Dynamic Data Assimilation
Prof. S. Lakshmivararhan
School of Computer Science
Indian Institute of Technology, Madras

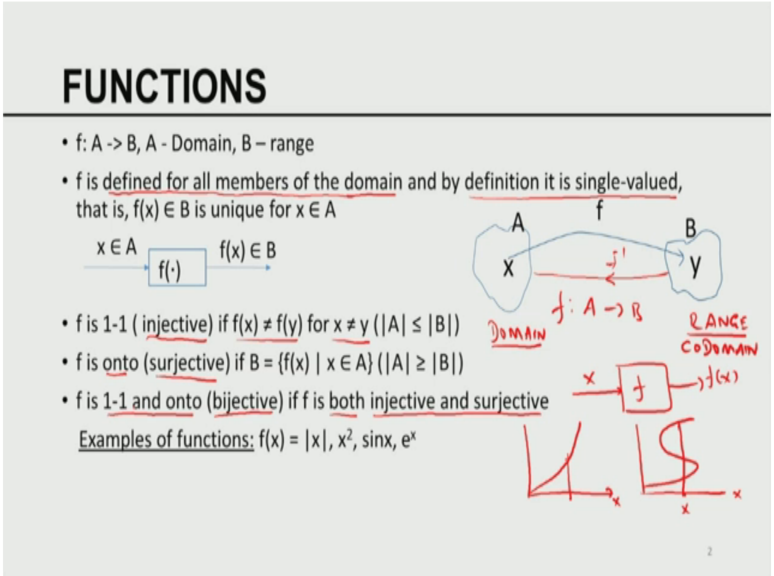
Lecture – 07
Multi-variate Calculus

In the previous 2 modules, we reviewed the fundamental concepts from finite dimensional vector spaces. And then various properties of matrices one would normally come across in the analysis of data assimilation algorithms. In this lecture we are going to be also providing a broad overview of the fundamental tools that we need from multivariate calculus.

The reason for including this is as follows. Almost all the students who take a bs degree in basic sciences or engineering in any part of the world, they have done calculus 1 2 3 4, in equivalent of calculus 1 2 3 4. They have been introduced to univariate, calculus, differentiation, integration, differential equations all in one variable.

But when you formulate a problem in a data assimilation framework, the problems have to be formulated using multivariate analysis. The state of a system is defined by a vector x. The dynamics of a system could be linear or non-linear for a linear system the matrix defines the state of state transition. So, we need to be very familiar with the multivariate analysis. The tools for multivariate analysis include understanding of (Refer Time: 01:47) space a thorough understanding of matrices and properties and also, a good facility with dealing with multivariate calculus which is an extension of the ordinary univariate calculus that, anybody who does a BS degree should be different way.

(Refer Slide Time: 01:54)



So, our goal is to be able to provide you a broad-based introduction to fundamental concepts from multivariate analysis. So, let us start with from fundamentals.

We are going to start with the notion of functions. To be able to define a function we need different objects. One is a set A, another is a set B. A is a, f is a function from set A to set B. We call a the domain of the function. We call a b the range of a function this is called the domain of the function. This is called the range of the function. So, I need a domain, I need a range, then I need a function what is function a function is simply an association of points that the domain with the points in the range.

By definition f has to be defined for every member of the domain. The value of f is the range we need not take all the values in B, that is where the distinction between various functions coming come to be. So, by definition f is defined for all members of the domain; that means, you can leave anybody here. By definition a function is also called single valued. What do you mean a single valued? If you think of function as a black box, if you give an inputs.

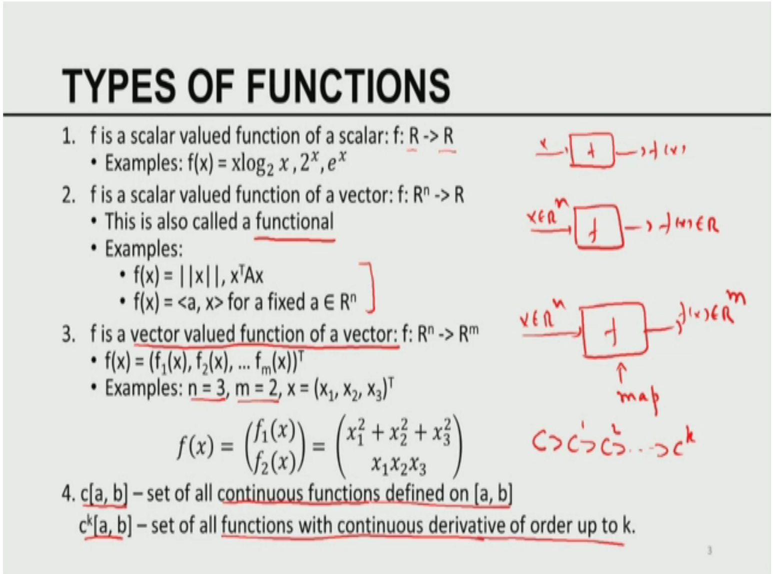
It gives an output f of x. For every x there is a unique f of x it has a single valued. What is the difference between the single valued function and a multi valued function? This is a single valued function, what is the multi valued function? This is the multi valued function if I took x. So, this is x, this is x, here for x there are 3 values. This is a single valued function. So, while in principle, one can have single valued functions on multi valued function in mathematics, we exclude multi valued functions from consideration. So, a when a mathematician says, let f be a function he already has the back of his mind, a domain, a range. A range is also called codomain, and an association between points in the domain of the codomain.

By single valued means it is unique. So, that is the broadest possible way one can define functions. Now there are special class of functions such as special class of matrices. So, f is a function is called one to one I calls it injective, what does it mean? If x is not equal to y f of x is not equal to f of y; that means, distinct points are mapped into distinct points in the range; that means, y is equal to x square. So, distinct values of x have distinct values of y.

So, that is what is called injective. F is called onto or surjective means, f is maps all the points of A on to complete set B. F is called one to one and onto it is also called bijective. It is both injective and bijective. So, let us give some examples of functions. F of x is a absolute value of x x square sin x e to the x. Sin x e to the x these are all examples of functions. And these are different classifications of functions.

We will more often be interested in one to one functions which are both injective and subjective. Because it is for these functions, we can have inverses. So, if f is a function from here f inverses a function from here to there that is that is the inverse function. In order that the inverse is defined we need to be able to have further constraints the constraint is f must be one to one or injective.

(Refer Slide Time: 06:36)



These are all facts essentially come from basic definitions of functions.

So now I am going to talk about other classifications of functions. F is a scalar valued function of a scalar. So, what does it mean? Here is a black box sorry, here is black box f. This is x, this is f of x. X is a scalar f of x is scalar that is what is called scalar valued function; that means, the domain is real the codomain is real. So, what are examples of functions which are scalar valued function, x is equal to x log x 2 to the power of x e the x, these are all scalar valued function of a scalar the input a scalar output is a scalar. So, you can think of f as a transformation as a black box.

Something goes in something gets out. F is a scalar valued function of a vector. So, in this case what happens f x belongs to R of n, but f of x comes in f of x belongs to R. It converts a vector into a scalar. So, that is what is called scalar valued function of a vector. Such a thing is also called functional; we have already seen the notion of a functional when we talked about vector spaces. So, what are examples of scalar valued function?

Given a vector x the nama of x a nam is a number associated with the every vector that is a number. So, a nam is a function is a scalar valued function. Quadratic form of x is a scalar valued function. Inner product of x with a for a fixed a that is a scalar valued function. So, these are all examples of scalar valued function of a vector. The input is a vector the output is a number. But in dynamical systems theory as well as in data assimilation, we are going to be interested in a third class of function which are called vector valued function of a vector.

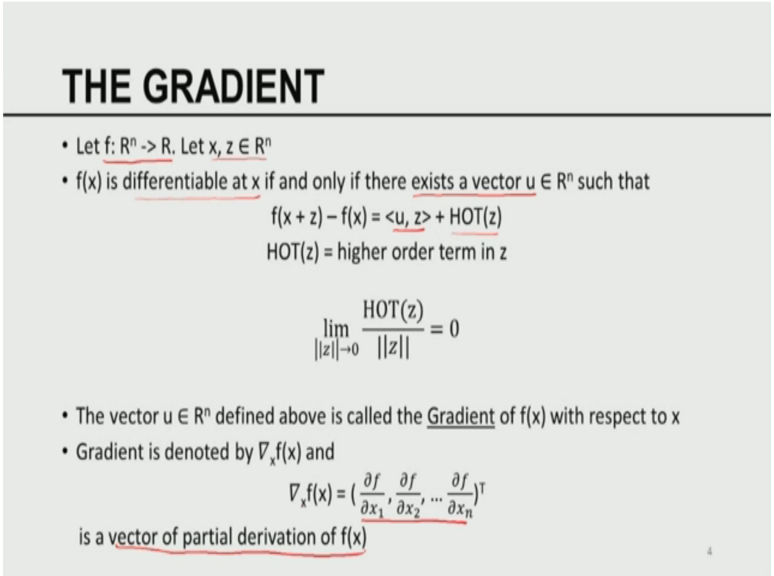
So, what does this mean here? I have a box f. X gets in, f of x gets out. X is a vector, f of f of x is also a vector. This is called vector valued function of a vector. In general f is also called a map map is a very technical term used in dynamical systems theory. Let f be a map what does it mean f is a scalar valued function of a vector input is a vector output is a vector. So, let us in general these 2 n did not be the same it could be a n vector this could be m that I wanted to see the difference.

The input of the vector output of the vector, then the vectors could be of same size are of different size. So, let us work an example give an example let n be 3; that means, input vectors are size 3 let m be 2 output vectors of size 2 the x is equal to x 1 x 2 x 3. So, f of x is f 1 of x f 2 of x. So, what is f 1 of x? X 1 square plus x 2 square plus x 3 square. What is f 3 of x? X 1 x 2 x 3. So, you give x 1 x 2 x 3 you get this vector given by this. So, that is what is called a map or a vector valued function. C a b denotes the set of all continuous functions defined over the interval. I can that is the huge (Refer Time: 10:09) an infinite.

Set c k of a b is the set of all continuous functions with derivatives of order up to k. If I said c a b continuous continuous function need not be differentiable, but in the second set c k a b it not only be differentiable, but also I would like you differentiable up to the order k. So, what does this mean? I have c, I have c 1, I have c 2, I have c k. Continuous function in the larger set. Differentiable continuous and differential from function be smaller, functions which are priced differentiable smaller.

So, you can think of a relation. These are super sets. C 1 is a subset of c c 2 is a subset of c 1 which is the subset of c. I am putting greater conditions on the behavior of the function. So, functions come in various shapes and forms, functions are of various types continuous function differentiable functions set of all continuous function set of all differentiable functions of order up to k, where k is an integer.

(Refer Slide Time: 11:20)



If k is infinity, what we call? Such functions are called an analytic functions. For example, polynomials are analytic functions. They have different they have derivatives of all order. Exponential function or analytic function they have derivatives of all orders and so on.

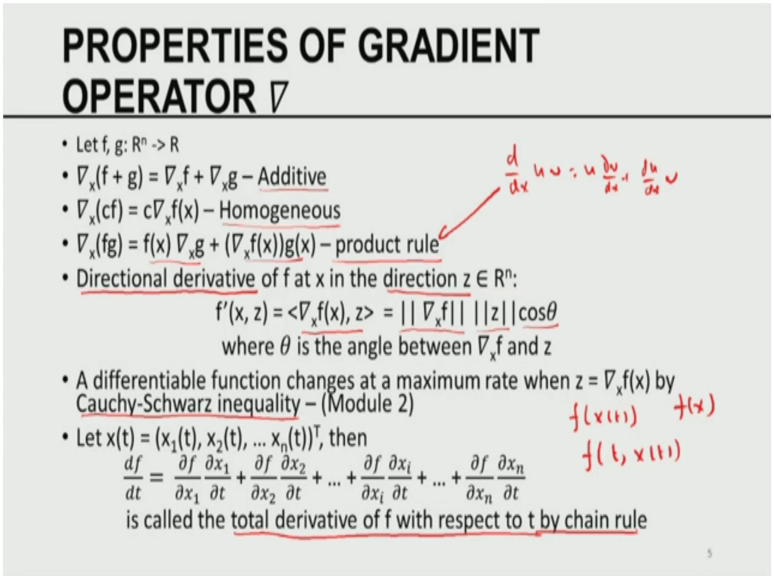
The that is a background, now I am going to introduce various concept that we would need in trying to do data assimilation algorithms, especially optimization algorithms. We are introducing the notion of what is called a gradient of a function. So, in this particular case we are concerned with yes, I am sorry we are concerned with a scalar. So, what is the starting point? Let f be a scalar valued function of a vector. The type 2, let x and z be 2 vectors in R n.

We say x f of x is differentiable at point f if and only if there exists a vector u such that f of x plus z minus f of x is given by u times z. See this z and this z are same. So, the definition is contingent on the existence of a vector. So, this is an inner product. Hot means, higher order terms higher order terms in z, and what is the property higher order term? The ratio of the higher order term to the norm of z, they go to 0 as z goes to 0, this is the limit. So, such a u is called the u is called the gradient of f of x with respects x.

So, this is the most general definition of a gradient. This gradient algorithmically can be computed as a set of partial derivatives of f with respect to x 1 f with respect to x 2 f with respect to x n. So, this is an n vector. The derivative is denoted by the inverted delta subscripted by f of x. This is called a gradient, you can call it. So, we use the term derivative for univariate gradient for the multivariate. So, even though f is a scalar valued function of a vector it is gradient is a vector in R n.

So, I wanted to be able to see the importance of introduction of vectors and matrices very soon. So, you cannot do multivariate calculus very well, until and unless you understand final assumption of vector spaces as well as matrices very well. So, what is gradient in simple terms? Gradient is simply a vector of partial derivatives. That is a simple form of being able to describe what a gradient is.

(Refer Slide Time: 14:11)



This operator inverted del, which is the gradient operative? It has lots of interesting properties. Let f be a scalar valued function. G be a scalar valued function the gradient of the sum is some of the gradients. So, a gradient is additive as an operator as an additive property. Gradient of a constant times the function is constant times the gradient of a function that is called the homogeneity property.

Gradient of the product has this rule which is called the product rule, which is the extension of the product rule in univariate calculus. We are used to d by dx of u v is u dv by dx. So, we already know this right d by dx of uv is equal to u dv by dx plus d u by dx times v. We call this product rule. So, this is the analogue of the product rule from univariate to multivariate.

In the multivariate calculus, we are also interested in another concept called directional derivative. So, f is given f is a scalar value function of a vector, I prespecified direction z, I would like to be able to compute how the function varies in this direction. So, it is called the directional derivative. That is defined by f prime x comma z x is a function. So, f of x is a function, f prime x comma z is the directional derivative of f of x along the direction z, and that is given by the inner product of the gradient with z.

By Cauchy Schwarz inequality, this inner product is equal to the nam of the gradient times the nam of z times. Cosine theta, theta is the angle between z and the gradient. You can readily see that. So, Cauchy Schwarz inequality we already saw in in in the previous lectures. So, this is essentially an application of Cauchy Schwarz inequality that tells you how the directional derivative the magnitude of the directional derivative can be obtained by computing these.

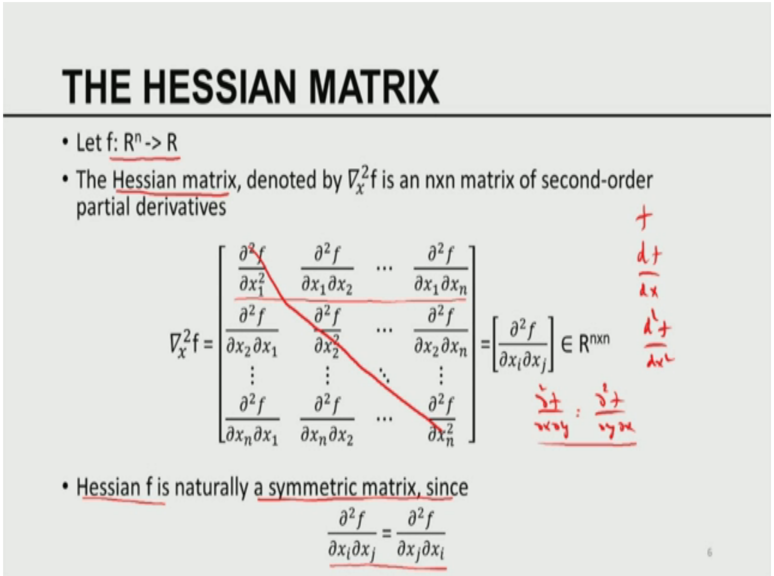
Now, we are going to be talking about a slightly related concept. Until now we assume x is a variable in itself, but here x is not a variable x is a function of another variable. So, x of t is a vector each component is a function. So, x n of t is a scalar valued function of t x 2 of t is a scalar valued function of a t x n of t likewise. So, I have a vector function each component is a function of t. So, if I'm going to be computing the total derivative of x with respect to the time t df by dt is, df by dx one times dx 1 by dt.

So, what is that I'm? Now talking about I have talking about f of x of t. So, if I am interested in. So, f so, I would like to talk about a couple of things now. F of t of x of t a these are all different functions. F of x so, let us talk about all these functions. F of x, x is a simple variable. F of x of t f is a function of x, x is a function of t. So, it is a function of a function. F depends on t in this case only implicitly.

In the third case f depends on t both explicitly and implicitly. So, this is no dependence, implicit dependence implicit and explicit dependence, in these cases we should be able to carry out the computation of the derivative of f with respect to t. So, derivative of f with respect to t is the total derivative. Derivative of f with respect to x 1 x 1 times derivative of x 1 with respect to t so on and so forth. So, this is called the total derivative of f with respect to t by chain rule.

So, chain rule, additive rule, homogeneity rule, product rule that we have learnt in basic calculus all carry over. There is nothing new, but the old concept take a new form when you go from univariate to multivariate, that is the idea.

(Refer Slide Time: 18:39)



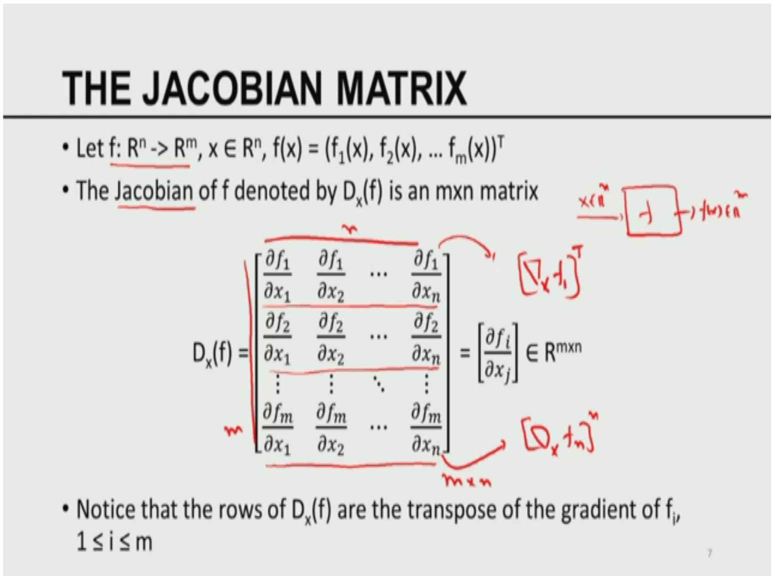
Next was the notion of what is called second derivative. If you f is the function of a scalar second derivative, we simply say the second derivative is simply given by. So, f df by dx d square f by dx square, we are done. But when f is a function a scalar valued function of a vector like this, x is not one there are n variables x 1 to xn. So, I can gradient is a vector, I can consider the second derivative matrix, the second derivative matrix look at this map, the first row of this matrix second partial second partial of f with respect x 1.

Second mixed partial of x 1, x 2 second mixed partial of x 1. With the x n likewise, each row such a matrix is well define this matrix is a special name is called the hessian of f. So, you can see matrices arise very naturally not only that, we know from basic calculus the mixed derivatives are essentially the same; that means, del f by del x del y is; I am sorry, del square x by del. So, del y del x the mixed partial derivatives. If the partial derivatives are continuous the mixed, partial derivatives are same. The order in which you compute the partial derivative is the material. So, given that this matrix is a symmetric matrix.

So, n by n symmetric matrices naturally arise when you consider the second derivative matrix which are called hessian matrix of functions of scalar variables. So, hessian is natural symmetric. So, symmetric matrix so, where do symmetric matrices come from? Symmetric matrices come from various directions. One of the simple ways in which symmetric matrices arise is by computing the second partial derivative matrix of a scalar valued function of a vector.

And this matrix a singular because the mixed partial derivatives are the same and that is what I told you a minute ago. So, this is called the representation of second derivative matrix for a function, which is a scalar valued function of a vector.

(Refer Slide Time: 21:15)

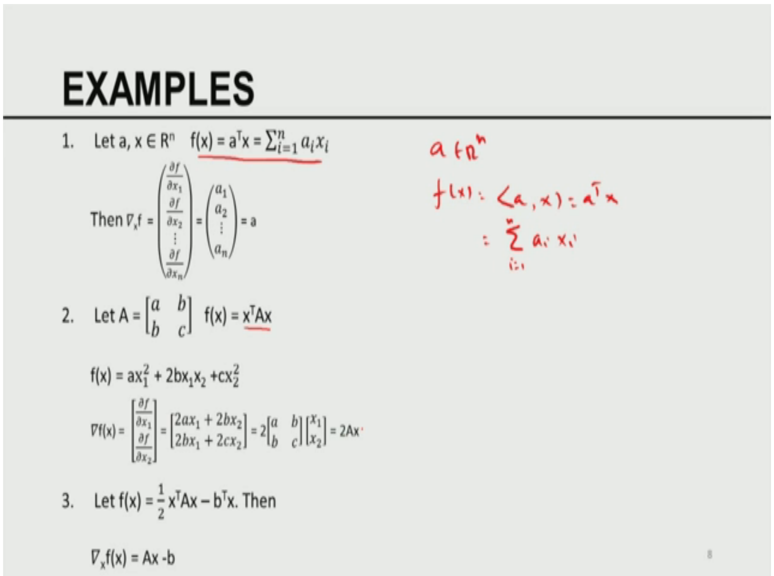


Now, we are going to vector valued function of a vector. We are going to move to the next level. We are so, let f be a function from R n to R m. Look at this now. I would like to be able to keep this picture of the back of the mind. What goes in is x what comes out f of x. X belongs to R n, f of x belongs to R m. So, f of x is a 1 of x f 2 of x f m of x, x is x 1 x 2 xn. So, there are m functions each of which is a function of n variables. I hope that is clear. Now what is that I can do now? I can take f 1 I am compute the partial derivative of f 1.

I can take f 2 I can compute the partial derivative of f 2. I can compute f m I can compute a partial derivative of f m. Now this partial derivative when written as a column is called a gradient. So, what is that? This is the transpose of the gradient. So, this is essentially transpose of the gradient of f of 1. This is simply transpose of the gradient of f of n. So, what is that we do now? We take component by component, we compute the gradient which is a column vector, you transpose it to a row vector, you stack these rows, there is one row for each component of f there are m such component. So, there are m rows, there are n variables.

So, this matrix is a m by n matrix, there are m rows there are n columns. There for if you have a vector valued function of a vector from R n to R m, this is the collective first derivative matrix for the entire function. The collective first derivative matrix is in general a rectangular matrix. This matrix is given a special name is called Jacobean. Such Jacobean of f Jacobian of f is defined only for vector valued function of a vector. Hessian of a scalar valued function of a vector. Gradient scalar valued function of a vector. So, these are all the various quantities associated with functions in terms of their derivatives.

(Refer Slide Time: 23:58)



Now, I am going to give some examples. Let a be a vector. Let x be a vector. A be a constant vector. So, I can define a function. So, look at this now. I pick a; I sorry, I pick a in R n I keep it fixed. So, I am going to define f of x equal to a x which is equal to a transpose x, which is equal to summation. ai xi, i is equal to 1 p m. So, it is a function of x. It is the scalar value function of x, because it is an inner product. The out output is a scalar input is a vector.

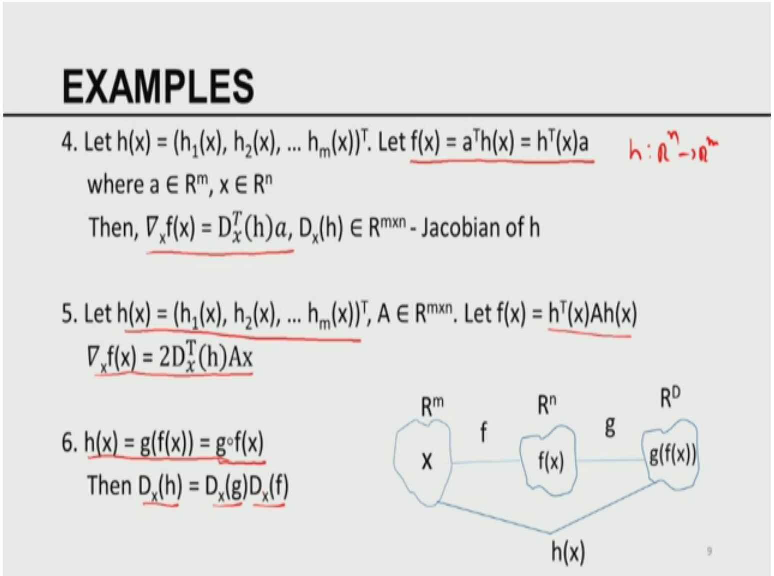
A is a common vector that transforms every input vector. So, what is that? This is simply a linear function because the right-hand side is linear in each component of that. So, what is the gradient of f of x? Partial of f 1 partial of f 2 partial of x n and partial of x 1 is a one partial of x 2 is a 2 partial of xn is a n. So, that is equal to a. So, we have enunciated the first rule of multivariate calculus, what is that? If f of x is equal to a transpose x, the gradient of f is a.

This is very similar to what the univariate calculus person does d by dx of e to the x is e to x d by dx of sin x is cosine x. We develop a table of differential coefficient of various standard functions. So, in data assimilation we need to have such table this is the first entry in the table. So, if f of x is a transpose x, the gradient of f is a. Now let us compute the gradient of x transpose a. That is a quadratic function, a is certainly we talked about with respect to quality function we need to consider only symmetric matrices.

So, let a b is a symmetric matrix f of x is x transpose ax. So, f of x is a x 1 square plus B x 1 x 2 plus c x 2 square. Let me compute the gradient of this f of x. Partial of f with respect to x 1 partial of f with respect to x 2. A simple calculation is shows is this vector. You can rearrange it at 2 times this matrix times this that is 2 ax. Therefore, we got the second entry into our table. If f of x is equal to x transpose ax it is gradient to 2 ax when f of x is equal to 1 half of x.

Transpose ax minus B transpose x the gradient is ax minus B by combining to 1 and 2. Anybody who has done 3 d war should immediately recognize that these terms very naturally occur in 3 d war. So, when you read 3 d war with a diagnosing that these are all tools from multivariate calculus, you will have more trouble now if you know this 3-d war will become as simple exercise. And that is the reason why I believe that it is necessary to understand many of these basic concepts before you start data assimilation algorithms.

(Refer Slide Time: 27:09)



My examples continue. Now I am going to I am getting a little bit more sophisticated. X is a vector; h of x is another vector. So, h is a function. So, in this case h is a function from R n to R m. So, h is a m vector, h has m component h 1 h 2 h m. Each of the components are functions of n variables x is the n variable x 1 to x n. So, let us fix a, a vector. So, let us define a function f of x is equal to a transpose h x or h x transpose a. A this is simply an analogue a very simple analog of what we did in the previous case. A transpose x here a transpose h of x. H of x is a any general non-linear function. So, what is gradient of this f? This gradient of f can be, this gradient of f is given by the transpose of the hessian of h times a.

You remember already we have talked about hessian. So, this is how you are going to be able to compute the gradient of this. Again, I am not in my class I will derive these things in this lecture we may not have time to derive all these things. But it is very necessary that each of you what these examples have that in your mind to be able to handle some of these things to develop there independence, and dexterity in trying to make these calculations and manipulations.

Now, let us (Refer Time: 28:59) given further. I am going to take the same h. Now I am considering h transpose a h. Look at this now. This h here we considered x transpose ax. Here we are considering h x transpose a h x. This is again very often come across you come across in in 3 d war especially with respect to the non-linear observation operator.

So, h is generally used as a non-linear observation operator I am using the same kind of notations and here. So, in this case what is the formula for the gradient of h the formula for a gradient of h, I am sorry gradient of f.

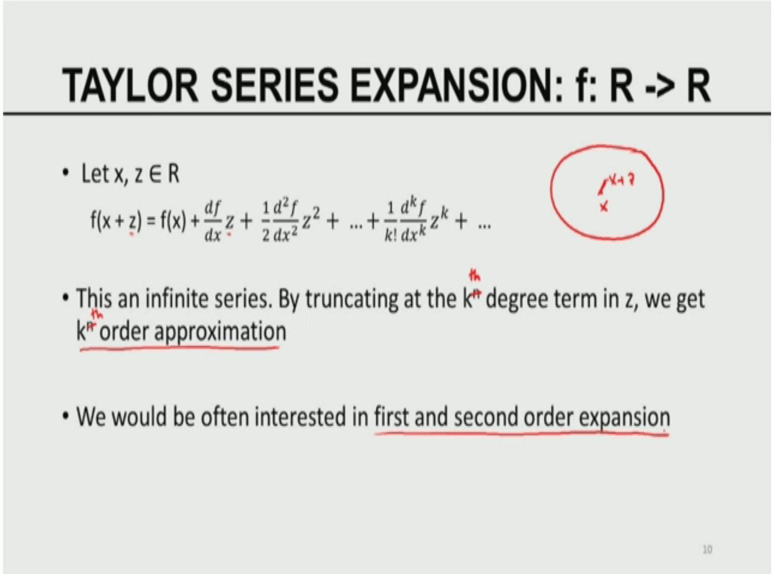
F is given by this is 2 times the Jacobean the transpose of the Jacobean of x times ax. If you it is very imperative to me to me these are all the nuts and bolts. Yesterday someone was observing after the class, who are the people who develop algorithms for data assimilation, those who understand and have good mathematical skills are the one who are going to be able to invent newer algorithms.

So, there are 2 ways one is to use somebody else's algorithm another is to be able to invent your own algorithms to be able to invent your own algorithms, we have to develop all kinds of mathematical skills. And that is one of the underlying purposes of doing um this preview of there many different tools.

Now, considered the next case, h of x is the composite function. Function of a function h of x is g of f of x. And this we denote as so, x you apply f first and then g first in terms of picture. There is R m there is R n there is R d. F takes you from R n to R m to R n g takes you from R n to R d h of x is. In fact, a bridge goes from R m to R d. So, h of s must be related to f and g.

In this way so, this tells you the relation between h and f and g. So, what is the Jacobean of h? The Jacobean of h is simply product of this Jacobean of g at x and Jacobean of f at f at x. So, this is a this is a kind of a chain rule for Jacobeans. This is again a fundamental results. These all are important things that we will apply when we talk about algorithms.

(Refer Slide Time: 31:44)



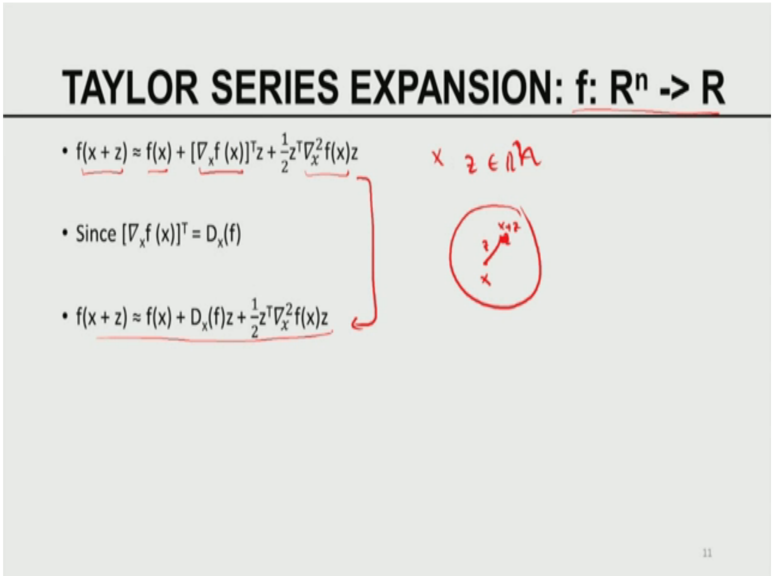
The next topic in multivariate calculus the notion of a Taylor series expansion, Taylor series expansion for a scalar valued function of a scalar valued function of vector valued function of a vector again, there are t 3 layers of Taylor series which are important. So, let x and z be real numbers, f of x plus z is f of x. So, what is the basic idea? The basic ideas as follows, if I have a domain if I have a point x, if I know f of x the derivatives of x all at the point x.

If I am given a point very close by x plus z, this is x this is x plus z, z is small, how can I infer the value of the function at x plus z, given the value of the function and it is derivative at the point x? That is the question Taylor answered. So, the value of the function at a neighborhood point is given by the value of the function at the point plus derivative times z. So, you can think of as a perturbation. So, the first derivative times the perturbation.

Second derivative times the square of the perturbation, k th derivative times the k th power of the perturbation. So, this is a Taylor series, in a small neighborhood under super appropriate conditions. This series will convert; this is one of the fundamental theorems in multivariate in calculus. This is an infinite series by truncating this infinite series at the k th degree term this is not n k th degree term. In z we can get the k th order of approximation, this is k th order of approximation.

So, normally we do not use k more than 1 and 2 we talk about first order approximation second order approximation. So, that is the general rule with respect to approximations in analysis. Either you compute exactly there are not to many things we can compute exactly in life, approximation is the order of the day. So, Taylor series-based approximation is often a very useful approximation computationally. So, this Taylor series is absolutely place of absolutely fundamental role in computation.

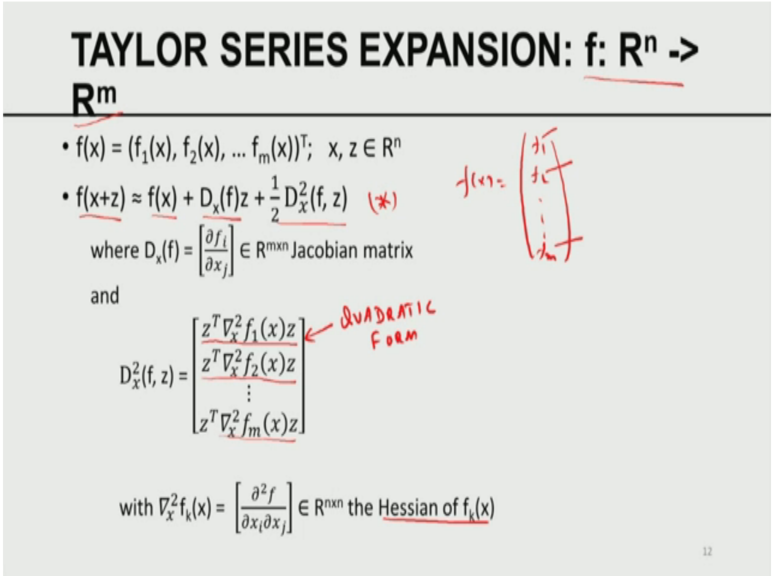
(Refer Slide Time: 34:00)



Now, I am going to consider the next class of functions. The next class of functions are functions which are scalar valued, the input is a vector output is a scalar. In this case x is a vector, z is a vector. Again, the z is a vector, which is R n. So, here again I have x I have z. I have point here the distance, this is z. So, this is x plus z, sorry this is x plus z. Again, if I know the value of the function at x and it is gradient and it is hessian, hessian is a second derivative I can approximate the value of the function of x plus z by this relation. So, this is called the second order approximation.

Second order approximation. We also know the gradient and the hessian are related I am sorry gradient to the Jacobean related by the transpose of each other. So, I can rewrite this using replacing the transpose by the by the Jacobean. So, it is this form we will use in our analysis. So, this is the second order Taylor series of it of scalar valued function of a vector.

(Refer Slide Time: 35:18)



Now, I am going to extend it further. The second order Taylor series for vector valued function of a vector, you can see there are so many intricacies in here. So, what is this f? So, what is f of x? F of x is stack you stack f 1 f 2 and f of m. Each component is independent. So, if I am going to be concerned with the second order expansion for f of x, what is that we do you com you compute the second order expansion of f of 1 f of 2 f of m stacking them together. That is, it very simple.

So, once you know how to compute the second expansion for a scalar valued function of a vector, you have conquered the Taylor series expansion for the vector valued function of vector as a vector valued function is simply a collection of m independent functions. Whatever you do for one does not affect the other. You do everything same for everybody. So, expand everybody in second order term stack them out together collector term you got that. So, that is the basic idea.

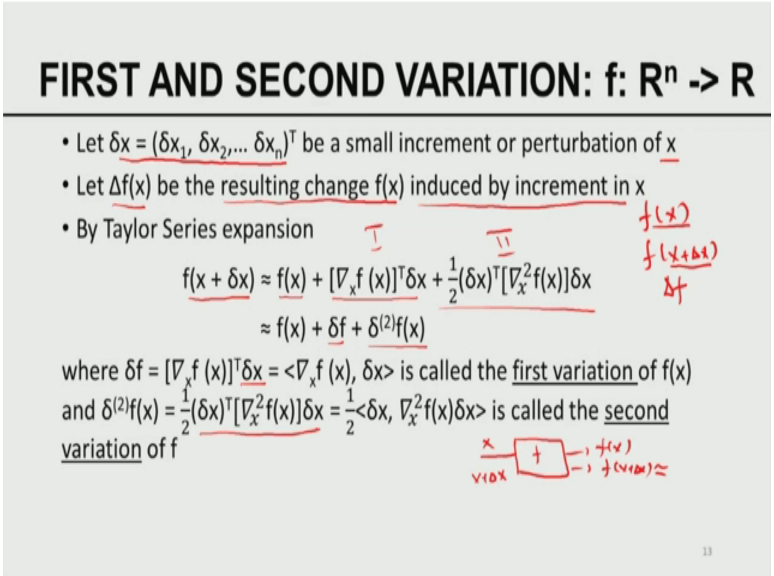
So, with that in mind the second order Taylor series expansion for this is given by this that is a Jacobean term, this is the second order term. This second order term is little bit more complex. Look at this now it is a vector. So, look at this now f is a vector. So, this is the vector, this is a matrix times a vector. This is another vector. One half of a vector and how this vector is design, this vector is given by this look at this now I have f 1. I have hessian of f 1 this is a quadratic form of hessian of f 1.

This is the quadratic from the hessian of 2. This is the quadric from the hessian of f m, stack them all together. So, you get the Taylor series expansion, simply by concatenating are putting together the Taylor series expansion for each of the component. Please remember these are all hessian. Also, please remember this is quadratic form. So, you can see quadratic form occurs in many different ways one of the natural ways of dealing with quadratic forms this second order Taylor series expansion of scalar valued function spectra valued functions and so on.

So, that is where these things come into play. It is unfortunate that once you finish bs, we take our special disciplines and masters electrical engineering mechanical engineering meteorology oceanography and other things. When we take oceanography our meteorology for example, they run you through lots of dynamics and so on which are very necessary.

So many of the meteorology courses are very strong and models. Some of the meteorology program very strong on collection of data, but there are not many there are not programs at all, where much emphasis is given data assimilation models are necessary data are necessary. But data assimilation is something beyond. In my view data assimilation is an engineering discipline, sitting inside the science of prediction. So, this aim of this course is to be able to bring out the mathematical underpinnings of this engineering discipline called data assimilation. Why do I call engineering discipline? Engineering always concentrates on developing a product what is the product forecast. The development of forecast product in my view is a branch of engineering. The product for public consumption, and I would like to be able to create a good quality product by doing a good quality engineering, which is called a data assimilation.

(Refer Slide Time: 39:03)



The next concept is the notion of variation. 3 d war, the war refers to variation 4 d war refers to variation. So, the notion of variational calculus, first variation second variation is fundamental to the development of many of the underlie algorithms. I would like to highlight some of the fundamental properties of this notion, of first and second variation within the context of multivariate calculus. So, let x be a vector let delta x be another small vector the small components.

We call this a perturbation vector or a small increment. So, x, x plus delta, f of x plus delta x, when x changes value of f also changes; change in x induces a change in f. The change in f is called delta f. So, what is delta f? Delta f is the resulting change in the value of f of x induced by the increment in x the increment in delta x. So, what is that you can think of now. There is a black box. That is f. If you give x it gives you f of x. If you gives if you give x plus delta x, it is going to give you f of x plus delta x, but f of f delta x, I would like to be able to express it in terms of f of x itself. I would like to be able to compute approximately what f of x. And that is where the notion of computing there the induced variation.

So, what is delta f delta f is the difference between the new value and the old value the induced perturbation. So, input perturbation induced perturbation delta f is called the induced perturbation this is the input perturbation. So, if f is a smooth function. Smooth function means why it is differentiable up to order 2 that is c 2. You remember the notion of c 2? C k functions.

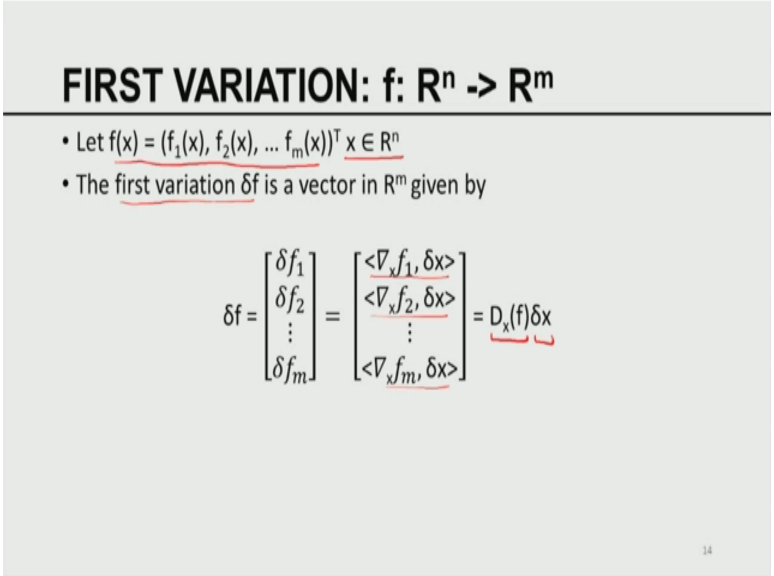
So, to be able to do Taylor series expansion your function you should not only be continuous, but also be differential at least once differentiable. In a function is k differentiable, I can conserve the k th third or Taylor series expansion. So, I'm assuming a very minimum a function is in c 2. If a function is in c 2 then I can compute the increment delta f 2 a second order accuracy. So, f of x plus delta x, that is the actual value of the function at the new point is approximately equal to the function value the old point plus increment one correction.

This is a second correction. This is called the first order correction. This is called the second order correction. This first order correction is denoted by delta f, the second order correction is denoted by delta within bracket 2 f of f. So, this is called the first variation, this is called the second variation. Likewise, I can conserve that kth variation. The larger the order of variation I can add more accurate the value becomes, if you chop off at any level, it is only an approximation that is why approximation symbols are important in here. What is delta f? Delta f is simply the inner product of the gradient with delta x.

What is the second derivative? It is a quadratic form delta x transpose this looks like x transpose a x, what is this? This is the hessian. Please remember, this is the hessian this is x transpose a x. So, that is called the second variation term. So, I am I have given you the definition of first variation, second variation. The first variation is linear in delta x. Second variation is quadratic in delta x.

Therefore, when you are talking about variational methods 3 d war 4 d war, we are interested in computing the increment suffered by the output resulting from increments the input. And that is where the notion of first variation second variation comes into being. And these are essentially. So, the variational calculus within this setup is derived out of the fundamental concepts that underlie Taylor series expansion. Here we are concerned with second up to second order in principle, I can also go up to kth order.

(Refer Slide Time: 43:41)



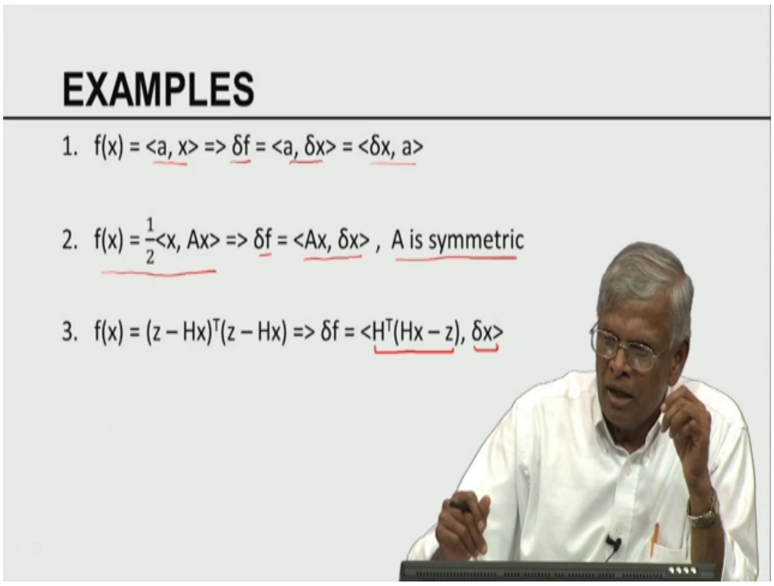
So, given this now I am going to give you formulas for computing the first variation much like I gave you formulas for computing the derivatives. So, this these are the tables of variational calculus just like tables of multivariate calculus tables of univariate calculus.

So, what are the differential coefficient of standard functions? What are the differential coefficients of standard function multivariate calculus? What all the first variation formulas for various cases and that is what I'm trying to. So, this these again help you to develop that skill to compute all these quantities which are fundamental to applications. So, let f of x be look at this now. X is an R n sorry, that is that is correct, x is in R n f is a vector of size m.

So, f is m vector n I m sorry, f is m vector x is n vector I'm now going to be concerned the first variation of delta f. So, the first variation of f is simply the first variation of f 1 f 2 f m. F 1 f 2 f m there are all independent. Compute the first variation f 1 compute the first variation f 2, stack them all together. F 1 you get you get the first variation of that. First variation f 1 is simply the inner product of gradient of f 1 with respect of delta x, gradient of f 2 with gradient of f m with delta. So, you get the formula and this delta x is a common factor.

And the resulting one is a matrix it can be very easily verified this is the Jacobean times this. So, the first variation of a function is related to the matrix vector product the matrix being the Jacobean the vector being the increment. So, this is a beautiful formula that we will use repeatedly in the derivations of 3 d war 4 d war. A that is the reason why they are called variational methods.

(Refer Slide Time: 45:54)



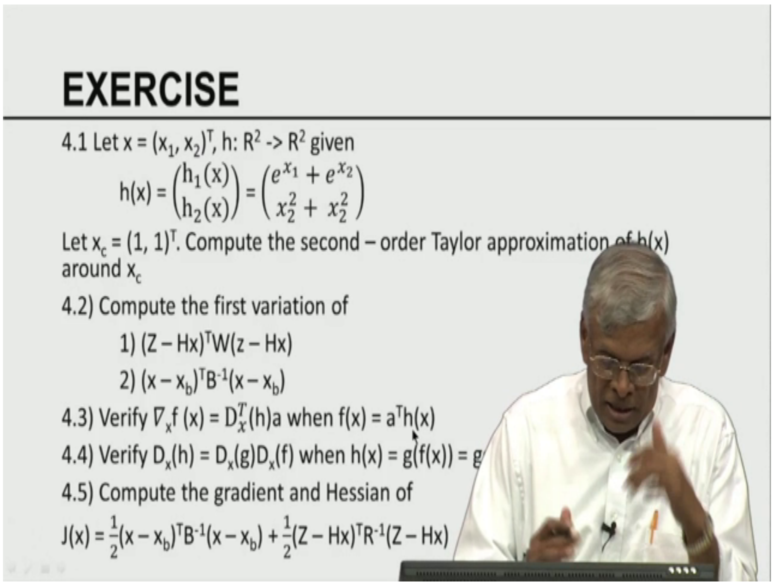
Here are some examples again. If f of x look at this now, I am using the same example through to compute the gradient to compute the hessian to compute the first variation to compute the second variation. The reason I am keeping the same example is because then you can see the interrelations between the gradient, and the variations I think it is it is the ability to knit together a picture how these relations are built is fundamental to a thorough understanding of what we are planning to achieve in this course.

So, if f of x is a scalar valued function of the vector. So, the first variation is simply the inner product of the first variation of a with respect to delta x. Again, this is the first formula in variational calculus. If f of x is equal to 1 half 2 of x transpose ax, delta f is equal to the inner product of ax with delta x here a symmetric; obviously, a symmetric because the quadratic. Form now the third example should be very familiar to all those who have done 3 d y z is the observation h of x is the model prediction z minus h of x is the error. So, this is the sum of the squared errors, this is the function that we are often using in least square methods to minimize.

So, given z given h we would like to minimize this with respect to x. So, this function is the cost function of the linear least square problem. So, I not to be able to compute, the solution for the linear least square problem I need to be able to compute. The first variation I need to be able to compute the gradient. I am giving the formula for the first variation of f is given by the inner product of this vector with that.

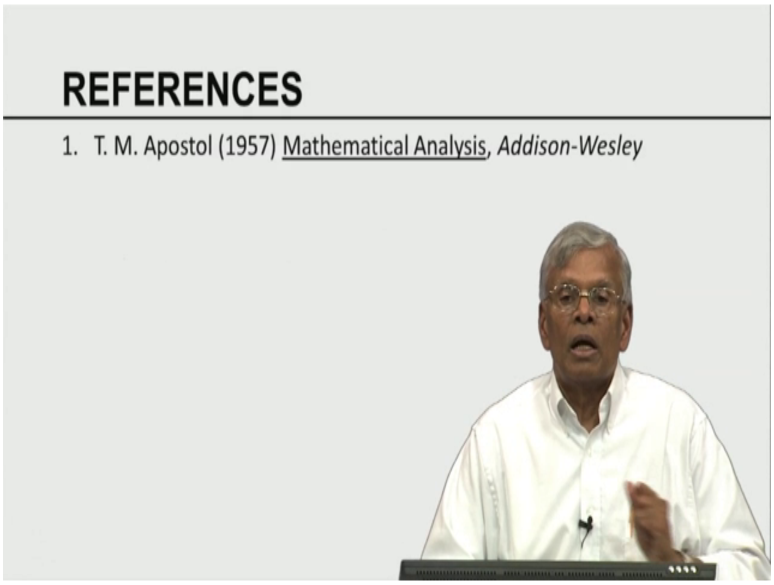
Again, I am I am these are simple exercises, but these are obvious. I did not go into the derivation of each of these things. I am trying to hit on various important themes. And you have to fill in the blanks for a thorough understanding of all these things. The aim of a course like this is not to provide all the details we will we will not be able to accomplish much if that properties. The aim is to be able to tell all the important concepts to see how things are knitted once you understand once you have a development bigger picture. Then you can the deeply into each of these. So, I would like you to be able to develop that begin deeply are as you go through the modules.

(Refer Slide Time: 48:30)



With this we come to the end of this part. I have given several exercise problems. These exercises are essentially extensions of the concepts we were we had talked about. You look at this as an example the first problem essentially asks you to compute the first order Taylor series and second order Taylor series, not for any arbitry h, but for a special h when the forms of h are given in a specific way. So, this will these are concrete examples if you did it you will have that final aha. Again, compute the first variations. Again, verify the different formulas that we have already talked about. So, doing this in in long hand in pencil paper, would help to complete the picture and here.

(Refer Slide Time: 49:22)



What is the standard reference for multivariate calculus? My favorite is a slightly older book, but it is still classic in my mind Apostle mathematical analysis. I have a copy of it. Whenever I get into difficulty which I do very often, I fall back on apostle. Apostle is a beautifully written book on multivariate calculus. With that we will conclude this discussion and overview of concepts and properties that are often used in from multivariate calculus in data assimilation.

Thank you.