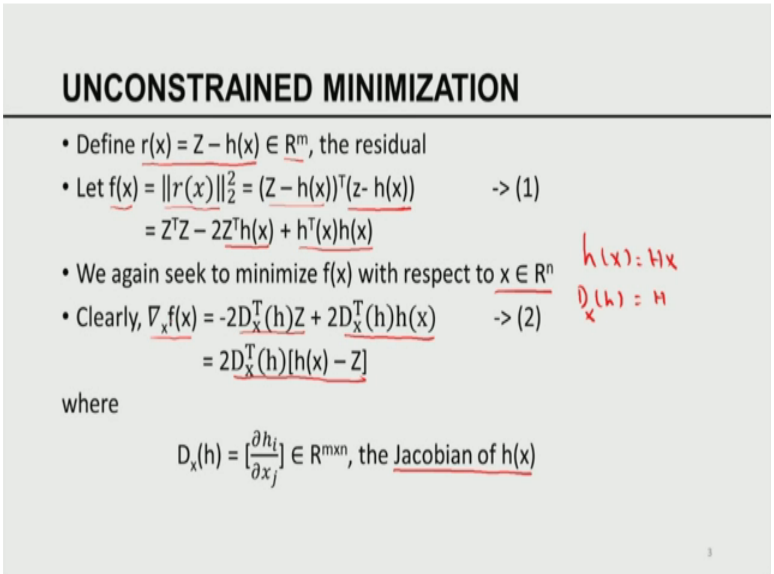
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture - 12
Deterministic, Static, non-linear Inverse Problems

In the previous modules we talked about static deterministic linear least square problems waited unwaited versions well posed and ill posed versions. We also saw the natural relation between the least square solutions and a geometrical interpretation of the least square solution and that theory was very beautiful in itself. But very seldom the problems that we come across in geosciences are linear some are linear some are made, but many of them are non-linear. So, in this module our aim is to be able to extend the concept of least square solutions to solve deterministic static non-linear inverse problems.

(Refer Slide Time: 01:10)



So, let h be a map that means, h is a vector valued function of a vector; x is a vector that gets into R m h of x that comes out is belonging to R m. So, you can see this is the vector valued function of a vector it is also called a map, in meteorological context in geophysical context is also called forward operator, it is a map from the model space. So, R n is the model space this is the observation space. So, h is a map from model space to the observation space, h has m components h 1, h 2, h m transpose whether x is where x is Rn.

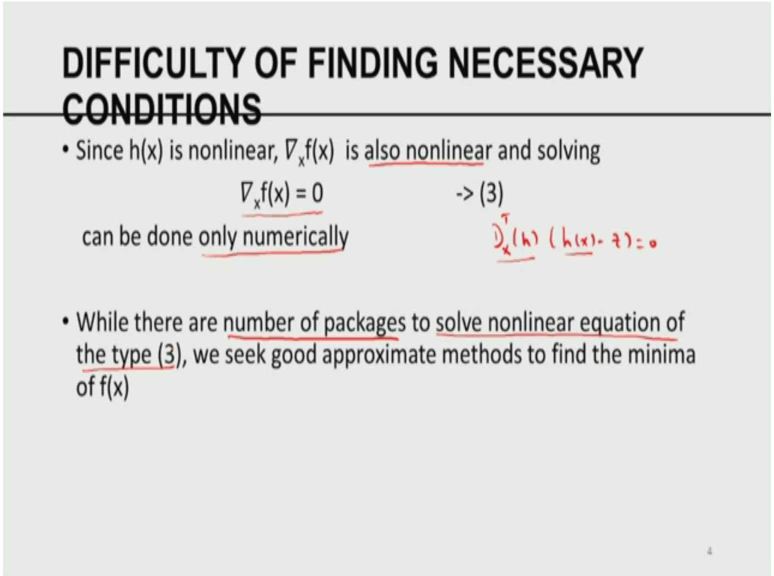
So, given z which is R (Refer Slide Time: 00:00), given the nature of the function h our problem is to be able to estimate x such that z is equal to h of x a that is the problem non-linear version of the least linear least square problem the linear problem what it that we said there is a matrix h, that goes from model space to the observation space. So, in that case we had the problem Z is equal to H of x, we have solved that problem now it takes the form Z is equal to H of x. This we have already done the our question is how to do these problems and that is our goal in this module. We are going to characterize this inverse problem again as an unconstrained minimization problem. Please remember that is exactly what we did when we did the linear least square problems.

So, we are going to follow the same track of ideas, the residual is z minus h of x, m is a vector x is a vector. So, you can think of the residual to be Z minus h of x and that is in R m. So, we can now concoct a function f of h which is the square of the norm of the residual. The square of the norm of the residual is given by z minus h x transpose times z minus h of x the only difference is instead of using a linear function I am using a non-linear function that is all about the differences if you multiply this and.

Simplify you get z transpose z minus 2 times z times transpose h x plus h x transpose h of x. So, that is a scalar each of this is a scalar function of the vector we again seek to minimize f of x with respect to x and there is no constraint on x x belongs to r and that is why is a unconstraint to minimization problem. A standard way to solve the unconstrained minimization problem is to be able to compute the gradient from the module on multivariate calculus we have already seen we have already computed gradient of terms like z transpose h which is equal to 2 times transposes the Jacobean of h times z, again from that module on multivariate calculus the gradient of h transpose h is transpose of the Jacobean of h times h of x, this can be succinctly written by 2 times the transpose of the Jacobian of h, h x minus z where please recall the Jacobian is simply a matrix of partial derivatives it is a m by n matrix Jacobian.

So, if h of x is equal to h times x linear the Jacobian of h is simply h. So, by specializing this we can readily get the least square counterpart of this. So, this is in that sense is a generalization.

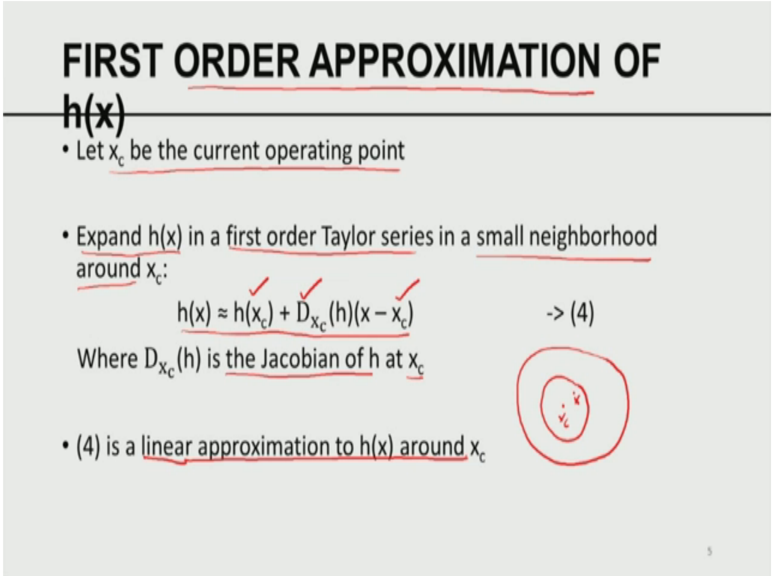
(Refer Slide Time: 05:48)



So, they have computed the derivative in order to be able to maximize in order to be able to minimize we want to be able to minimize f of x, please we want to be able to minimize f of x in order to be able to minimize f of x I have to be able to equate the gradient to 0. If I equated the gradient to 0 as you can readily see I get a non-linear equation where is the non-linear non-linear the equation comes from? There is a Jacobian of h times h of x minus. So, I will rewrite this now the solution is essentially given by Jacobian of x with respect to h, h x minus z must be equal to 0.

Now, h is non-linear, D h is also non-linear, product of 2 non-linear functions are non-linear therefore, the gradient is a non-linear function and we have to solve a non-linear system of equations. The only way to solve a general non-linear system of equations is to solve them numerically. So, one way to be able to solve the non-linear least square problem is to be able to compute the gradient and use the well established procedures from numerical analysis to be able to solve the system of equations that is one way. So, I would like to summarize by saying there are number of packages available, and these packages can be used to solve this type of equation 3, and using that we can find the minimum of f of x the solution of this equation 3 corresponds to satisfies the necessary condition for a minimum and then in order to be able to guarantee they are minimum we have to compute the hessian, we have to evaluate the hessian at the roots then we have to test whether the hessian are positive definite. Once the hessian are positive definite then we have minima in general a non-linear function may have many roots for this equation therefore, that could be multiple minima so, this going to be computationally a very challenging problem.

(Refer Slide Time: 08:15)



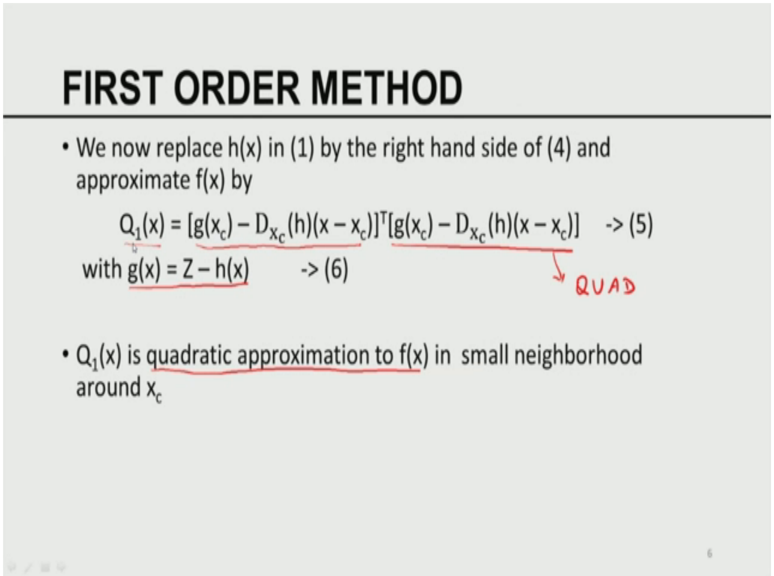
So, to get around that what is that we are going to be doing? We are going to be looking at an alternate method, we are going to be seeking good ways to approximate the non-linear least square problems, and that is what we are going to now describe. The approximation we are going to be talking about is called a first order approximation. First order approximation to the function f of x. So, what is the basic idea? Let us pretend I now know where to start the solution that is the current operating point. Generally engineers and scientists know the range within which the solution yes they may not know the exact solution, but it is supposed to be in this box are in fear.

So, x c current operating point is some point that we already know which is not too far from the solution. So, that is contingent on our prior knowledge of this problem now what do we do? We try to expand h of x in a first order Taylor series in a small neighbourhood around the point x a. Again going back to our module on multivariate calculus, I am going to be expanding h of x. So, you can think of the domain like this, this is the current operating point x of c. I am considering a small enough neighbourhood around this I am now considering a point x in a small neighbourhood around x of c.

So, x of c is the current operating point, x is the point I would like to move from x c to x. If x is close x c, I can express h of x by a first order Taylor series which we have already seen in our module on multivariate calculus. So, h x of c is the Jacobian of h at the point x c which is. So, I can. So, given h I can always compute the Jacobian if you give me x c I can evaluate the Jacobian matrix numerically. So, the numerical value of this Jacobian matrix is known, h of x c the value is known, x c is known. So, you can simply see is simply a function of h of x that is the vector is the vector function. Not only it is the function of x, you can also see it is the linear in x. So, it is a linear approximation to h of x around x c. So, what is that we have done? H of x is different globally, but I have replaced the global h of x by a local linear approximation using a first order Taylor series in a small neighbourhood around the current operating point.

So, that is the key to this argument. So, what is that we are going to do? Instead of solving the problem globally we are going to solve the problem locally and keep making local improvements with the hope that these local solutions and local improvements will ultimately eventually lead to the global solution. So, in here we have converted a non-linear problem to an associated linear problem by invoking to the first order Taylor series using Jacobian.

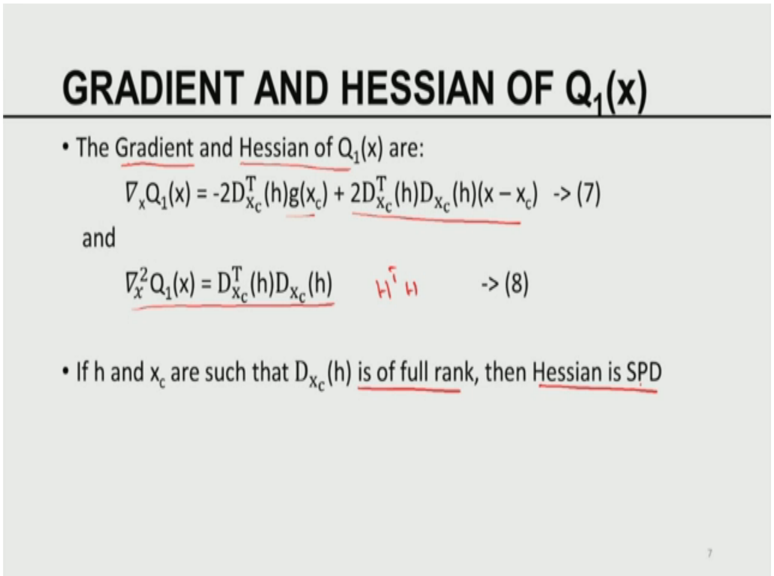
(Refer Slide Time: 11:51)



So, now if you go back to our previous slide, my f of x consists of h of x h of x. So, what is that we are going to do we are going to replace these h of x in equation one I will come back here. So, now, replace h of x in equation one by the right hand side of 4 the right hand side of 4 is simply a linear approximation to h of x. So, if I did that, I am going to get a function Q 1 of x. Q 1 of x is an approximation to my f of x in a small neighbourhood. So, f of x is a global function Q 1 of x is a local approximation to the global function. You can readily see Q 1 of x is given by a linear part and a linear part multiplied together. So, this is a quadratic approximation. So, this is a quadratic approximation.

Where g of x is equal to g minus h of x instead of writing z minus h of x c, I am. So, g of x c will be h of x c. So, g is a change of variable for z minus h of x. So, this quadratic approximation of f of x in a small neighborhood around x of c, what is that we are going to now look at? We are going to be looking for minimizing Q 1 of x.

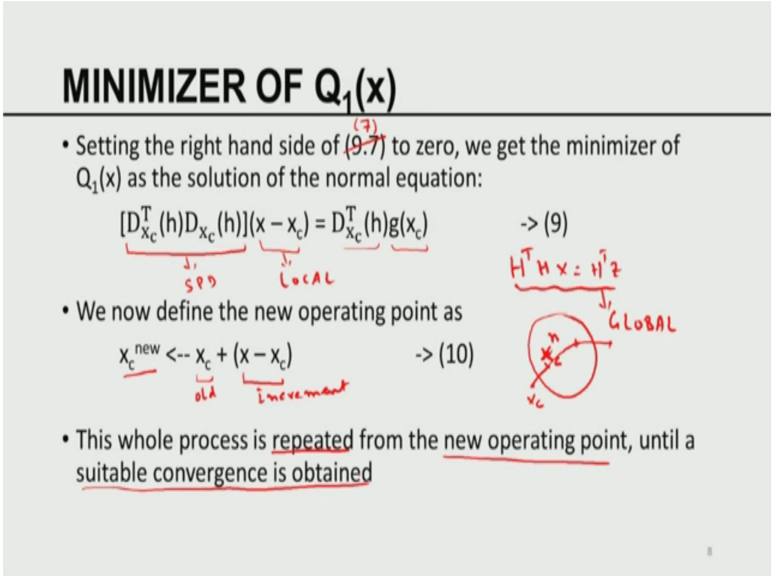
(Refer Slide Time: 13:34)



If I want to be able to minimize Q 1 of x, I can readily compute the gradient; I can compute the hessian of Q 1 of x again by the results in the module relating to multivariate linear algebra, the gradient of Q 1 is given by these expressions the hessian is given by this expression.

Now, please recognize the hessian is the transpose of the Jacobian times, the Jacobian evaluated the point c this looks like H transpose H. So, there the inverse of it exists if h is full rank, here the inverse of it exists if the Jacobian is of full rank. So, if the Jacobian is of full rank, then the hessian is positive semi definite. If they have Jacobian is full rank and the hessian and the hessian is positive semi definite the equation to see the that the solution of the equation the gradient setting at 0 in 7 must give the optimal or the minimum solution.

(Refer Slide Time: 14:43)

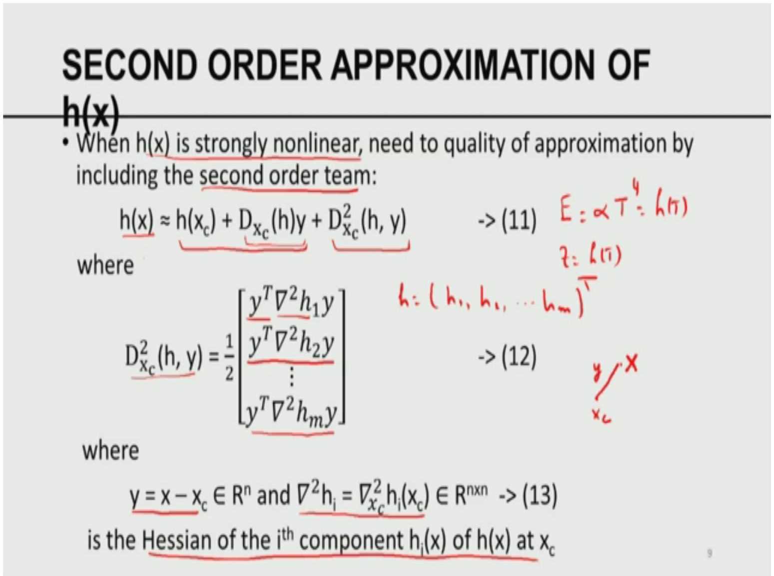


Therefore by setting the right hand side of se 7 equal to 0, I do not have yeah I do not have to say 9.7, it is simply 7. By equating the right hand side of 7 to 0, we get the minimizer of Q 1 as a solution to the normal equation. The solution to the normal equation is given by this matrix times this is equal to this matrix times g. Now I would like you to look at the structure, this structure is very similar to H transpose H x is equal to H transpose Z that we saw in the linear case. There we got the global solution here I am getting the local solution.

So, by solving this and this matrix is symmetric and positive definite. So, it is non singular. So, I can express x minus x c is equal to the inverse of this matrix times the right hand side. So, I am by solving this I am going to get x minus x c, I originally got x c. So, originally I started with the point x c, I went to a point x. So, if I add these 2 together I go from x c to x the optimal x that minimizes the linear approximation. So, my new operating point is equal to the old operating point. So, this is the old operating point plus the increment, that gives you a new operating point then I repeat the same show around the new operating point which is x c new.

So, now I will consider another small neighbourhood around this new neighbourhood I will consider another x, I will go from here to here. So, I went from here to here then here to here, then here there to there. So, by moving sequentially from operating point operating point operating point, I am moving towards the global solution this whole process is repeated from the new operating point, until a suitable convergence is applied. So, what is the key here? I am converting the difficult problem of solving a non-linear least square problems globally, to a sequence of simple linear least square problems we already know how to solve linear least square problems. So, the trick is if you know how to solve one problems very well, I can convert other problems to one that I know how to solve and using the algorithms for solving the linear problem, I can continue to solve the non-linear problems iteratively. So, you can see the difficulty of non-linear problems essentially comes from our inability to look at the global solution at one juncture in one shot, I am trying to build the global solution by sequence of local solutions that is what we saw the first order approximation.

(Refer Slide Time: 18:27)



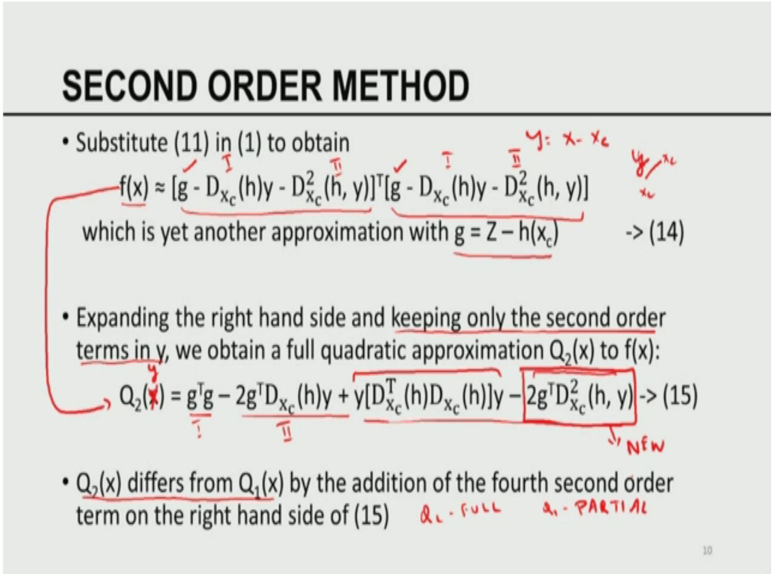
Now, I am going to go to the second order approximation just to be able to tell you if my function h of x is strongly non-linear, what does it mean if it involves logarithmic functions, exponential functions, trigonometric functions or fractional powers of different quantities of interest. For example, what is one typical non-linear function? In the case of satellite meteorology, the energy radiated is equal to alpha to the power T to the power of 4 the temperature that is a very strongly non-linear function. The energy radiated you see is proportional to the fourth power of the temperature.

So, if I Z in this case Z is equal to. So, this is equal to h of T. So, this is equal to h of T, T is the state variable. So, x plays the role of T and Z is equal to h of T in this case very strongly non-linear. In this case linear approximation will not come very very handy linear approximation has lot of errors, in order to be able to improve the quality of approximation, I am now going to go from first order to second order terms. Again you can readily see all these things are related to first order Taylor series, second order Taylor series for vector valued function of vector is all the things that we have already covered in the module on multivariate calculus. So, in this case in addition to the first order term, I am going to have a second order term. The second term additional second order term improves accuracy of the Taylor series approximation, the second order term depends on the hessian.

So, this second order term is given by a vector please understand h is the vector, h of x c is the vector, this D of h D of h of x c is a Jacobian matrix evaluated x c; y is the vector. So, this is a vector this is again a vector, the vector is given by now look at the following now you already know h is a function, which it which consists of h 1, h 2, h m transpose. So, you take h 1 compute it is hessian that is the matrix, that is a quadratic form with the hessian of h 1 this is the quadratic form with the hessian of h 2 this is a quadratic form with respect to hessian of h m, 1 over half the half comes from the second order term Taylor series coefficient. So, this whole vector will go in here and how do we express this? I would like you to be able to think of it like this, this is x of c, this is x the difference between them is y is y. So, y is the distance between the current operating point and any other point x in the neighbourhood of it.

So, y is equal to x minus x c that is a vector that is a vector in R n, del square h of i is the hessian of the i th component h i. So, the hessian of the ith component of h i, with this I get a reasonably good expression for the second order Taylor series for h of x and that is given by 11. Look at the notation the notation could be a little complicated for some of us who are not familiar with dealing with ah Taylor series expansion. So, it is very imperative we understand the Taylor series expansion for multi vector valued functions of vectors, in order to be able to get the complete total understanding in here what is happening around the current operating point x c.

(Refer Slide Time: 22:32)



Now, what do we do? So, we are do going to do the same thing. So, I got an approximation for h of x in 11. So, I am going to substitute 11 in 1 to be able to obtain, a new approximation for h of x again we are dealing with approximation. So, this is one term this is another term, now what is y? Please remind yourself y is equal to x minus x c x c is known. So, I can recover x if I know y.

So, I am simply talking about the increment y. So, so f of x is now expressed in terms of y again g has a previous value of Z as Z minus h of Z. So, expanding the right hand side; the right hand side has now 3 terms if you multiply the whole thing, you get an approximation in terms of I would like to call it y sorry that is not x it should be y. So, we expand the term if you expand the term look at this now this is a the g is a constant term this is the first order term, this is the second order term in y, this is a constant term, this is a first order term, this is a second order term.

So, each one of the factors are quadratic functions if you multiply 2 quadratic functions, you are going to get a fourth degree term in the components of y. So, what do we do? We expand, but keep only the second order terms in y. So, what is mean what do you mean by second order terms? Second order or the second degree in the components of y, there will be third degree term there will be fourth degree term we are going to neglect the third degree term and the fourth degree term why we are allowed to neglect the third and

[redacted]

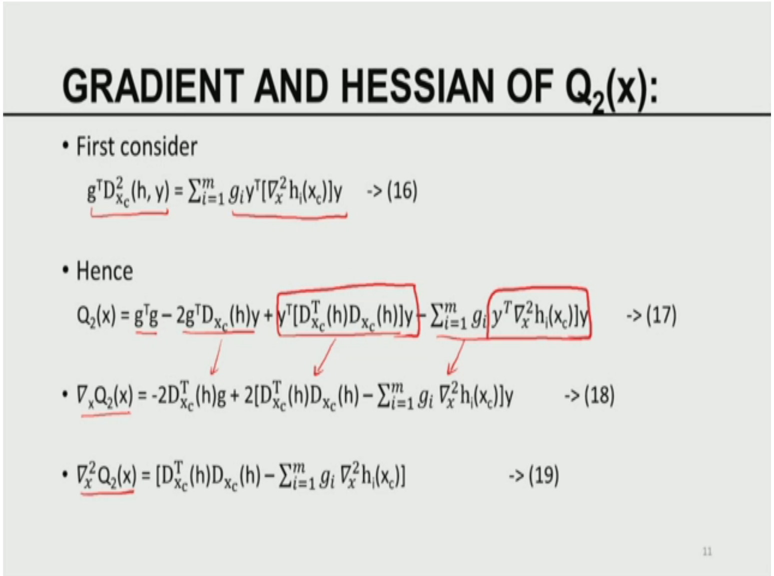
fourth degree term. If x c goes to x c h is y is small. If y is small y square is smaller y cube is even smaller y to the power 4 is even much smaller.

[redacted]

So, we are simply invoking to the order of magnitude scaling passes involve in here. So, we are only going to keep the dominant term up to the second order we believe the third order term and fourth order term are essentially very small. So, by keeping only the second order approximation, I get an approximation of f of x as Q 2 of y. Q 2 of y is given by these terms I would like to look at this term for a minute. This term is of degree 0 the first term is of degree 0, the second term is a linear term the third term is essentially a quadratic term look at this now, the third term this is quadratic in y and by the definition of second order term that is also quadratic in y, the sum of 2 quadratic terms is a quadratic terms. So, Q 2 of y is quadratic, Q 1 of y it was also quadratic, but in this in the case of Q 1 of y I did not have this term. So, this is the new term that comes into play if I use the second order approximation

So, this is new. So, that is why we are going to call Q 2 as a full quadratic approximation and Q 1 as only a partial quadratic approximation that is simply a mathematical fact that comes out of this analysis. So, if you drop the second order term, you Q 2 becomes equal to Q 1. So, that is the important nesting that we have to look at that. So, quadratic approximation is obtained by simply adding a second order term, which is the last term in equation 15. So, that is a summarise in the following discussion, Q 2 before from q I with the addition of the fourth which is the second order term on the right hand side a 15. So, I had f of x which f of x by a second order Taylor series, I substituted that I will get second degree term, third degree term, fourth degree term we dropped that every term larger than the second degree. So, Q 2 y is the total or full quadratic approximation of f of x.

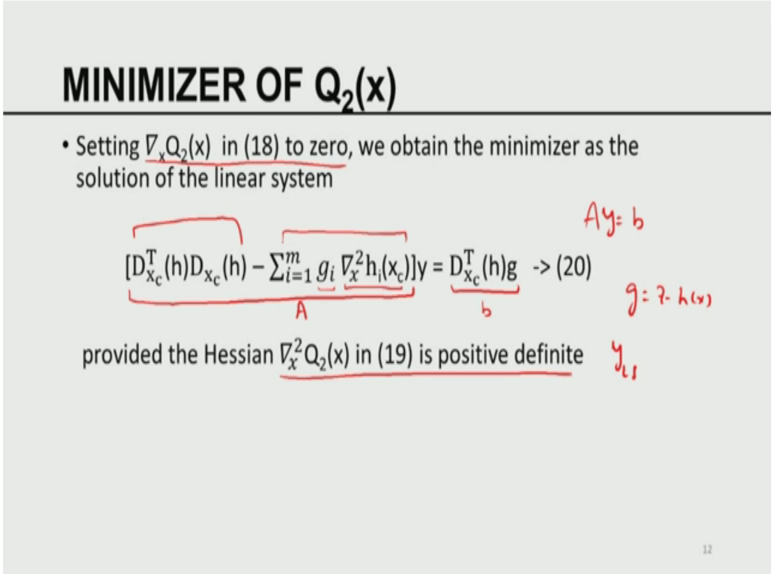
(Refer Slide Time: 27:20)



Now, the problem becomes very simple, I have a quadratic function, I would like to be able to compute that del or the gradient of the this quadratic function in 15 and that I do in stages I am considering Q 2 g transpose D square h of y this is the term that we have added is a new term that comes in to Q 2 that was not in Q 1, this term if I expand it. So, g is a vector D square is a vector; I am talking about the inner product of 2 vectors. The inner product of 2 vectors is given by the sum of g I times the quadratics formula; hence Q 2 of y now can be replaced by this constant term first degree term one second degree term and this second degree term. So, you can readily see the quadratic function coming in here.

You can also see the quadratic function coming in here, these 2 are quadratic that is the linear therefore, I can compute the gradient of Q 2 again we are invoking to the module on multivariate calculus; the gradient of this gives rise to this, the gradient of this gives rise to this, the gradient of this gives rise to this, and I am going to equate this gradient to 0 to get my solution, I also get the hessian of Q 2 this is the gradient of Q 2. So, I have computed all the required quantities in order to be able to compute the solution I simply need to be able to set the gradient to 0.

(Refer Slide Time: 29:00)

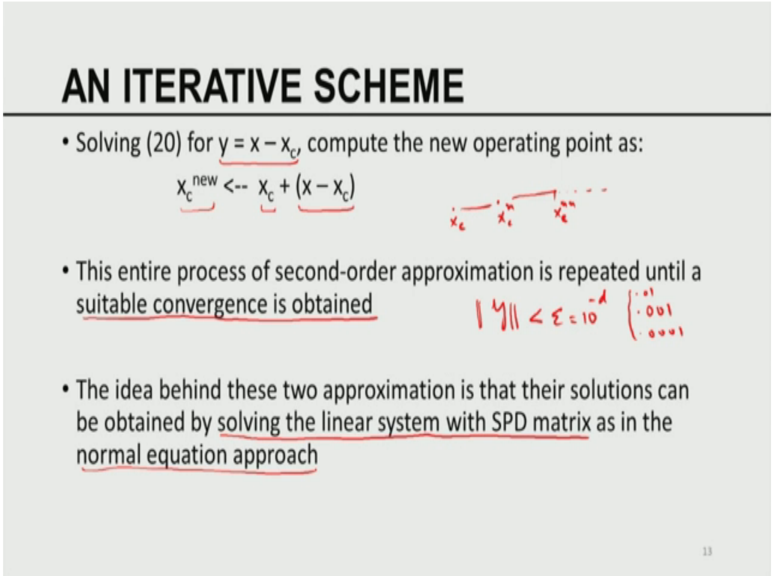


I am going to have to set the gradient to 0; the setting the gradient to 0 gives rise to a linear system where the system matrix is given by this. So, this is like A y is equal to b where this is the matrix A, this whole thing is the vector b and I am going to solve for y.

Now, I would like to look at this matrix. So, this is the Gramian that comes out of the Jacobian. This is the term that comes out of the hessians of the components of that g i is are the constants the g is as your if you are if you recall g is equal to z minus h of x. So, g i is our constant. So, this is each one of these are matrices. So, this is the linear combination of matrices multiplied by g i. So, this is the matrix this is the matrix, I can solve this matrix equation. The solution of this matrix equation is going to give me a y least square, that least square solution is it will indeed be if that solution will indeed be a least square solution provided this hessian term is positive definite.

So, this is essentially another way of looking at the approximations to the non-linear problem using second order approximations.

(Refer Slide Time: 30:42)



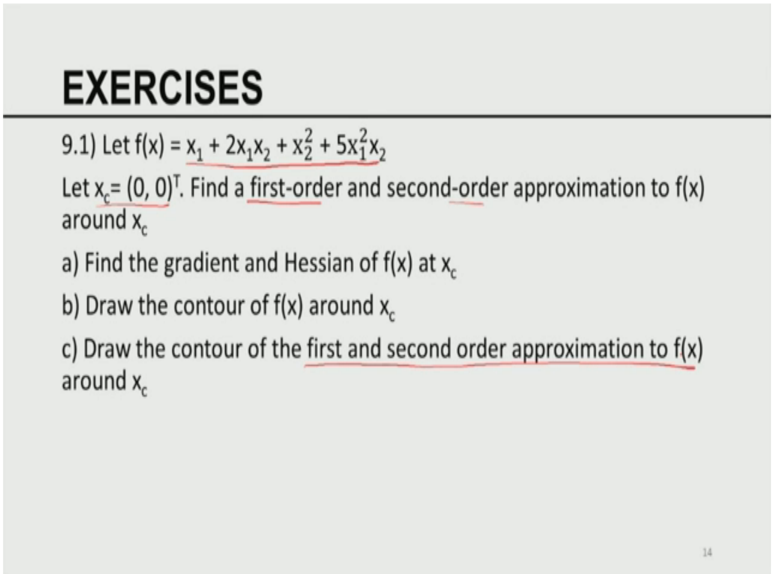
Again we are going to solve 20 for y, if I solve for y please remember y is equal to x minus x c. Once we have x minus x c we can compute, we can add x minus x c to y that gives you a new operating points. So, here again I am going to go sequentially I have x c then I am going to x c new operating point, from here I am going to go to another new operating point we can call it new new and the conversion goes on. So, I am solving a sequence of local minimization problems by using clever partial quadratic form approximations are full quadratic form approximation to the function f of x.

So, the entire process of second approximation is repeated, until a suitable convergence is obtained. How do I say is suitable convergence? When do I say the convergence has occurred, if the norm of y you compute y if the norm of y is less than a pre specified epsilon and what is that epsilon? Epsilon could be 10 to the power of minus d and what could be that? That could be you can set the criterion any way you want 0 0 1 or 0 0 0 1 or 0 1 these are all typical values of epsilon one could utilize and. So, if d is large your approximation is better. If d is small the approximation is screwed. In some problems if the model is not a perfect model it is not worth worrying about exact solutions you can afford to ah get reasonably ah good neighbourhood, but not nearly exact d could be need not be too large in such cases.

So, it all depends on how well you believe your model is, how well you believe your method should be your method need not be more accurate than the model. So, more accurate solutions are needed only when the model is more precise. I wanted to be able to think of this consistency between goodness of the model versus goodness the solution that you obtain by virtue of data simulation. So, the idea behind these 2 approximations is that their solution can be obtained by solving linear system with SPD matrices, using the normal equation approach. So, what is the basic idea here? We developed the expertise in solving linear least square problems, once we have a good expertise in solving linear least square problem we are readily extending that expertise to solve non-linear least square problems by approximating the non-linear function using either the first order Taylor series are the second order Taylor series. So, that is the idea.

So, understanding linear least square problem is fundamental and if you do and if you understand it very well if you have developed programs to solve the linear least square problems, you can readily apply them to solving non-linear problems, but non-linear problems are not solved in one shot they are solved repeatedly. So, it is a sequential approach to solving non-linear problems using a sequence of linear problems.

(Refer Slide Time: 34:08)



With this we come to the end of this module, there are a couple of different exercises. These are very useful exercises, I want to emphasize couple of these f of x is a very simple function I would like you to consider x of c as a starting point, compute the first order and the second order approximations for f of x around x of c, find the gradient in a hessian of f of x at that point, draw the contours of f of x around x of c, also draw the contours of the first and second order approximations around x of c, you can see how these contours approximate the solutions to the problem as we progress from one operating point to another operating point, another operating point the you can understand and appreciate the progress of the local solution towards the global solutions. With that we come to the end of the discussion of solving non-linear least squares problems.

Thank you.