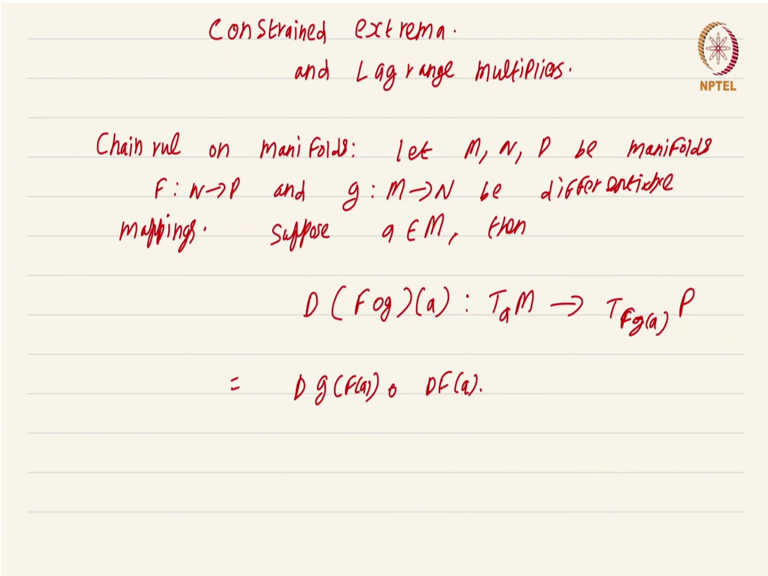
Real Analysis II
Prof. Jaikrishnan J
Department of Mathematics
Indian Institute of Technology, Palakkad

Lecture - 22.4
Constrained Extrema and Lagrange Multipliers

(Refer Slide Time: 00:23)



In real life problems that involve finding extrema, often there is a constraint being imposed. For a toy example of this, consider the amount of money that you have that is a fixed number that is a constrained and you want to maximize your profits by using this money wisely.

So, this is a toy problem. So, there is a fixed constrained the amount of money. Typically, this constraint is the resources you have and the function that you want to maximize is some sort of profit or output or something like that.

So, the previous results about extrema cannot deal with this because it detects extrema only at points, where the function is defined on an open set. So, when you have a constraint, its not clear whether the point that you find out for this function, an extrema point whether it satisfies that constraint is not clear.

So, we are interested in finding out constrained extrema and the method of Lagrange multipliers allows you to figure out at which points you could possibly have a local extrema subject to some constraint.

Now, the idea is that constraint is going to define a manifold and you try to find the extrema of this function restricted to that manifold. So, in the proof we will require a chain rule for differentiable functions on manifolds. So, let us prove that first before proving the central fact about Lagrange multipliers.

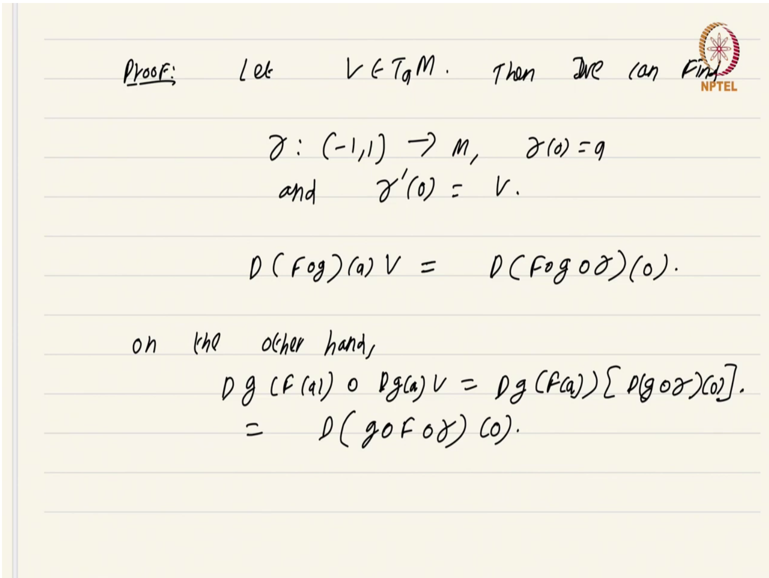
So, this is chain rule on manifolds on manifolds. So, the chain rule states the following. Let M comma N comma P be manifolds. It does not matter what the dimensions are. Let F from N to P and g from M to N be differentiable mappings, be differentiable mappings.

Suppose, the point a suppose a is a point in M, then D F composed with g at a this map is actually a map from as we all know, its a map from T a M to T F compose with g a P. This map is nothing but the map D F composed with D g at a and D F is taken at g of a ok.

So, these two maps, the derivative of the composition is nothing but the composition of the derivative. So, let me write this out in a clear way. So, this is D g at the point F of a composed with D F at a. So, we have an elegant analog of the chain rule for manifolds.

Now, the proof of this requires two exercises that I have given in the notes relating the tangent space to curves. So, I am going to take it for granted that you have solved these exercises.

(Refer Slide Time: 04:00)



So, let us give a proof assuming that exercise. So, what this says is let V be a point in the tangent space of M, then those exercise sets would have told you that we can find a curve, then we can find; we can find gamma from minus 1 1 to M, gamma of 0 is a and gamma prime of 0 is V.

So, essentially, this says that the velocity vectors of curves exhaust the tangent space at a of the manifold M. So, any tangent vector for any tangent vector V, you can always find a curve whose velocity vector at 0 is exactly this vector.

So, this was one of the exercises that I have given before; if you have not solved it, solved it now. It will deepen your understanding of the tangent space. Now, the second exercise that I am going to use is the fact that this velocity vectors behave well under the derivative map.

More precisely, if I push forward this velocity vector that is if I push forward V under the map F composed with g, this is the same; this is the same as the velocity vector of the composed curve; that means, this is same as D F circle g circle gamma at 0.

So, F composed with g composed with gamma is going to give you a curve in the manifold P such that at 0, it takes the value F composed with g at a, the velocity vector of this curve is nothing but the push forward under the map under the derivative map of the tangent vector V, ok. So, this is the second exercise that I had left for you previously. If you have not solved this, please solve this also.

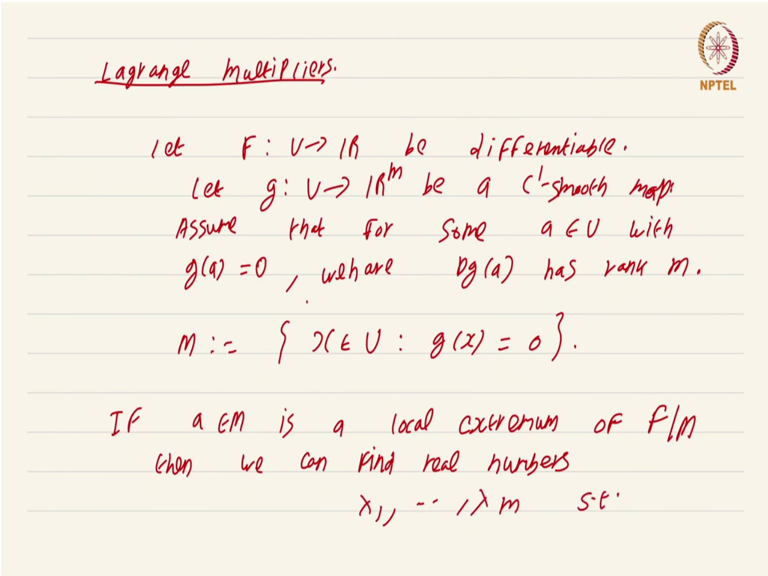
Do not take it for granted ok. Now, on the other hand, on the other hand, again applying this exercise, we know that D g at F of a compose with D g a at V, I apply this exercise that the push forward of the velocity vector is nothing but the derivative of the composition of the curve with the function. I apply it only to g now. I apply it only to g. Here, I had applied it to F compose with g, now I apply it only to g to conclude that this is nothing but D g at the point F of a acting on D g compose with gamma at 0 ok.

So, this just says that this I mean I am the push forward of the velocity vector V under the derivative map of g is nothing but the derivative of the composed curve g composed with gamma at 0 ok.

Now, in fact, this is nothing but again applying the same result again, but this time you replace V by the vector D g composed with gamma at 0 and the curve gamma by g composed by gamma and applying the same result again, this is same as D g composed with F composed with gamma at 0.

So, this time, this quantity acts as the vector, g composed with gamma acts as the curve and this g function g acts as the function to which I am push for I am using to push forward the velocity vector ok and you see notice these two are same and this is true for all velocity vectors. So, the chain rule on manifolds is proved. So, with the chain rule at hand, the Lagrange multipliers method is not that difficult to prove ok.

(Refer Slide Time: 08:10)



So, Lagrange multipliers, I will state the precise result. I am fairly certain that you have seen this and spent at least a few hours in your undergraduate basic multivariable calculus course solving constrained extrema problems. So, as I emphasized in this course, the focus is on the theory, the computational aspects are neglected and left as some simple exercises. So, there are a couple of exercises in the notes to remind you how to use this particular Lagrange multipliers method in practice. But I am not going to be focusing on that.

So, let F from U to R be differentiable. So, this is the function I want to maximize or minimize; typically, maximize. Let g from U to R be or R m, I am going to put more than one constraint be a C 1 smooth function, C 1 smooth map function; it is fine ok. So, this g function is supposed to act as the constraint.

So, what we are going to do is assume that for some a in U with g of a equal to 0, we have D g at a has rank m. So, it is a full rank. The derivative map is a full rank matrix at this point a.

So, essentially, what we are doing is we are assuming that this map is a nice map and we are going to now consider the 0 set of this map which is going to be a manifold at least near this point a. So, just consider M to be x in U such that g of x equal to 0. So, what has happened is because this map g satisfies the condition that the derivative at a is full rank, near the point a this set M, this set M will be a smooth manifold will be a C 1 smooth manifold, that is going to help us a lot.

Now, the conclusion is under the setup if a in M is a local extrema or extremum of F restricted to M, so that means, I am focusing only on those points in U that already satisfy the constraint. So, look at an extremum of F restricted M, then we can find real numbers.

(Refer Slide Time: 11:19)

$$DF(a) = \lambda_1 Dg_1(a) + \lambda_2 Dg_2(a) + \cdot + \lambda_m Dg_m(a).$$

here $g = (g_1, \cdot\cdot, g_m)$.

Proof: we can find an open set $W \subseteq \mathbb{R}^n$ s.t. $W \cap M$ can be locally parametrized by a C^1-smooth fn.

$\phi : V \to M \cap W$ where $V \subseteq \mathbb{R}^{n-m}$ is open and ϕ satisfies all the hypothesis in the third criterion for a manifold.

Take $\phi(0) = a$.

We can find real numbers lambda 1 to lambda m such that the derivative D F a is nothing but a linear combination of the derivative maps of the component functions of g plus dot dot dot plus lambda m D g m at a ok. So, here g is g 1 to g m. Of course, when you would have learned this in your multivariable calculus course, all the D's that appear in this particular equation would have been replaced by the gradient grad ok.

So, you would have learnt it in terms of gradient vectors. I am intentionally phrasing it in terms of derivative maps; both are in a sense equal and just taking dot product with the gradient is going to give you the derivative map ok. So, I am going to use the fact that this at this point a, this M is going to be a smooth manifold.

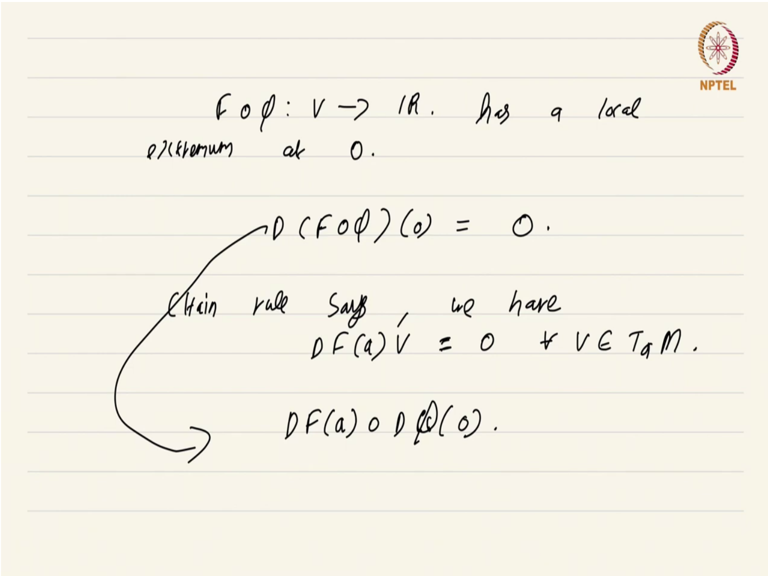
So, the condition of full rankedness is going to tell us that we can find; we can find an open set open set W in R n such that W intersect M. So, locally this M at near this point is going to be a manifold.

So, it can be parameterized, can be locally parameterized by a C 1 smooth function, C 1 smooth function which I call phi defined from V to M intersection W, where V subset of R n is open and phi satisfies. So, actually V is a subset of V is a subset of R n minus m; V is a subset of R n minus m and phi satisfies all the conditions, all the condition all the hypothesis rather; all the hypothesis in the third; in the third criterion for a manifold ok.

So, all this is saying is this is just a straightforward application. The fact that such maps exist such a map phi exists is a straightforward application of the implicit function theorem and no pun intended. The proof of this is implicitly contained in the proof of the implicit and the proof of the equivalence of the three criterion for set to be a manifold within that embedded within that there you can see a proof of this ok.

So, now that you have parameterized this piece by an open set, it seems like we can go ahead and apply the usual results about maxima minima ok.

(Refer Slide Time: 14:54)

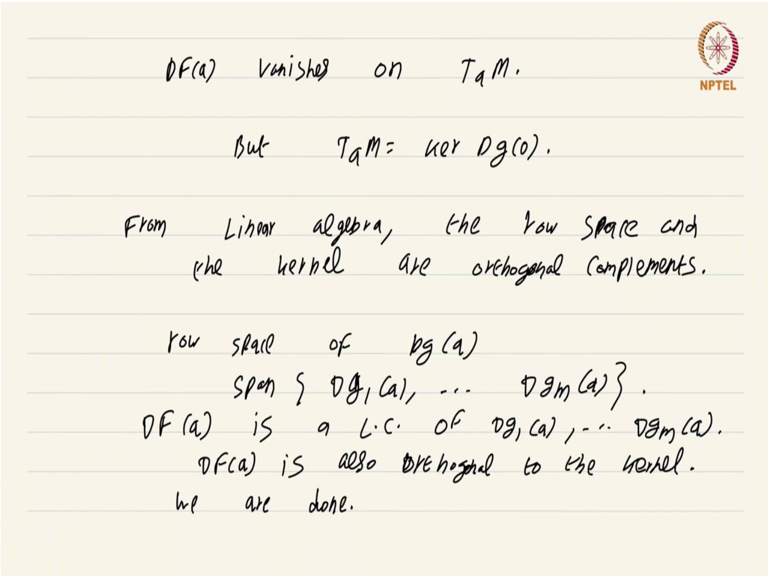


So, what you do is consider F composed with phi this starts from V and ends at R ok. Now, we are already given that phi of 0, let us just take phi of 0 for convenience to be a ok. This means 0 is in V. So, that notation will become simpler ok. So, phi of e has a local extrema local extremum at 0 by hypothesis ok; that means, the derivative of D F compose with phi at 0 must be 0 ok. Now, how does this help us?

Well, what is the chain rule say? Well, the chain rule says, chain rule says, chain rule says we have D F a acting on V is equal to 0, is equal to 0 for all V in the tangent space of a at M ok. Why is this the case? Well, think about this the chain rule will tell you, the chain rule will tell you that this D F composed with phi at 0 is nothing but D F at a composed with D phi at 0; D phi at 0. But the tangent space is nothing but the image of this map D phi.

So, D F a must have the image of D phi at 0 in the Kernel which is just another way of saying that D F a acting on V 0 for all V in T a of M ok.

(Refer Slide Time: 16:55)



So, essentially, essentially what this shows is D F a vanishes on T a M right. But there are several ways of expressing this tangent space; another way is to express it as the kernel of D g at 0 ok; the kernel of D g at 0. Now, we are going to use an elementary fact from linear algebra, linear algebra from linear algebra, the row space; the row space and the kernel and the kernel are orthogonal complements.

So, if you have a matrix, then the row space and the kernel of that matrix are just are orthogonal complements; not orthogonal matrix. Orthogonal matrix means something entirely difference a different; are orthogonal complements ok. So, elementary linear algebra will tell you that the row space and the kernel are orthogonal complements ok.

Now, what is the row space? What is the row space of D g a? Well, if you write down the matrix of D g a, you know that the row space of D g a is nothing but the span of grad g 1 at a comma dot dot dot grad g m of a ok. Now, what do we ultimately want to show in terms of gradients, we want to show that gradient of F of a gradient of F of a is a linear combination of grad g 1 a comma dot dot dot grad g m a right.

So, since the row space and the kernel are orthogonal complements and the fact that this gradient of F of a is also there in this orthogonal complement, is also orthogonal to the kernel orthogonal to the kernel. That is what we just established saying that D F a vanishes on T a M is saying that gradient of F of a dot product with any vector V in T a M is 0 which just says grad gradient of F of a is also in the ortho is orthogonal to the kernel.

But saying that gradient of F of a is there in the orthogonal complement of the kernel just means that it is there in the span of gradient of g 1 at a dot dot dot gradient of g m g m a which is just a refreshment of the conclusion that we need. So, we are done; so we are done.

So, this proof was utterly straightforward. Just uses some elementary linear algebra and the chain rule. So, I will not again I repeat; I will not emphasize the computational aspects in this course that is better done at a more elementary level. So, this concludes the derivatives portion of this course. We will move on to integration from now.

This is a course on Real Analysis and you have just watched the video on Constrained Extrema and Lagrange Multipliers.