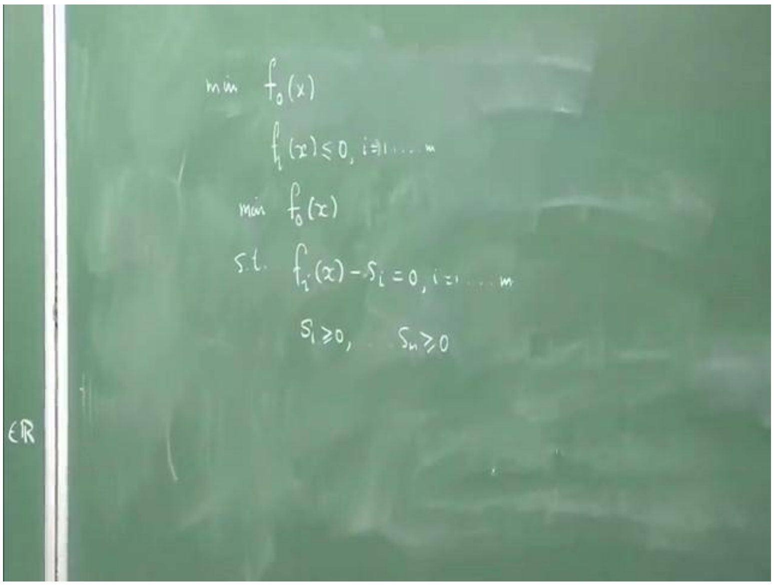
(Refer Slide Time: 27:52)



Now, the question is this is quite very elegant simple did not have to write much no inequality constraints suppose I have the inequality constraint issue the problem can I use this idea to actually develop the Fritz John multiplier rule for inequality constant. Can I apply the Lagrange principle? So, what I would like to show here just like a Jan Brinkhuis and the great optimization theorist Vladimir Tikhomirov has shown in this book optimization insights and applications I have already named this book, but, I really want to show you this book. It is a Princeton university publication of 2005 and a very very beautiful book it gives a very beautiful insight into optimization, very deep insight I would say.

And what I would also like to note that in this book they show that it is the Lagrange principle, which is the guiding principle of all of optimization this Lagrange principle this one and that and lot of problems can be actually solved by using Lagrange principle. Let us now try to develop the inequality constraint john multiplier rule let us make a trial through this Lagrange.

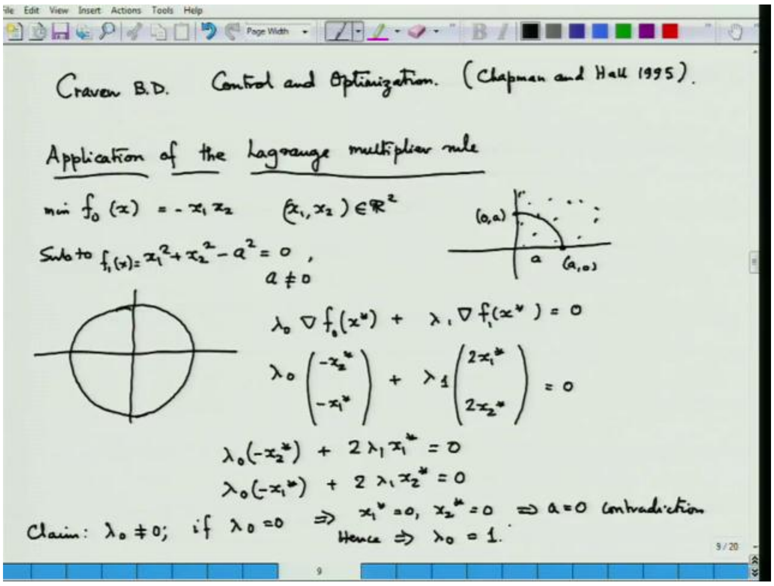
(Refer Slide Time: 28:34)



So, now if I am making it less than zero, once you do that remember I can always introduce what are called slag variables, that is I can always introduce non negative variables right which will make them non zero. But the biggest problem that you will lie is that when you want to try to apply the Lagrange principle on this. Then what would happen is the following; if you want to write this is as a inequality constraint problem, you have to write this as but S 1 is greater than equal to 0, (( )) so, I have added m variables and added m constraints also. Now, these are again inequality constrains. So, the question is that I cannot try and write down if I want to use the Lagrange principle and try to write down the john multiplier rule it is really not possible. So, here we need to have a lot of geometry which is essentially related to sets given by inequality constraints.

So, that is why Fritz Johns derivation of the whole thing was so so, important it was a 1948 paper rejected by Duke, published in a memoir volume celebrating the 16th birth day of Richard Curand of who is a famous mathematician. In fact John is a famous mathematician known in partial differential equations, now this again needs some more convex geometry a more geometry. Because, again inequality constraints even if you change the functional constraints to equality, inequality constraints are actually reappear in the form of the slack variables, which must maintain non negativity to actually be a slack variable so, this procedure cannot be done. So, now what is important is that we will take some examples and try to use the Lagrange principle we will take these examples from Brinkhuis and Tikhomirov so and try to apply and see what we can get.

(Refer Slide Time: 31:18)



So, let us look at the first problem minimize f naught(x), minimize minus x 1 x 2, where x 1 and x 2 belongs to R power 2 so, this is a problem in two dimensions. So, here f1, f1(x) so, it is a unit, it is a circle of radius a and you are only expecting x i strictly greater than 0. That is you want the problem to be here so, once you basically want problem to be minimized over this so, basically you are excluding you are taking everything in the non negative (( )), but, you are excluding the points (a, 0) and (0, a). Now, once you do anything you restrict your things to in an open set then, it does not matter you will get back the Lagrange multiplier.

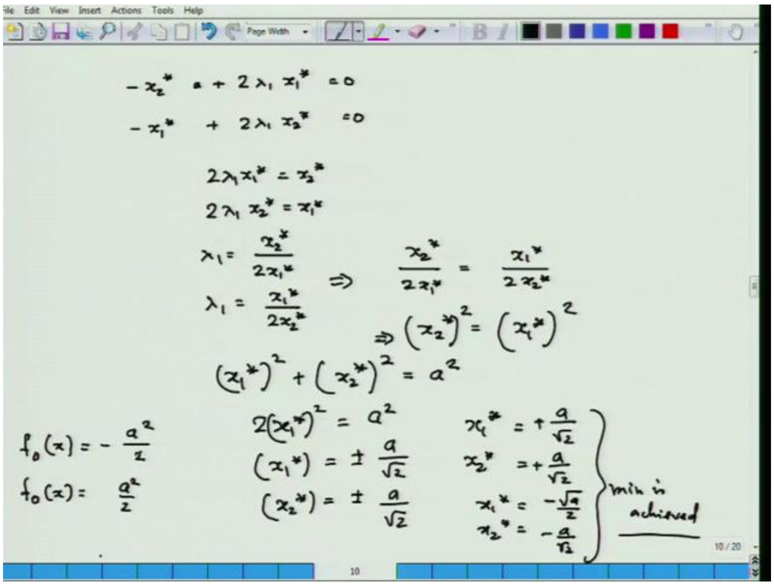
How you would get back would need a lot of more deeper things which we will not go into at this moment. Let us try to first learn, how to apply the Lagrange multiplier rule and we will see that we are actually getting in most cases the multiplier lambda naught to be one because we are getting normal Lagrange multipliers. First case is to know the existence of a solution. See if I look at this problem from this point of view, if I consider this whole set and I do not take the fact that xi has to be strictly greater than 0 if I do not bother then this set is a compact set. It is closed and bounded and over it you will have a minimum, that is over a compact, set a continuous function has a minimum now here we are not putting a is not equal to 0.

So, there will be a minimum somewhere here, on this let us forget this extra condition I do not think that you really need to bother about this extra condition at this moment just take this problem. Now the there will be a minimizer somewhere so, the multiplier rule says that there would exist a lambda naught grad of f x star plus lambda1 grad f x star, f1 x star. Now lambda naught this one would become grad of f x, f x star is here, minus x 2 star, minus x 1 star plus and lambda naught lambda1 are not equal to 0.

Assuming that there will be a we are now sure about the existence of a solution, because we are minimizing a continuous function over a compact set. So once you are sure then we are telling that let x star be that minimum and then that would follow the Lagrange multiplier rule and in fact it is a global minimum, if you take this one. So, this would now amount to the following. Lambda1 does not have signs so, I can take minus lambda1 also, it does not matter just for easiness of the calculation never mind is equal to zero. Another equation is lambda naught so, if I say if lambda naught I will now claim that lambda naught is not equal to 0.

Now, if I say if lambda naught is equal to 0 then, from the above two equations it would imply that x 1 star is equal to 0 and x 2 star is equal to 0. Because by the Lagrange multiplier rule lambda naught, lambda1 both cannot be 0 lambda1 is 0 so, lambda 1 is not equal to 0 so, this will happen. But if this happens, then it will mean s square is equal to 0, but I have taken n not equal to 0; so I cannot really take that these two equal to 0 either so, this is hence, this would imply that a equal to 0 contradiction. Hence, it implies that lambda naught is equal to 1 it is just taking lambda 0 equal to 1 so, then that what would it give me, so let me just lambda naught is 1.

(Refer Slide Time: 38:24)



So, minus x 2 star plus; 2 lambda 1 x 1 star is 0; minus x 1 star plus 2 lambda 1 x 2 star is equal to 0. Now we will eliminate the lambda one to get the results about so, from here lambda1 sorry from here we will get 2 lambda 1 x 1 star is x 2 star; 2lambda just a moment 2lambda1, the first equation will give me 2 lambda 1 x 1 star is x 2 star; and 2lambda1 x 2 star is x 1 star. So, lambda1 is x 2 star by 2 x 1 star; and lambda from here we also get from the second equation; we get lambda 1 is x 1 star by 2x 2 star so, then this would imply that x 2 star by 2 x 1 star is equal to x 1 star by 2x 2 star; and this would finally, imply that x 1 x 2 star square is equal to x 1 star square.

Now, these has to be feasible so, which means x 1 star square plus x 2 star square; is equal to this. Actually, now if I want that these are to be positive we can have positive, negative all those things now if we really want that we will restrict it to the positive part let us see what happens. Now once you have this, what do you have you have this as a square now you will have 2 x 1 square as a square.

So, x 1 star is plus minus root over a by 2. Similarly, x 2 star is equal to plus minus a by root two. Now you have to determine which will give you the minimum. Basically, if you take a positive and negative combination then you will get a negative number you take x 1 star to be a by root2; x 2 star to be minus a by root2 so, you really have to see what are the points where the minimum is achieved on the circle. So, if you want only x 1 and x 2 to be both positive then you will have to take x 1 star; see what you have to find the so f naught(x); suppose, if I take both negative then what I will get minus of a square by two.

If I take both positive then I will get minus a square by 2; if I take one positive one negative then I will get so, I take this positive and that negative so, I will get plus and minus would be minus and then there is a minus so, I will get a square by 2. So, this is giving me the maximum value of minus x 1, x 2, and this is giving me the smaller value. So, only when I have both positive or both negative so, x 1 x 2 is this or x 1 star x 2 star is; these are the points where the minimum is achieved so the two points where the minimum is achieved. Because, if you take a plus minus combination then you get a plus thing so, that is a bigger quantity than what you get and we know that a minimizer exists so, among these points these four possible combination, we will get the minimizer.

So, these are the two points where the we get the minimum value; so, these are the minimizers. So minimizers must you know satisfy this all this Lagrange conditions and minimizer exists. So, the minimum value among this four critical point objective, objective function value is minimum at which point we have to find that among these Lagrange points or points satisfying the Lagrange multiplier rule so, for minimum value is obtained where the objective function is getting is achieving the minimum value that is in this particular case. So, this is our example and we will carry out some more examples in more details in the next class.

Thank you very much.