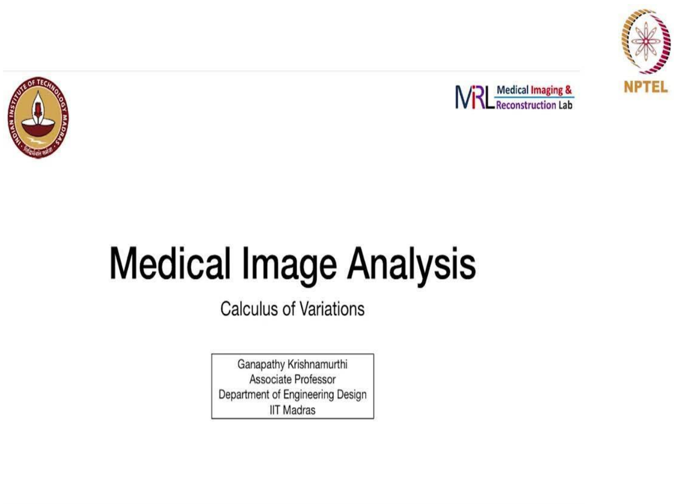
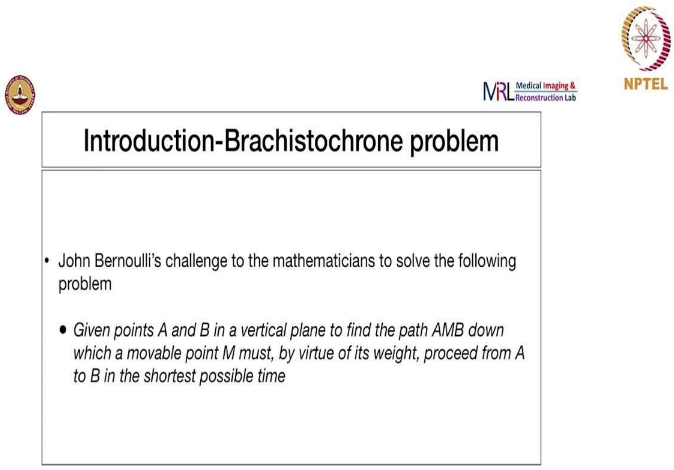
Medical Image Analysis
Professor Ganapathy Krishnamurthi
Biomedical Engineering Design
Indian Institute of Technology, Madras
Lecture 28
Calculus of Variations

(Refer Slide Time: 00:16)



Hello, and welcome back. Today's class we will talk about calculus of variations, which is a fundamental topic, which is necessary to understand partial differential equations based techniques for image segmentation. So, this is the calculus of variations is what gives rise to many of these techniques. We will have a brief look at this method just to understand where it comes from, and the fundamental Euler-Lagrange equation.

(Refer Slide Time: 00:46)

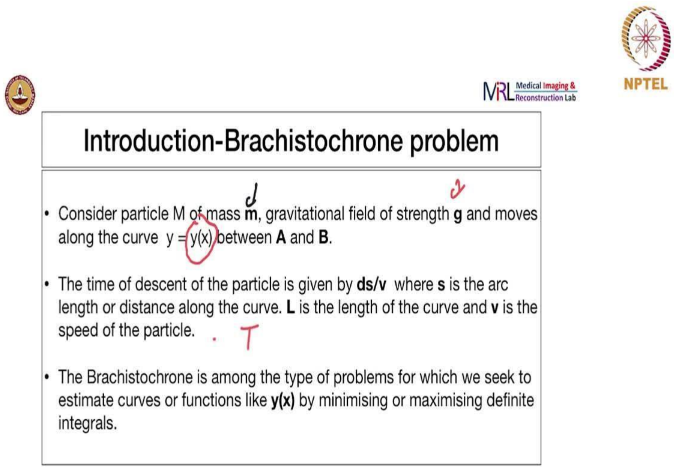


So, we will look at the historical origin of the calculus of variation. So, in the early 1600s John Bernoulli, one of the mathematicians in Europe, renowned mathematician in Europe at that time, posed the so called Brachistochrone problem. So, the problem statement is given below, I will just read that out. So, given points A and B in a vertical plane, find the path A and B down which a movable point M must by virtue of its weight, proceed from A to B in the shortest possible time.

So, basically, it is a point mass, mass m, small m, and it is going to travel in the vertical plane, we look at the images in the next slide, from A to B starting point is A, ending point is B. And you have to find the, the idea is to find the path of that particle B, we will take it as bead on a string. So find the path of the particle or a point and it is of course it is moving by virtue of it. So, it means that the only external force is the gravitational force, it is acting downwards.

And the idea is to optimize for the time. So you have to find out the shortest possible time, the path which gives you the shortest possible time. So, this is like an optimization problem, except that the solution is not just one number, or a coordinate, but rather an entire path or a curve or a function. So, we will look at the details of this problem in the next few slides.

(Refer Slide Time: 02:15)



So, here is the slightly more detailed overview of the problem. So, you consider the particle M of mass m. So, the small m is the mass (poor notation, sorry about that), but the gravitational field of strength g, so, this is the acceleration due to gravity, g and it moves along the curve, the curve is parameterized by x. So y is a curve, $y(x)$ is the equation of the curve between two points A and B. And of course, it is in a vertical plane.

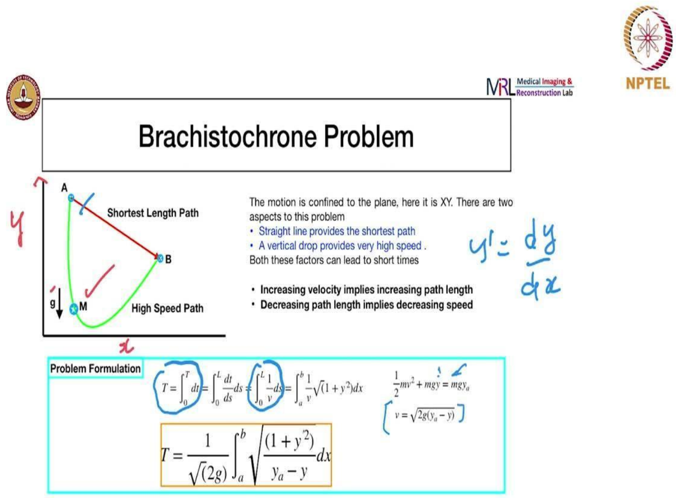
And the time of descent of the particle is given by ds/v, so ds/v is basically if the particle travels along the curved path by an infinitesimal length ds is called the arc length and the time it takes is given by ds/v, v is the speed of the point mass or the particle and ds is the length along the curve. So, it takes a curved path for the general path is the curved path.

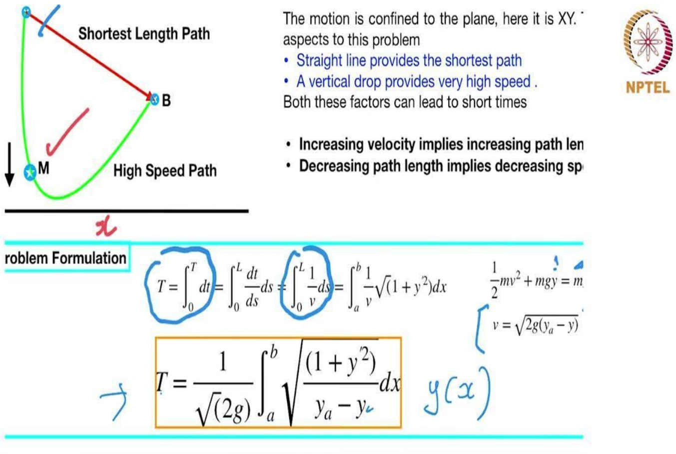
So, if it travels a distance ds around that curved path then the time taken by it ds over v where v is the speed. L is the length of the curve and v is the speed of the particle. And you can also denote the total time taken if you think of it as t that will be total capital t will be the total time taken. So, this is the problem statement.

So we are trying to optimize for time. So we have to find the path that gives the shortest time. So if you looked at optimization problems before, for instance, in neural networks, we are looking for parameters that provide the smallest loss function. In a lot of cases in optimization, we have some form of cost function and usually try the estimate parameters in an optimization problem so that the cost function or loss function is small.

Here the loss function, if you want to think of it is the time, time taken. Time taken should be just as minimum as it can get. And what parameter are we trying to estimate here, we are not trying to determine a parameter rather, we are trying to determine a entire curve or a function. So, these are the kind of problems that is concerned in the field of, that you are concerned with in the field of calculus of variations.

(Refer Slide Time: 04:39)





So, we have slightly more detail look, because I have not shown you any pictures so far just talking about these things. So without any loss of generality, we can say this is the vertical plane and think of this as the y axis and this will be x. So, and you are, you are trying to get from A to B in the shortest possible time. So, what is the path? What is the curve that will take you from path or curve that will take you from A to B.

And this M here, this is the point mass you are talking about g is the acceleration due to gravity. So, we are trying to estimate this equation of this path and we will call this $y(x)$. Now, how do you go about doing this? So, if you do it intuitively, then, you would immediately say, let us do the shortest path, the shortest path is the straight line between A and B, that is what is given here, this is the straight line. So, if you look at it, this is the straight line.

So ideally, you would think of the shortest path straight line path would give you the least time. But also think about the other way, if you let the particle drop a bit vertically quite a bit, then it will gain speed because it is motion under gravity it will gain speed. So, you are not giving it any extra or whatever, just dropping it. So, it will gain speed and so then if you think about it, in which case, the time still can be reduced because you have higher speed.

But the problem with this approach is that if you let it drop so that it gain speed is that, you have a longer path. So, if you want to increase its speed or velocity, you need to get a slightly longer path, but it will be very high speed, so then you have a good chance of getting reaching that quickly from A to B quickly.

On the other hand, if you are just choosing the shortest length, then maybe velocity is not that great. And typically, is like that. And so, then you might not, even though the time must might be reduced, because the path is shorter. But they are still not very clear. So, there are two things at tension here, longer path, high speed, shorter path, low speed.

So we do not know which is better. Now you have to find a compromise between these two. So, many of the problems, we will talk about segmentation problems, we will talk about, this kind of tension will be there between two different quantities that we are trying to minimize. And those situations is where, calculus of variation comes in handy. It provides you a systematic way to arrive at correct solutions.

So in this case, what you are trying to figure out is the equation of this curve. So mathematically, the form, the problem is formulated as follows:

$$T = \frac{1}{\sqrt{2g}}\int_{a}^{b}\sqrt{\frac{(1+y'^2)}{y_a-y}}dx$$

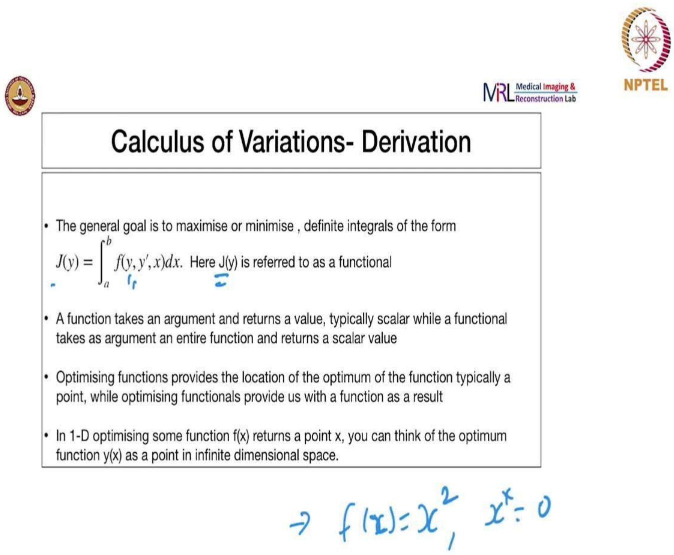
So, we want to minimize the total time taken, which is what this integral gives you and we saw that dt can be returned as ds/v and that in turn can be written as $\frac{1}{\sqrt{1+y'^2}}$. So what is y',y' is nothing but y' is dy/dx, I will just write it out here.

This is a very common derivation, you can actually Google this and you will find it everywhere, Wikipedia pages for it. So, I am not going to go into detail but eventually, you have to do a few substitutions basically, for the velocity also, because the v, you can get by conservation of energy. So if y_a is the, maybe I will have to zoom in a bit. So, y_a is the position of the starting point so, this is the starting point coordinate of the starting point y_a. If it drops a distance y, then it gains some kinetic energy.

So initial energy is purely potential. So, using conservation of energy, you can evaluate an expression for v from this equation above, and then substitute back in the integral. So that will give you this integral expression for the time, a definite integral expression for the time, which is $\frac{1}{\sqrt{2g}}$, an integral a to b and whatever is in respect of this of course this on the numerator, this is y'^2. Just keep that in mind. Maybe I will zoom in a little bit, just to make sure.

So, this is the integral that we want to optimize. So, you want to find the minimum time, so this integral is what is referred to as the functional. So, this is a function. So, functional takes as input a function, in this case, a curve and returns a scalar value. So, in this case, it is the total time taken, and we want to minimize the functional to arrive at an optimum path, which is basically y we want to figure out $y(x)$. That is what we need. So we will, we will revisit this once you finish derivation. If you want to minimize time, which is basically this is the functional then you have to find a path y. So it is entire function.

(Refer Slide Time: 10:03)



So, the general goal here is to maximize or minimize definite integrals of this form. So, if you note, it is a definite integral a to b and f here, the integrand is a function of y, y', which is

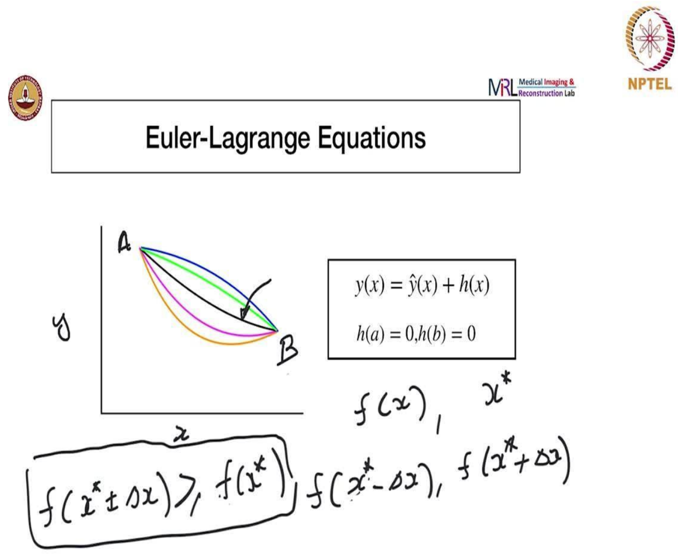
dy/dx and x. This J of y itself is referred to as a functional. So, like I said earlier, a function takes an argument. For instance, if it is a 1D function, defined on the real line, let us say, takes as x an input on the real line, and returns a value which is typically another scalar on the real line.

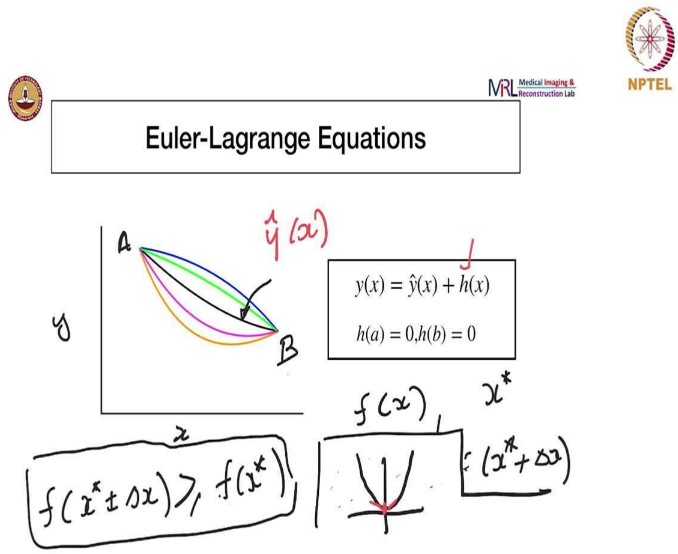
Typically, a scalar, while a functional in this case takes as input a function, the input to this function is y, y is a function, it is a curve, but it is still a function and returns a scalar value. So, when you optimize functions what it does, it provides the location of the optimum. So let us say f of x is x squared, let us say you are optimizing something like x square. Location of the optimum is x star equal to 0.

That is where the function, $f(x)$ takes its lowest value. So, in this case, when you optimize this function, what you are trying to do is to estimate y the function itself, so you want to estimate a function, which will optimize this function. Now there are some other aspects like this is actually defined in infinite dimensional space, because if you look at this typical function optimization, let us say 1 D, f of x equals x squared.

Now x is just in 1 dimension, but when you consider you are trying to estimate for an entire function, then the problem is defined in infinite dimensional space. Here, again, infinite dimensionality comes from how you represent the function and in what basis you represent the function. So for instance, if you use Fourier basis, or you are using Fourier series of Fourier transform, there are infinite number of terms. So consequently, you are in an infinite dimensional space, making the problem a lot harder.

(Refer Slide Time: 12:08)





So just to reiterate, so let us consider this picture here and I will tell you what are we are trying to explain here. So let us assume that the solution to the problem we studied. So that solution here is the x and y, solution we wanted. So the solution curve, which is basically the curve that you want, estimate from A to B, put A here and B there. So, B here, and let us say the black curve is the correct solution. The curve that gives you the shortest time of travel from A to B.

So that is the black curve. I have to use color that I have not used here, so I will stick with black, so black curve is the correct solution, that is the one. So we all know that when you work with minimization functions, etcetera, let us say we have some, we are not talking about a different topic. So let us say we have this function $f(x)$ and this is defined on the real line. So x is probably minus infinity to infinity, take on all values on the real line.

And $f(x)$ is a scalar valued function output is also a scalar. So, it returns to scalar. So let us say x * is the point at which $f(x)$ attains a minimum, we will just talk about minimum. So x * is the x, x * is x at which $f(x)$ becomes minimum. So, if you look at any perturbation around x *, so if you look at, let us say, x *$-$ Δx, and also you can do $f(x$ *$+$ $\Delta x)$, small perturbation.

So if you look at this, then you see that for all these values, when you move to the left a bit or to the right a bit, since x * is the minimum. Well, there are other assumptions here, I am assuming that there is only one minimum for $f(x)$, the global minimum or no local minimum, etcetera.

We have made all the necessary assumptions. But then still, we are only looking at the neighborhood of a minimum. So, x *$-$ Δx and x *$+$ Δx if you move either to the left or right, then $f(x)$ in that case, for all if you $f(x$ *$-$ $\Delta x)$, and $f(x$ *$+$ $\Delta x)$, and everything in between except for x * will be greater than the value at $f(x$ *$)$. So, all value of $f(x) >= f(x$ *$)$, so this is true.

If your x * is the minimum, you show the minimum value of $f(x)$, and it is at x *, then if you perturb a bit perturb the value of x around x *, then all values of $f(x)$ will be higher than that, will be greater. So, it is very simple if you let us say, if you draw, let me wipe this out and draw so it is easier. Very simple function like this parabolic function. So if this is your minimum there, if you move to the right or to the left, the function value increases. That is what I am trying to say.

So, similarly, but then this is in the real axis, of course, you can also consider it in multiple dimensions x, y z or x1 x2 up to xn. And the argument remains the same. If you perturb it a little bit around where it attains its optimum value in this cases it is minimum value, the x values, the coordinate values, then the function value will increase. So how do you do similar perturbations in, when you are trying to estimate function?

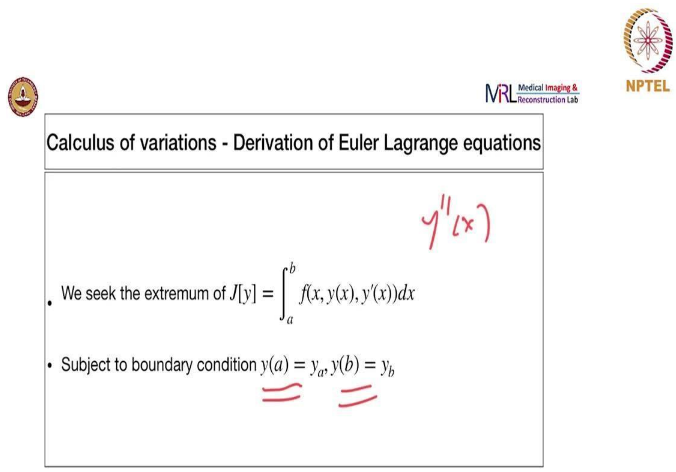
So, if this is our correct optimal function, this black curve is our $yhat(x)$. This is the path that gives you the shortest time in our problem we are studied. But in general, for any such definite integral, we saw, let us say this is the function that gives you the least value of the functional, then how do you perturb this, so you perturb this by adding a test function, h of x, that is how you do, here you added Δx, you are adding $h(x)$.

And it is of course satisfy the boundary conditions here, because boundary conditions is that, in our case the particle or the point mass starts at A and ends up and B. So, the boundary condition is satisfied. So, which means $h(a) = h(b) = 0$. So, they are at the ends of the path, your boundary conditions on it and this is true for any other problems, they will always be a boundary condition, and you have to meet them and this is one way to meet it.

So $h(x)$ is a think of $h(x)$ as a small perturbation to $yhat(x)$, which is the correct or the correct curve that we are seeking. This is how we perturb it. Because it is like you said it is, it is an infinite basis for infinite dimensional space, the only way you can perturb $y(x)$ is by adding another test function to it.

So this is, of course, this is the important point that you have to understand, in the calculus of variations, and how do you perturb this function. We are not done yet. So, we will have to look at what will further calculations are required in order to arrive at the so called Euler-Lagrange equation.

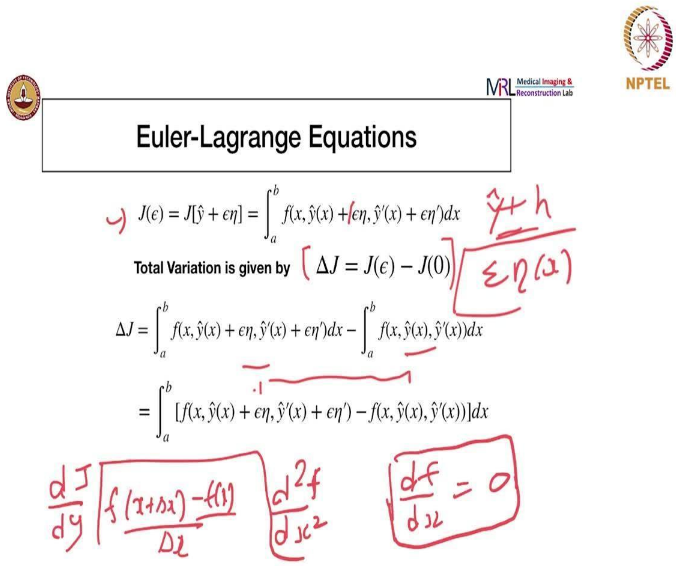
(Refer Slide Time: 18:12)



So we seek the I call it the extremum of $J(y)$ so but you have to pass this in your mind saying, we seek to find some y such that $J(y)$ attains a minimum value or you can also say maximum value, but always change the sign and do that. And there are boundary conditions to be met such that $y(a) = y_a$ and $y(b) = y_b$, these are typically given. So you understand that.

And the integrand is some function, which is a function of x, y of x and y prime of x. Sometimes you can also have higher derivative, so $y''(x)$, we will see that this actually happens in some of the image processing problems. So, we will see about that when we get there, so we seek the extremum of $J(y)$ that is what I call it extremum. Once again this optimization literature has different definition etcetera. So, I am just being very informal, so do not hold me to this. So, I assume that we are just trying to find out y such that $J(y)$ is minimal, this functional is minimum.

(Refer Slide Time: 19:16)



So, like we said, we have to perturb it around the optimal solution. So, which is basically we assume that $yhat$ is that function for which $J(y)$ attains minimum value. So, we then try to perturb it, that is what we do. This is again, a familiar trick. However, when you also try to take derivatives, then you do something like that. So, we have as infinitesimal decimal change in the arguments, and then we try to see how the function changes.

Similarly, that is what we are trying to do here except that we have to parameterize carefully. So we saw that, we are trying to, in the previous slide, we tried to do $yhat + h$. That is much harder to do, actually. Because what do we eventually want? We want to have estimate this, you want to estimate. So that is what you would do, if you want to find out the maxima or minima of let us say, an arbitrary function, I am using f, here all the while, but you feel free to put something g or if you want h also.

So if you write it as df/dx, and set it to 0, this will give you a maxima or minima and you would look at the sign of the second derivative like and you will also examine the sign of d^2f/dx^2 and that will give you whether it is a maxima or minima. But either case this is the first step, you have to set this to 0.

And the way you do that, now, if you want to start on principle and you will perturb $f(x)$, left hand side is $f(x *+ \Delta x)$, etcetera. So we want to write dJ/dy and this is hard to compute using just this kind of trick. But then we have to perturb everything by h and then h'. And then you have to say h is small. So, because when you do f right, this kind of derivative, you will do by $f(x *+ \Delta x) - f(x)$, something of this.

I do not want to write too much around here just to give an idea. So, you will do $(f(x *+ \Delta x) - f(x))/\Delta x$. That is how we estimate the first derivative. So, this is the limit of Δx and 0, that gives you df/dx. So that is the idea. So, but then to do that, here, you have to do something or we have to reparameterize so that it is a lot easier to do.

So what you do is instead of h, you write $y(x + \epsilon\eta(x))$, so that is what we have here, this term so h epsilon eta of x. So $\eta(x)$ can be any function, because we assume that it is typically well behaved that it does not have some nasty discontinuities. But generally, η is any function, $\eta(x)$ is any function and ϵ is a very small number with ϵ is a very small increment that you add on.

So this way, you can perturb around y hat, you can perturb around y hat and see how this changes. So now, J is no strictly a function epsilon. Why is that? Because y hat is fixed, we know assume that y hat is the correct curve that we are seeking. And the curve for which J of y is minimum, or $J(yhat)$ is the smallest value $J(y)$ can take and we perturb it with $\epsilon\eta$, we fix η also, we say it is a test function.

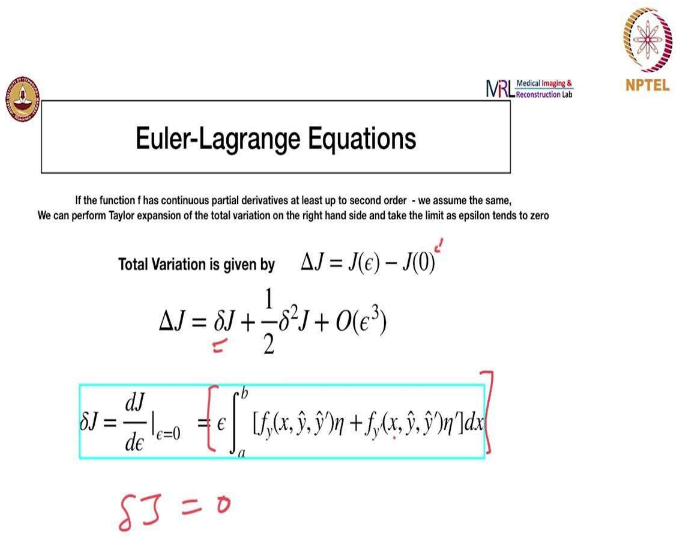
It can be any function without loss of generality, we can say it can be any function η and which of course, like I said, it has desirable well behaved function, we do not want some pathological function, something that goes into infinity and all that we do not take, we just say, it is a function, which well behave as the derivatives first and secondary derivatives exist.

And then so then we can just use that to perturb. So now that is also fixed, so the only thing that we are changing is ϵ, so that we are actually searching in the neighborhood of y hat, just making sure that for any such value of epsilon, your function takes, J is the functional takes higher values than at $J(yhat)$.

So then we have calculated what is known as the first variation. But before we do that, define what the total variation is? Total variation is given by this expression, $\Delta J = J(\epsilon) - J(0)$. So, what is $J(0)$, just put epsilon equal to 0 with just the functional, which is estimated at $yhat$ and $yhat(x)$ and $yhat'(x)$ So now we can write the total variation in this form, I have just substituted for J of epsilon here, and J of 0 here.

Now that the second step is of course, bring them under the same integral, make them one integral under the same integral. And if you look at this, we can just then now you do your usual tricks like you can do a Taylor series expansion of $f(x)$, y hat of x plus epsilon eta y hat prime of x plus epsilon eta prime. So, you can think of $f(x)$ as a function of several variables, and you do a Taylor series expansion. So, and then once you do that, you will see that this term will cancel out and what is remaining, this what we will refer to as the you can call it the first variation.

(Refer Slide Time: 24:09)



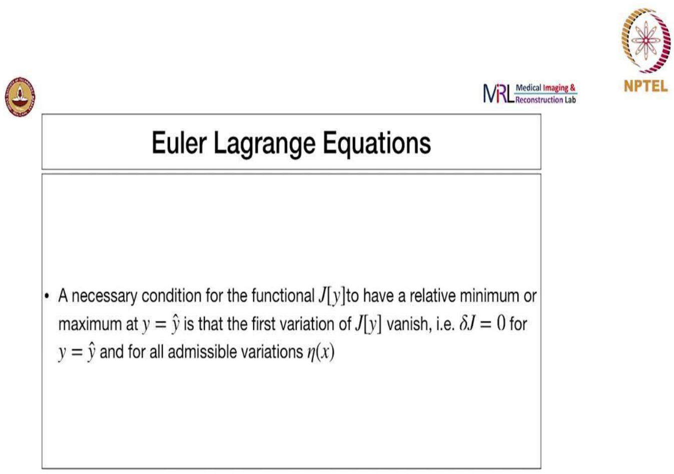
The left hand side also, the total variation can also be expanded in Taylor series and once again, the J of 0 will then get subtracted out, because that is what we are expanding around, around J of 0. And once we do that, then we get this term which is known as the first variation. First variation, which is nothing but $dJ/d\epsilon$ at evaluated value epsilon equal to 0 and it will turn out that it is this expression on the right.

And in order for the minimum to occur at y hat, we should get done so that ΔJ should be 0, it is not too hard to see because you see the Δ, the expression for ΔJ is an odd function of ϵ. So, if we change the sign of ϵ then what happens is automatically changes the sign here, ΔJ. See, $\Delta J = J(\epsilon) - J(0)$, this is the minimum, $J(0)$ is the minimum.

Any other perturbation should only increase the value of J. So, but if I change the sign of epsilon, you can show that this will become a positive number and if I change the sign of epsilon will change the sign of the value of δJ. So, that means that you can get higher or lower values. So that is not correct. So, then you have the set $\delta J = 0$. So we can show that

$\delta J = 0$, that is okay make a theorem you can prove. And once you set that to 0, then the next step.

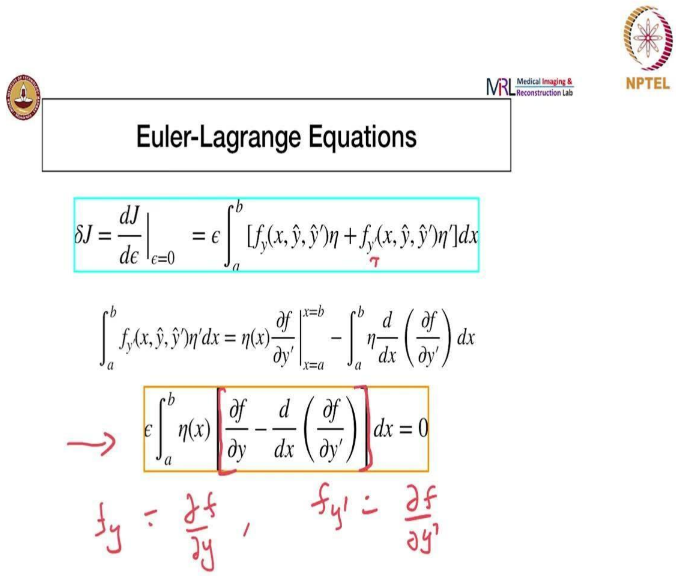
(Refer Slide Time: 25:54)



So a necessary condition for the functional $J(y)$ to have a relative minimum or maximum at y equal to y hat is that the first variation vanish. This is like a theorem. I am not going to prove it, but I just gave you like a kind of hand waving argument because δJ which is called the first variation is an odd function of ϵ. So by changing the sign of ϵ, we can make it lower than $J(0)$, which is not true because $J(0)$ corresponds to y hat and that is we know that when we have assumed that they were defined in this that it is the value at the correct or the optimal solution of y.

So then, by changing the sign of epsilon, we can make the total variation, positive or negative. So, to prevent that, the only way that can be prevented is by making $\delta J = 0$. So, the theorem once again, is for the necessary condition for the functional to have a relative minimum or maximum at $y = yhat$ is that the first variation of $J(y)$ vanish, that is $\delta J = 0$ for$y = yhat$, and for all admissible variation $\eta(x)$.

(Refer Slide Time: 27:10)

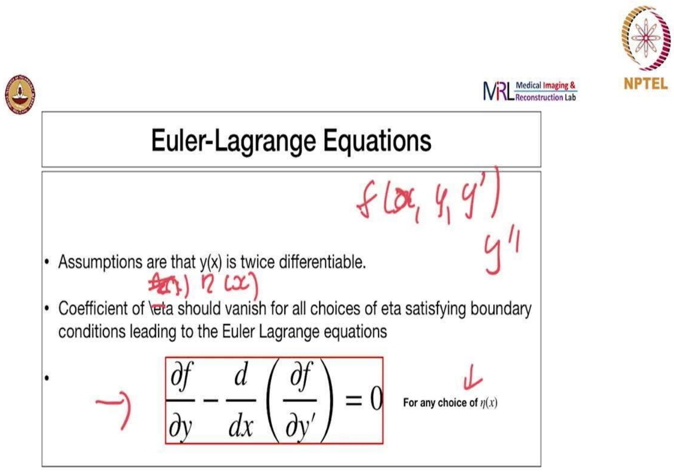


So once this is done, then we can go here and look at what happens if we set this to 0. Again, it is other way of looking at it is that you are setting the first derivative to 0. Easier way to look at and set that to 0. And once you set that to 0, one other thing that we have to do is we have to simplify this expression.

So, this f of y prime once again, sorry, if I have not mentioned this earlier, f of y, f underscore, f subscript y is nothing but $\Delta f/\Delta y$ and f subscript y prime is $\Delta f/\Delta yhat$. So we assume that, so we can do integration by parts, so we take the second expression, we can do integration by parts, and use the boundary condition, to eventually come at this expression.

Now, all you do is focus on the term in the brackets, it turns out that we can prove, in order for the integral to be 0. For any data x, this term should go to 0, the term that I have put in red bracket should go to 0. That is another theory. See we did integration by part for the $df(y')/dy'$ and if you have y integrated, we will have a derivative with respect to y w prime when f is a function of y w prime also, then I will have another term, we will have to do integration by parts twice.

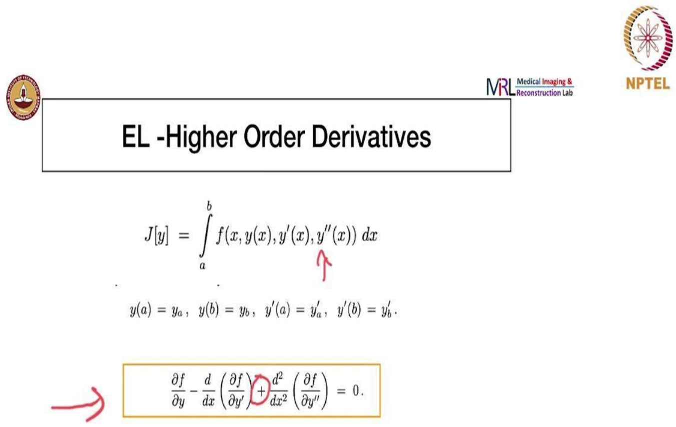
(Refer Slide Time: 28:34)



So the coefficient of η f x should vanish for all choices of eta satisfying boundary conditions leading to the Euler Lagrange equation. So, these are the Euler-Lagrange equation for any choice of eta. So now, this equations are partial differential equations and this is what used to actually estimate y, see, that is a function of both y and y'. And in this case, because we started out with the f strictly as a function of x, y and y'.

Sometimes it is also a function of y double prime, which is d squared by dx square, in which case there will be an additional term with a change in sign. So, we will look at that in another slide. But that is one of the things that you have to remember, it can also be a function of higher derivatives. And for that and what happens is that in order to solve that if I go back to previous, and so there will be one more term here, we will look at that later.

(Refer Slide Time: 29:46)



This one here as I was talking about if $J(y)$ and where f is a function of y'' also here, then you will end up with an equation like this. Of course, the sign will be flip, not that the sign is flipped, that is because you will do integration by parts twice and there will be a minus sign, extra minus sign which will make this positive. So, this is the Euler-Lagrange equation and this what is the driver for all the methods that rely on all the PDE based methods and that rely on Calculus of variations.

So we will look at something called Snakes Active Contours that is one of the earliest techniques that was derived using the Euler-Lagrange equations. And then subsequently other techniques also came about and they used something called other technique called level sets to make computation simpler and more efficient. So, we will look at that in the subsequent videos, so that we will stop here.