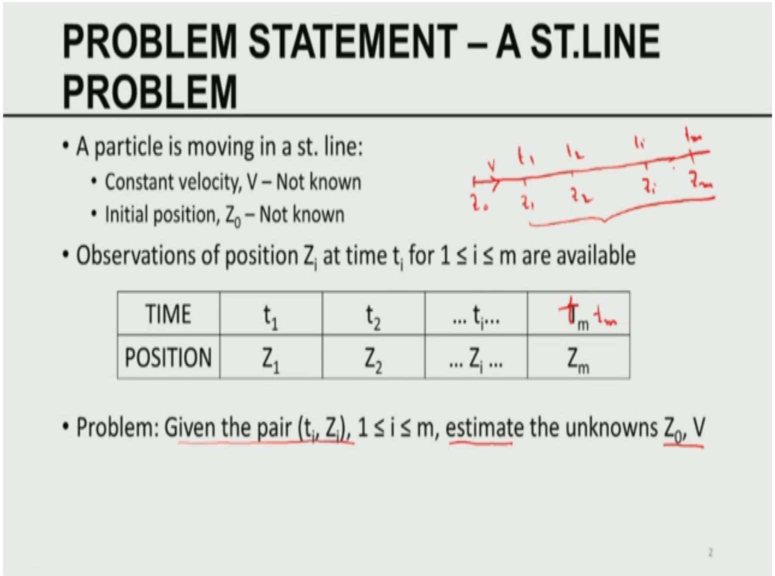
Dynamic Data Assimilation
Prof. S Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture -09
Deterministic, Static, linear Inverse (well-posed) Problems

In this lecture, we are going to start the discussion of solving inverse problems. We are doing this for the first time in this lecture series. So, to get started I am going to consider the simplest version namely, static, deterministic, linear inverse problem. We also have attached a qualifier well posed problem. We will describe what a well posed problem is as we develop the details of the statement of the problem.

(Refer Slide Time: 00:55)



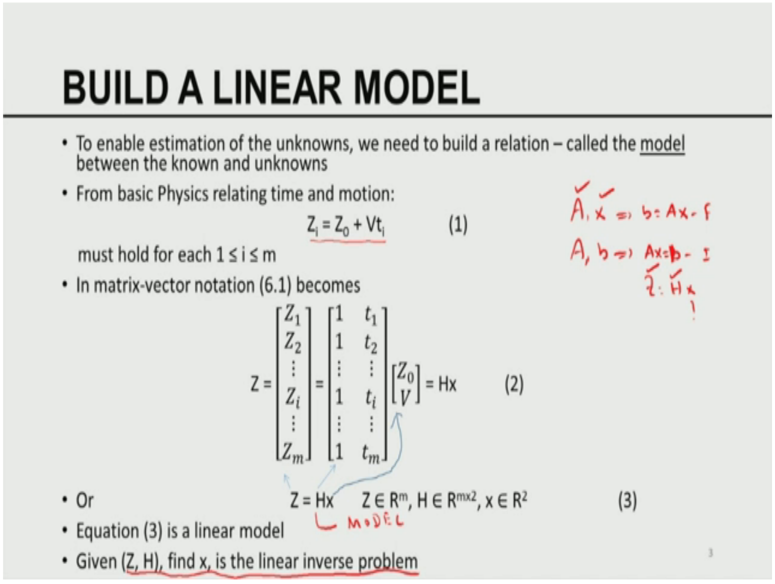
I would like to start by describing an example. This is called a straight line problem. Suppose, a particle is moving in a straight line the particle with the moving at the velocity V it started an initial position Z naught. We do not know where it started, we do not have Z naught, we do not know what the velocity of the motion is. We can only observe the position Z i of the particle at time t i. Let us assume, we are going to measure the position of the particle at times t 1, t 2, t 3, t m, where t 1 is less than t 2 less then t 3 less than t i less than T m. In this case the t here must be a lower case t, a lower case t m.

So, let the particle pass through the position Z 1 at time t 1, let the particle pass through the position t 2 at time at time t 2, Z i at time t i and Z m at time t m. So, what is the statement? The statement is the following; we have a set of observations of the time and position the pair t i Z i for i is equal to 1 t m, in other words we have m pairs of time versus position. This is the position at which the particle appears at the Z i is the position at which the particle appears at time t i. So, knowing t i Z i for i running from 1 t m, our aim is to estimate the unknown Z naught and V. So, this is the data that is what we need to find.

So, you can conjure up the particles like this, it is moving in a straight line. It started at the position Z naught, it is travelling at the velocity V, at this is Z 1, this is Z 2, this is Z i, this is Z m, this position is t 1, t 2, t i, t m. We can only observe the position of the particle at various times. We do not know the velocity, we do not know the position it started, and we would like to be able to estimate the velocity and the position knowing a bunch of m observations; that is the problem.

Why this is called an inverse problem? Let us talk about it in a moment.

(Refer Slide Time: 03:46)



In order to be able to formulate the inverse problem, I need a mathematical model. The mathematical model is one that relates the known to the unknowns. In this case the unknowns are Z naught and V. The knowns are Z i and t i; the model is now based on simple basic physics. From basic physics, we know that if a particle started at position Z naught and travelling at a constant velocity V, the position Z i, that it would be at time t i is given by the simple relation from basic fundamental physics. So, this relation relates the unknown Z naught and V to the known Z i and t i. So, I have m values, I have m equations like this. So, this I can rewrite the vector matrix notation. So, Z 1 is equal to Z 1, Z 2, Z i, Z m are vector. Z is the vector of all positions at m different times, Z naught and V are the unknowns.

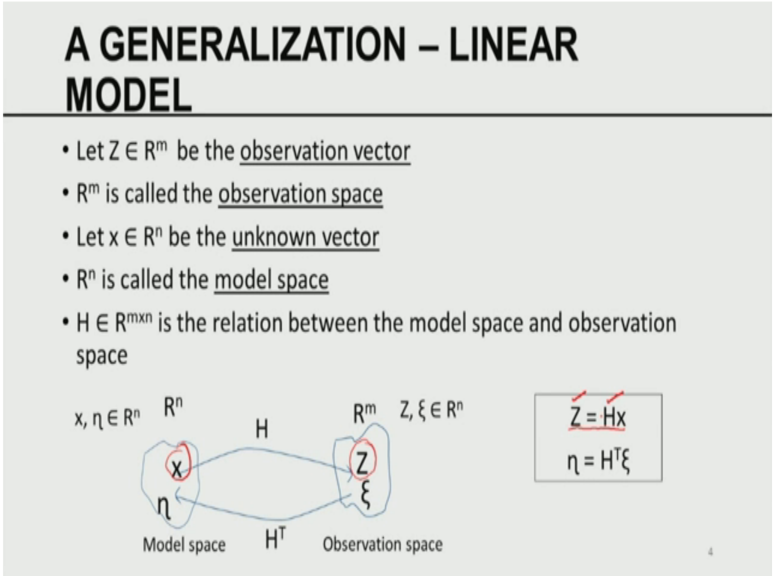
By using matrix vector multiply you can see if I multiply the first row by this column I get Z naught plus Vt 1, Z naught plus Vt 2, Z naught plus Vt i, Z naught plus Vt m, each one of them corresponds to the position at various times. So, I call the vector Z 1 to Z m as Z, I call this matrix with the first column 1, second column t 1, t 2, t m as h, I call the vector Z naught and V as x. So, x, the components of x there are 2 components; first component is Z naught, second component is V. The unknown is the vector of size 2, the known is a vector of size m, H is a matrix; H is m by 2 matrix. It has m rows and 2 columns.

So, this problem can be stated as Z is equal to H of x. So, Z is equal to H of x, is the mathematical relation that simultaneously captures all the positions there are observed at m different times. So, Z is a m vector, H is a matrix which is m by 2, x is a vector which is R 2. Now, please go back to our definition of direct problem and inverse problem. Given A, given b ah I am sorry, given A let me erase. The given A, given X, computing b is equals to AX, that is, the when these 2 are given computing this that is the forward problem. Given A, given b computing the solution AX is equal to b, that is the inverse problem, we have already seen that in the several classes.

Therefore, here this problem Z is equal to H of x; H is known, Z is known, I need to find x. So, this problem is an inverse problem in the sense of the inverse problem that we have talked about. So, given Z and H, find x. This is an example of a linear problem, is an example of an inverse problem, is an example of a linear inverse problem. The unknown is x, the unknown does not vary in time, because V is constant the position where it the particle started is also a constant. So, it is a static problem. The relation Z is equal to H x is the model equation. This model is a static model. Therefore, we want to term this as static, deterministic, linear inverse problem. This is the simplest of the problem that one could formulate.

So, what does it is tell you? Based on a bunch of observations, I do a mining, I build a data, I build a model from the data. The mining rule that helps us to build the model is the basic relation in physics, Newtonian laws. So, using the Newton law I fit a model. Once I have a model, I know what are knowns, what are unknowns. It turns out this problem as a problem is an inverse problem.

(Refer Slide Time: 08:21)

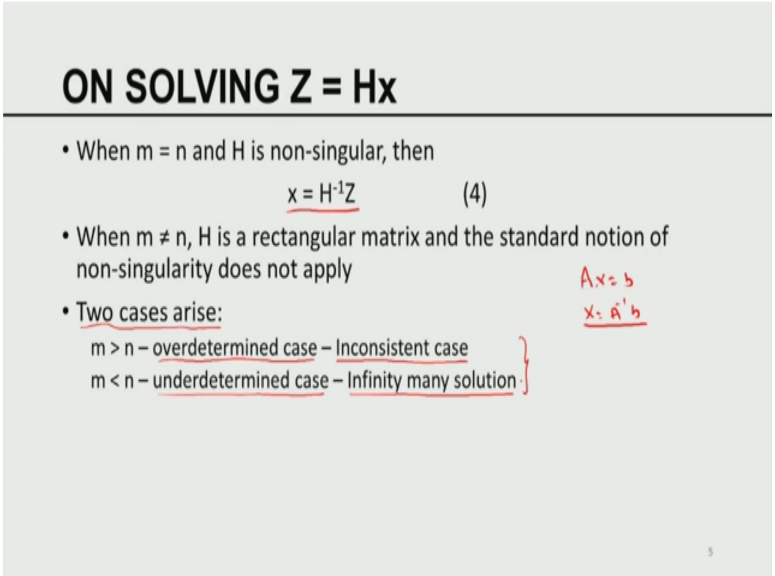


So, let us create some nomenclature. Z belongs to R m. R m is a set of all vectors of size m. So, that is what is called the observation space in here, you can see the observation space R m. So, Z is called the observation vector, R m is called the observation space. Likewise, x is called the unknown vector.

In the previous case x has 2 components. I can generalize that to n components. So, if the unknown x is going to have x 1 through x n, also R n is a model space, H matrix that R of m cross n is the relation between the model space in the observation space. So, if I have x in the model space, my H maps the x into Z. This relation between the model variable and the observation variable is given by the matrix H. So, H is the known matrix.

So, let us come in here. So, Z is equal to H of x, Z is known, H is known. So, if I generalize the particle moving in a straight line as an inverse problem the general linear static, deterministic, inverse problem is to solve given Z, given H find the x such that Z is equal to H of x. This is the first statement of the inverse problem based on a very simple problem in physics.

(Refer Slide Time: 09:55)



On methods of solving Z equals to H of x. If m is equal to n and H is non singular, from matrix theory we already know x can be written as H inverse Z, but in inverse problems solve them is the case when m equals to n. In the case of particle moving in a straight line the n was 2, m could be many. It could be less than 1, less than 2, it could be equal to 2, it could be greater than 2. So, we need to be able to consider a general case.

So, in general, H is a rectangular matrix. H is a matrix of size m by n, m need not be equal to n. So, the standard notion of singularity and non singularity the matrix is an attribute of a square matrix. There is no concept of singular or non singular, rectangular matrices. So, when there is no notion of singular, non singular rectangular matrices, I cannot even define when the solution exists and so on.

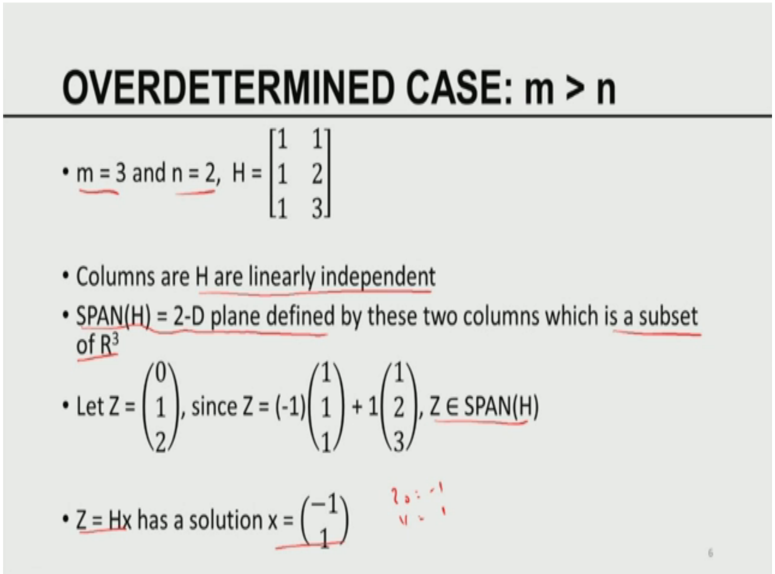
So, we need to consider a case which is harder than solving linear system A x is equal to b. So, solving linear system A x is equal to b, when A is n by n matrix, b is a n by 1 vector, when A is non singular I simply write x is equal to A inverse b; somehow, that I cannot do because A in this case is H. H is not a square matrix, I do not have even the concept of non singularity of rectangular matrix.

So, this problem, this linear inverse problem even though it is the simplest problem it does not fit in some of the standard problem that we study in linear algebra. So, we need to do develop a theory far beyond what the first course in linear algebra teaches us. In order to examine the solution concept for this we it is useful to define 2 cases, when m is greater than n, when m is less than n. Please remember; m is the number of observation, n is the number of unknowns. So, if m is greater than n, it is called an over determined system, when m is less than n it is called an underdetermined system.

We are going to show that in the over determined system the system is inconsistent. What does it mean? There is no solution for this problem; the simple is the system is inconsistent. In the case of underdetermined problem, there is no one solution there is infinitely many solutions.

But, in the case of A x is equal to b when a is non singular there is a single unique solution. So, we are now dealing with the problem that does not have that may not have a solution or that may have either infinite solution. So, if these are the 2 classes of problem that linear inverse problem gives rise to. So, linear inverse problems are more difficult than solving linear systems.

(Refer Slide Time: 13:03)



So, let us consider a over determined case to examine why this kinds the system is can be inconsistent. So, let us take an example of m is equal to 3 and n is equal to 2. Let us consider a case of H. The first column is 1, 1, 1; second column is 1, 2, 3. What does it mean? t 1 is 1, t 2 is 2, t 3 is 3. We can think of the particle moving in a straight line I am observing at time 1, 2 and 3.

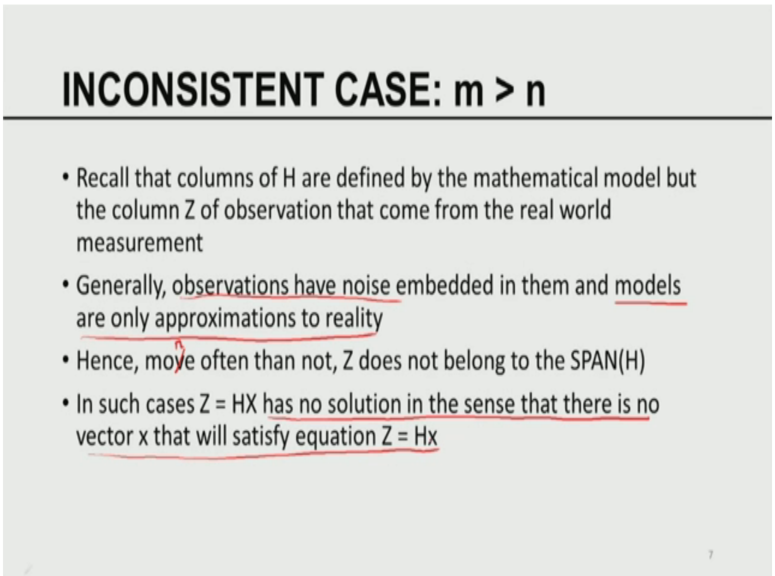
In this case if I look at the column the first column is 1, second columns 1, 2, 3. There are only 2 columns; these 2 columns are linearly independent. Why? No one column can be express as a multiple of the other. So, in here the columns of H are linearly independent; that means, if columns of H are linearly independent I can consider the span of H, span of the columns of H.

Please recall; in the module on finite dimension vector space we have defined the span to be the set of all linear combinations of vectors. Here the vectors are columns of H. So, span of H is equal to in this case these 2 vectors are linearly independent. So, 2 vectors each of size 3, define a plane. So, if this defines a plane which is a subset of R 3. So, R 3 is a 3 dimensional space, the span of the columns of H defines a plane embedded within that 3 dimensional space. It is in this space, we have to do perform certain computations.

Now, let us consider I have an observation which is 0, 1 and 2. Since, this vector Z can be expressed as minus 1 times the first column plus 1 times the second column. We can see Z can be expressed as a linear combinations of the 2 columns; that means, Z belongs to the span of H. If Z belongs the span of H, the solution Z is equal to H of x a unique solution. So, Z naught is equal to minus 1, V is equal to plus 1.

So, this is a case where I can solve an over determined system, but solves them such a case arises in practice.

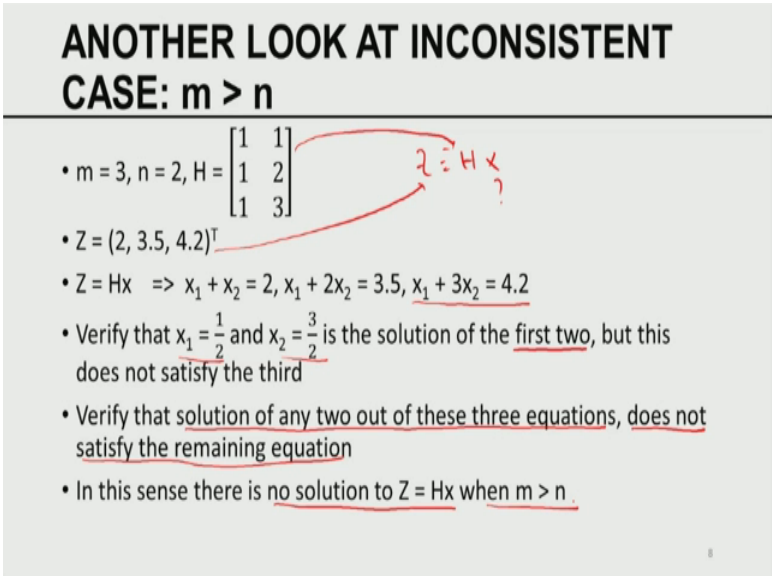
(Refer Slide Time: 15:36)



Recall that the columns of H are defined by the mathematical model. The column H comes from the basic physics equation, but Z is a column of observations that come from the real world measurements. The mathematical model describes the real world, but the reality is given by observation. Generally, when we say observations also have noise in embedded in them. They are corrupted by noise, observations have noise embedded in them, and so, there are 2 things. Observations always have noise and models are always only approximations of reality.

So, these are 2 fundamental facts. Hence, more often than not, so, this is more often than not. Hence more often than not, Z does not belong to the span of H. If Z does not belong to the span of H then there is no solution in the sense that there is a vector x that will satisfy the equation Z is equal to H of x. Therefore, in principle when m is greater than n the equations are inconsistent. Inconsistent means what? There is no solution that can make the left hand side is equal to the right hand side.

(Refer Slide Time: 17:03)



Let us take another look at the inconsistent case by giving little bit more specific example. I consider the same H, but I consider a vector for observation which is slightly different from the one that they had. Previously, I had observation 0, 1, 2; now, I am going to pick an observation 2, 3.5, 4.2.

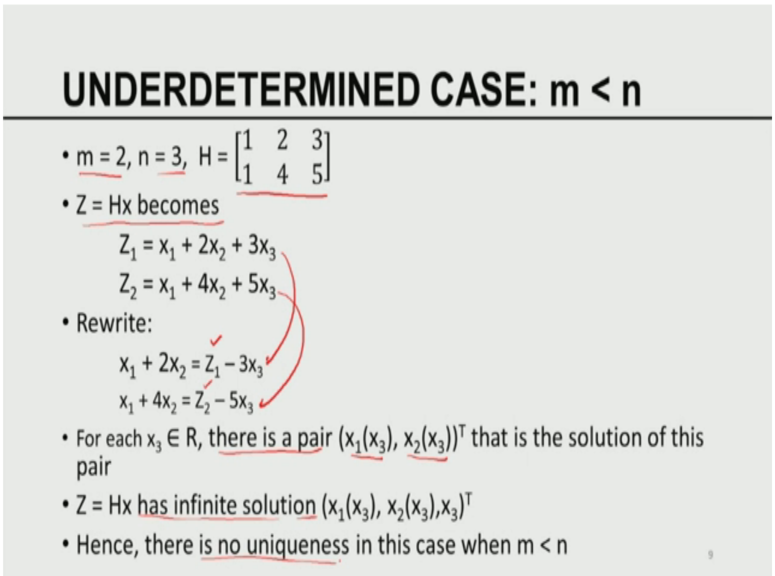
It could a current practice we should all offer this possibility. So, I would like to ask myself a question, does there exist an x such that Z is equal to H of x, when H is given by this and Z is given by this.

So, I want to ask myself the question does there exist a H. Let us explore this little bit further. So, the first equation tells you x 1 plus x 2 must be 2, second equation tells you x 1 plus 2x 2 must be 3.5, the third equation tells you x 1 plus 3x 2 is equal to 4.2.

So, if I pick the first 2 equation, let us consider the first 2 equations. I have 2 equations, 2 unknowns. If I solve the 2, I get the x 1 is equal to 1 half, x 2 is equal to 3 by 2. But, this solution or the first 2 does not clearly satisfy the third one. So, if you talk any subset of 2 equations and solve them and substitute in the third, the third is not satisfied. So, this is true whether you solve 1 and 2, 1 and 3 or 2 and 3.

Verify the solution of any 2 of these 3 equation does not satisfy the remaining equations, that is an important thing. So, in this sense there is no solution to Z is equals to H of x when m is greater than n; that means, in the case of over determined system when I have more. So, what do you mean by over determine system n is the number of unknowns to be estimated, m is the number of knowns. If the number of knowns m is larger than the number of unknowns n the system is over determined, in this case the system may not have a solution. So, that is a difficult situation to be m.

(Refer Slide Time: 19:12)

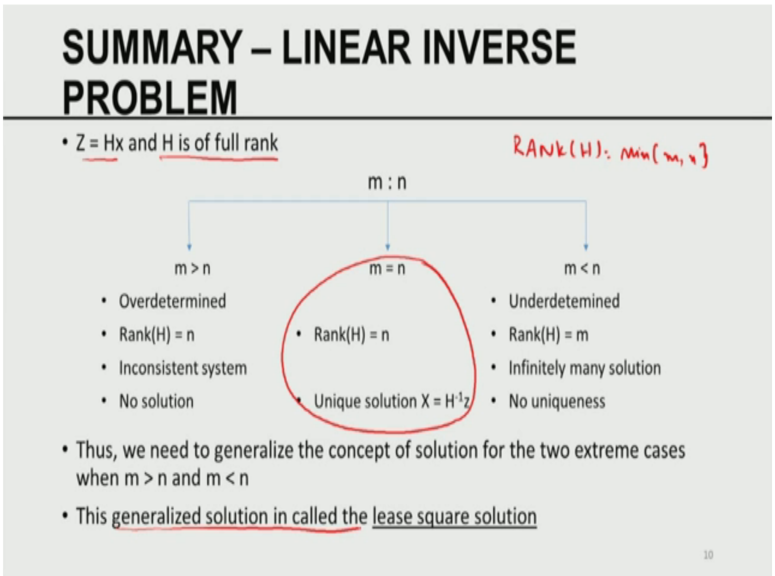


Now, let us worry about the under determined case. Let m is 2, n is 3, in this case I am assuming a H is of this form. I would like to be able to solve the equations Z equals to Hx. In this case Z 1 is given by this; Z 2 is given by this. Now, what is that I can now do? I can take the first 2 variables are on hand and kick the third variable to the other side. So, I can rewrite the first equation like this, I can rewrite the second equation like this. The determinant of this system is not 0. Therefore, I can solve these 2 equations, but if I solve these 2 equations, let us look at the right hand side. Z 1 and Z 2 are given to us; x is something to be found.

So, I am going to express x 1 and x 2 in terms of x 3. x 3 is a free parameter now that are infinitely different values x 3 can take. So, for each value we assigned to x 3, I can find a corresponding x 1 and x 2. Therefore, there is a pair which is x 1 of x 3, x 2 of x 3; that means, both x 1 and x 2 are functions of x 3, because x 3 occurs on the right hand side. There are infinitely many choices for x 3; therefore, there are infinitely many solutions. So, in this case there are infinitely many solutions there is no uniqueness.

So, in one case there is no solution, in other case there are infinitely many solutions. So, we are in between a devil and the deep sea. This is the typical nature the inverse problem. Inverse problems are generally harder, that is why, in training in colleges we generally learn to solve forward problems; because, forward problems are lot easier to solve. Once you learn how to solve forward problems using the knowledge gained in solving the forward problem then we can hope to solve inverse problems efficiently.

(Refer Slide Time: 21:21)

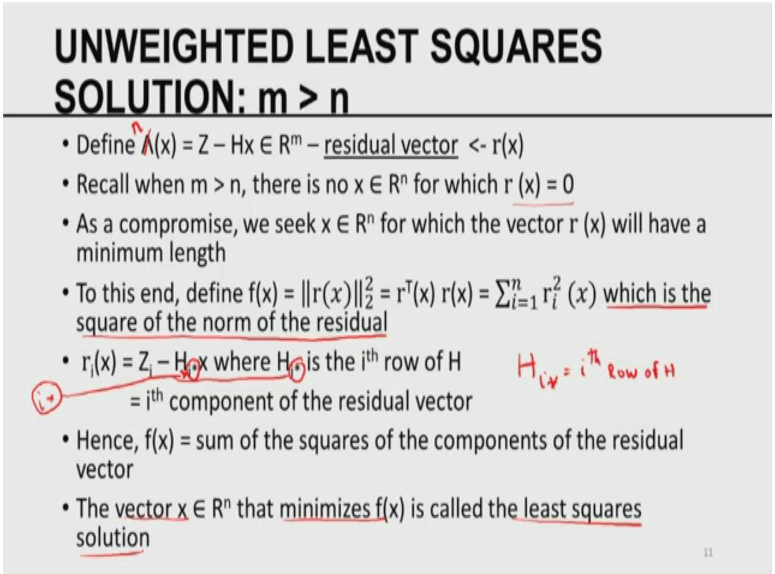


The summary of the linear inverse problems now; Z is equal to H of x. H is the matrix of full rank. So, please understand when H is m and n, the rank of H I want to remind you rank of H is equal to the minimum of m and n.

So, when m is greater than n is the minimum, the rank of H is n. There is an over determined case is there inconsistent system, there is no solution. This is the summary. When m is equal to n, the rank of H is n, there is a unique solution. When m is less than n the rank is m there are infinitely many solution non uniqueness. Generally, in a linear algebra course we essentially deal only with this case. These 2 cases are too difficult. We solve the over determine problem underdetermine problem using the method of least squares.

So, what is the least square solution? A solution is the left hand side must be equal to the right hand side. The least square solution is a solution that may not force the left hand side equal to right hand side, but in sense, we still call it is a solution. It is a generalized solution. So, least square solution is a generalization of the concept of solution, therefore, lease square solution is a very special class of solutions that one has to develop to solve over determined and underdetermined cases.

(Refer Slide Time: 23:03)



So, now that we have seen the formulation of the problem. So, linear static, deterministic, a linear least square problem have 2 versions; one is the under determined, another the over determine. Now, I am going to move towards developing a strategy to solve the problem.

So, what is the method? The method is called unweighted least squares solution. I am going to consider the over determined case. Yeah, this is r, little r of x. It is not lambda, r of x. Define r of x is equal to Z minus H of x; Z is a vector, H x is a vector; the difference is a vector that vector is m vector. That vector is called the residual vector. If the residual vector is 0, Z is equal to H of x, but we have seen we often cannot have there is a residual vector to be 0, in the case of over determined and under determined.

So, when m is greater than n there is no x for which r of x is 0. So, as a compromise what is that we do? The value of r of x for a given Z and H value r of x depends on x. So, r of x is a vector, when r of x is 0 we get the classical solution. So, what is the generalization of the classical solution? For every x, r of x is a vector. Every vector has a length I want to be able to find an x for which the length of this residual vector is a minimum.

If the length of the residual vector is minimum means I am trying to force the right hand side to be as close to the right hand side as possible. We cannot make the left hand side and the right hand side exactly equal, we can bring them as close as possible. This notion of being close instead of being equal is the generalization that comes from the concept of least squares.

So, as a compromise we seek a vector x belonging to R n for which the vector r of x will have a minimum length. So, we would like to formulate the problem mathematically. So, that I can develop an algorithm to that end I am going to define a function f of x. So, what is f of x? f of x is the square of the norm of the residual vector. Now, you can see the norm of the vector comes into play.

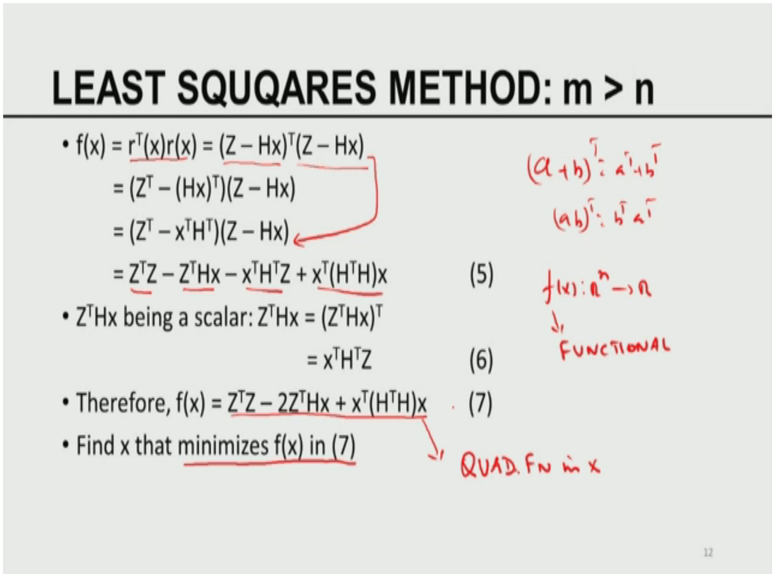
The square of the norm of the vector is simply the inner product of r with r, r of x with the r of x, r transpose r and that is equal to sum of r i square, i is equal to 1T n, so, which is the sum of the square of the norm of the residual. So, f of x is a function of x that represents the sum of the square of the residuals. It is called the r square of the norm residuals.

So, what is r? So, r is a vector, it is m components; r 1, r 2, r 3, R m or r i, i is the ith component of r. r i is equal to Z i minus H of i star. So, this is H of i star, what does i star means? H of i star is the ith row of H. So, this should be i star in the same line. So, same thing is continuing here is i star is in the same line that is ith row of H. So, the inner product of the ith row of H and x, when subtracted from Z i is the ith component of the residual vector. So, f of x is the sum of the square of the components the residual vector. We want to find a vector x that minimizes f of x, that minimizing x is called the least square solution.

So, I would like to comment on this a little bit. We have a case where we already know that there is no solution. Even though there is no solution I would like to be able to look at a generalized concept of a solution. The generalized concept of a solution is that value of x for which the length of the vector, residual vector, r of x is minimum. So, we have converted the problem of solving a linear least square problem into one of optimization problem.

So, that is where the optimization comes into play. So, now, you can see where of the knowledge of (Refer Time: 28:09) vector space, knowledge of norms of vectors, knowledge of minimization and all the things comes into a hue. That is where the importance of module 2 on mathematical preliminaries becomes fundamental to the persuasion of data assimilation problems.

(Refer Slide Time: 28:27)



So, f of x is equal to r transpose x times r, this r is Z minus H of x we already know. So, this is Z minus H transpose Z minus H. We already know the following a plus b transpose is equal to a transpose plus b transpose; we also know ab transpose is equal to b transpose a transpose. These are the 2 formulas I am going to utilize. So, I first distribute the transpose, then I use the product rule.

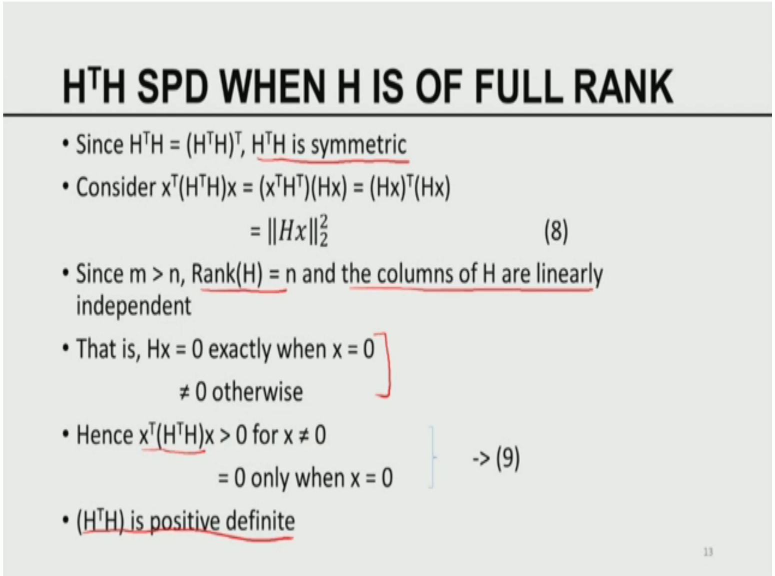
So, this product becomes equal to this product. Now, I am go there are 2 terms in here, I am going to multiply. They are going to be 4 terms. It turns out each of these terms are scalars. Now, look at this now what is f of x? f of x is a function from R n to R. So, f of x is a functional. f of x is a scalar valued function of a vector. f of x maps R n to R is a functional.

So, each of these, this is a scalar, the sum of all the scalars. This is the quadratic function in x. You can you can readily see, this is the quadratic function in x. This is the linear function in x; this is constant with respect to x. Now, it turns out if you consider this Z transpose H x that is a scalar transpose of a scalar is itself. Therefore, Z transpose H x is equal to Z transpose H x transpose, but the transpose of the product is the product of the transpose is taken in the reverse order which is this.

So, the transpose of the second term is the third term, transpose of the third term is the second term. These 2 terms are equal. So, I can reduce that 4 terms to 3 terms by saying f of x is equal to Z transpose Z, 2Z transpose H x plus x transpose H transpose H x. Now, H transpose H that is a Gramian, you may remember that. A transpose A, AA transpose they are Gramian. So, this a Gramian matrix and this is also quadratic function in x. So, this is the quadratic function quadratic function in x.

So, we have converted the problem of estimating the unknown as one of minimizing a quadratic form in 7. Therefore, I want to be able to estimate the unknown. The estimation of unknown is recast as a minimization of a quadratic function. So, you can see the importance of all the things that we have seen in module 2.

(Refer Slide Time: 31:30)



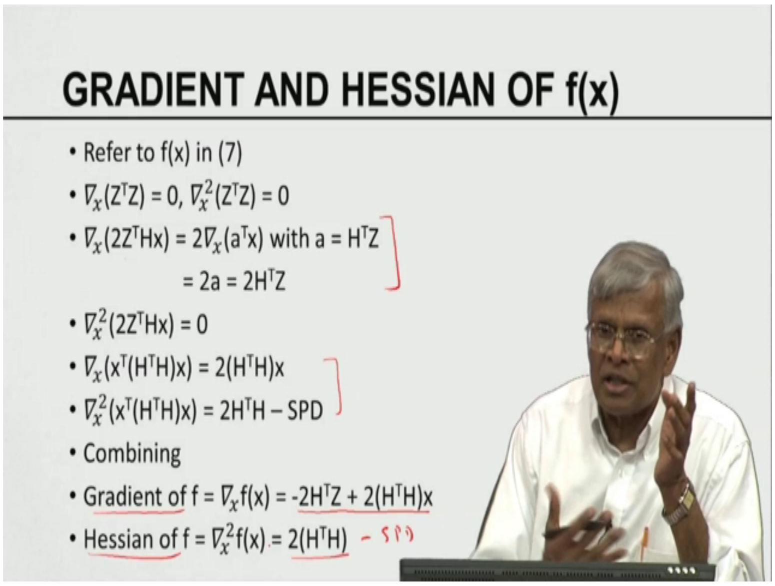
Now, I would like to be able to explore this objective functional little bit further H transpose H is equal to H transpose H transpose, you can readily see. Therefore, H transpose H is symmetric. So, I want to first show that this matrix is symmetric. Sorry, I would like to, that is correct. Therefore, this matrix is symmetric.

If you look at the previous term the quadratic term is x transpose H transpose Hx. So, I am considering x transpose H transpose H of x. I can rewrite this in this particular form, x transpose H transpose H of x. This can be written as Hx transpose H of x and that is equal to H of transpose H of x the norm of square.

So, when m is greater than n, the rank of H is n and the columns of H are linearly independent. Therefore, Hx is 0 exactly when x is 0, Hx is not equal to 0 when x is not equal to 0. These 2 comes from the linear independence of the columns of H that comes into play therefore, this quadratic form is greater than 0 for all x not equal to 0 is 0 only when x is equal to 0. This implies directly H transpose H is not only symmetric it is also positive definite.

So, this quadratic function is a positive definite quadratic function. Therefore, what is that we have now? We have accomplished number of things I would like to be able to consider this f of x, a constant term, linear term, quadratic term. Quadratic term is symmetric positive definite quadratic form.

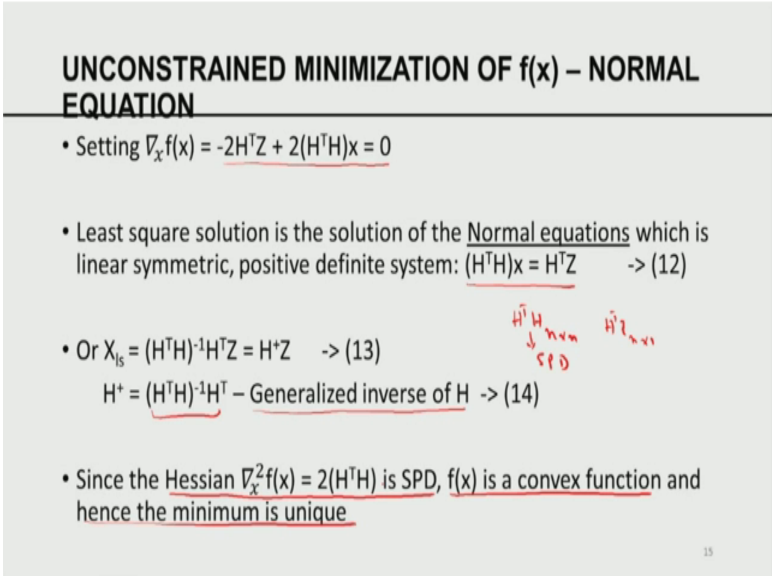
(Refer Slide Time: 33:35)



If you want to minimize, I am going to compute the Hessian of the gradient. So, compute the gradient that there are 3 terms gradient of the sum is the sum of the gradients. So, gradient of Z transpose Z with that of x is 0, second derivative the Hessian is also 0, gradient of 2 times Z transpose H x is equal to 2a, 2 times H transpose Z. This can be computed I would like everyone to be able to verify this using the formula that we have already derived in the class on multivariate calculus.

The second derivative of this term is 0. The first derivative of the quadratic form is this. The second derivative of the quadratic form is also this. If you combine all these results term by term, I get the gradient of H is equal to this term. I get the Hessian of H to be this. The Hessian is already symmetric and positive definite. Therefore, if I equate the first derivative to 0 and solve it that solution must be a minimum because I am equating the gradient to 0 and at the place where the gradient is 0; the Hessian matrix is also positive semi definite. It satisfies the necessary and sufficient condition for the minimum. Therefore, I have found the minimum of the objective function which is f of x.

(Refer Slide Time: 35:03)



So, by equating the gradient to 0, I get this. So, I can transfer the negative term the other side, cancel 2. So, the optimal solution is given by the solution of a linear system H transpose H x is equal to H transpose Z. Now, please understand H transpose H is a n by n matrix. H transpose Z is a n by 1 vector, x is also a n by 1 vector. So, we are called upon to solve a symmetric positive definite this. So, this is a symmetric positive definite system. Such system in least square methodology is called normal equations. These are the set of normal equations.

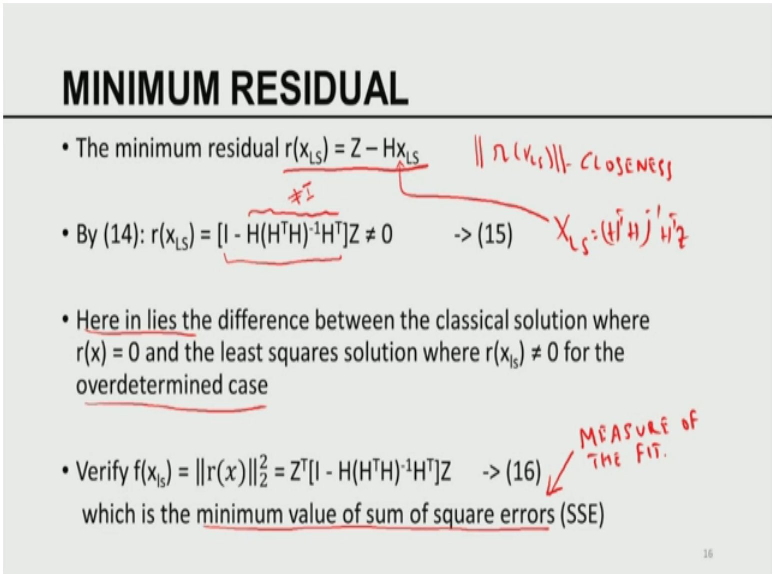
Therefore, by solving this and this matrix is symmetric and positive definite. So, I can solve this by taking the inverse therefore, the least square solution x LS is H transpose H inverse times H transpose Z which I try to write H plus Z, where H plus is equal to this and that is called the generalized inverse of H. Do you remember, when we talked about matrices we talked about the general notion of generalized matrices, generalized inverse of matrices.

So, here we have for the first time in trying to solve a lease square problem have come across the notion of a generalized inverse of a matrix and this minimum is this point to the solution defines the minimum because the Hessian is positive definite and f x is a convex function and hence the minimum is unique. The convexity the function guarantees uniqueness of the minimum. Positive definiteness of the Hessian tells you the minimum is well defined and the function is convex therefore, it is unique, it exists. So, if we have in principle solved the linear least square problem. The solution of the linear least square problem is given by equation 13. So, this is the least square solution.

So, the definition of least square solution intrinsically relates to the definition of generalized inverse of matrices. Now, look at that we have talked about 2 types of generalization; one is the generalization of the notion of the solution itself. The classical notion of the solution is left hand side is equal to right hand side, but here the generalization is the left hand side is close to the right hand side. They aren't equal, but close. We have also generalized the notion of inverse of a matrix, from inverse of a square matrix to the generalized inverse of a rectangular matrix.

So, H is rectangular matrix, H plus is called the generalized inverse of H, when H is a full rank the generalized inverse as an exact expression. The exact expression is given by equation 14, which is H transpose H inverse H transpose. So, we have introduced lots of newer concepts, generalized the old concepts to accommodate the solution for over determined problem and in this process we have demonstrated that all the mathematical tools are used, many of the mathematical if not all are used in the derivation of the least square solution and you can see the least square solution is a solution to an optimization problem.

(Refer Slide Time: 38:50)



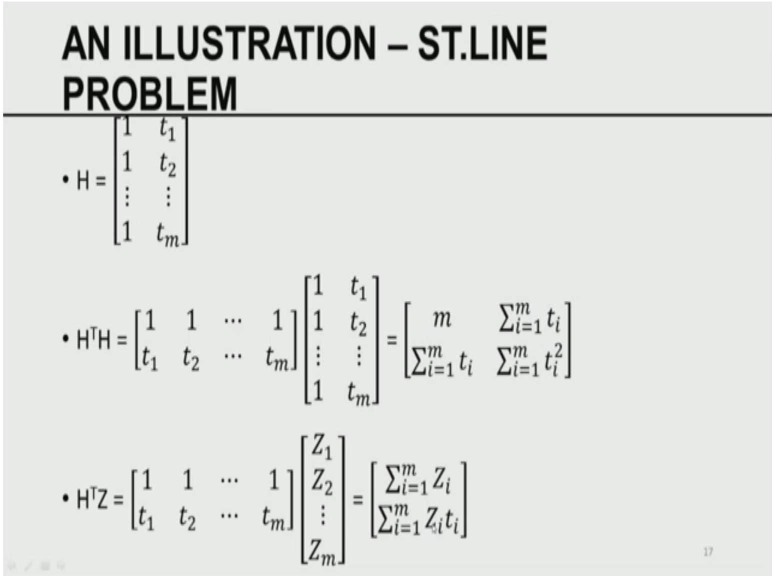
Now, if I have. So, least square solution is not a solution in the classical sense. The left hand side is not equal to the right hand side. I said it has to be close; I want to find out how close they are. So, I am going to substitute x LS in terms of x. So, Z minus H of L of x is the residual at the minimum.

This residual is a vector and what does the theory guarantees? This theory guarantees this is the residual whose length is the minimum and what is the length of this minimum length vector? I am we can readily compute the norm of this. So, the norm of r of x LS gives you the measure of closeness. So, this is the measure of closeness between the left hand side the right hand side.

And how do we show the residual is not 0? Now, let us look at this now, x L of S is equal to H transpose H inverse H transpose Z. So, if I substitute this in here and if you simplify you have Z minus H transpose H inverse H transpose Z. So, this is a matrix, this is the vector. In general, this matrix is not equal to identity. So, long as this matrix is not equal to identity this is not 0. Therefore, the least square solution does not guarantee equality between left hand side, right hand side. The left hand side is not equal to right hand side, but the difference between the left hand side and right hand side, the length of the difference, the length of the residual vector is the minimum.

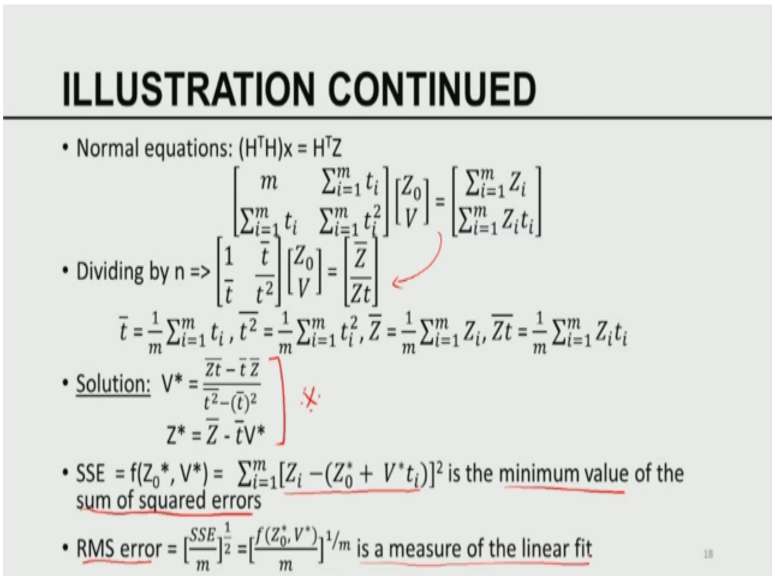
So, here in lies the difference between the classical solution where r x is 0 and the least square solution where r x is not 0, for the over determined case. It is the best we could do. Now, if you substitute x ls f of x, we get the minimum value of the sum of the squares. So, that is what is called the minimum value, that is, a measure of the fit between the model and observation. So, this is a measure of the fit between the left hand side and the right hand side.

(Refer Slide Time: 41:56)



An illustration let us go back to the particle moving in a straight line. H is all 1, t 1, t 2, t m. I can compute H transpose H, this is H transpose, so, I have H transpose H. I multiply them, I get this matrix. I have H transpose Z, I multiply this I have this matrix.

(Refer Slide Time: 42:21)



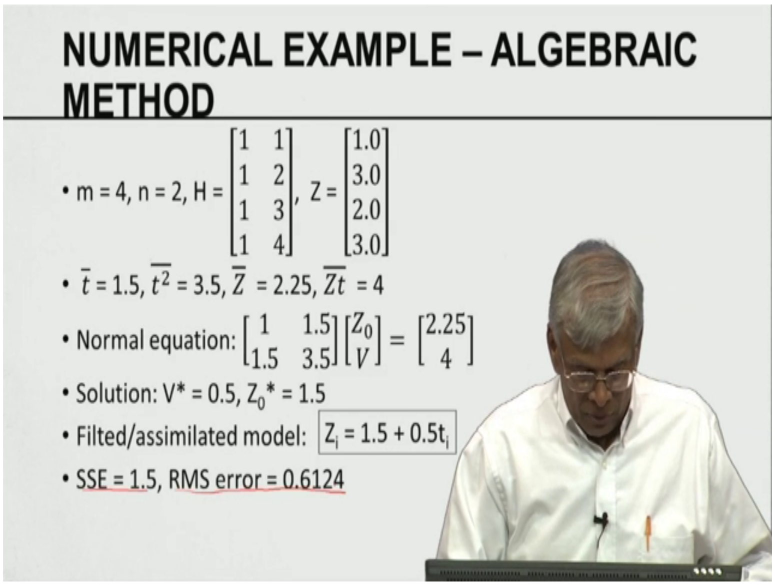
So, the normal equations are H transpose H x is equal to H transpose Z. H transpose H is this matrix. Z naught V is the matrix is the column vector x. H transpose Z is given by this. Now, dividing both sides by m, t bar is the minimum, t square bar is the average of the, t bar is the average of t i, t square bar is the average of t squares, Z bar the average of Z's. Z t bar is the average of the product Z i t i.

So, by dividing both sides of this equation by m this equation becomes, this equation reduces to this equation, where t bar, t bar square, Z bar they are all defined in here. This is a 2 by 2 system, we can explicitly solve it. The solution for this system is given by this. This is an important expression.

So, I got an expression, I got an estimate for V star. What is V star? V star is the least square estimate of the unknown velocity. What is this Z star? Z star is the least square estimate of the initial position. So, if I substitute this in my f of x I get the sum of the square residuals. The sum of the square residual is given by this formula and this formula tells if you replace Z naught by Z naught star V by V star it is the minimum value that is possible.

Now, we are going to define what is called the RMS error. So, SSE in above is the sum of the squared errors. Sum of the squared error divided by m is the average sum of square errors. If I take the square root it is a square root of the average of the sum of square errors that is called the RMS error. RMS error gives you a measure of the linear fit. If the RMS error is large, the fit is loose. If the RMS error is small, the fit is tight. The looseness and the tightness of the fit it all depends on the goodness of the data, the goodness and availability of the data.

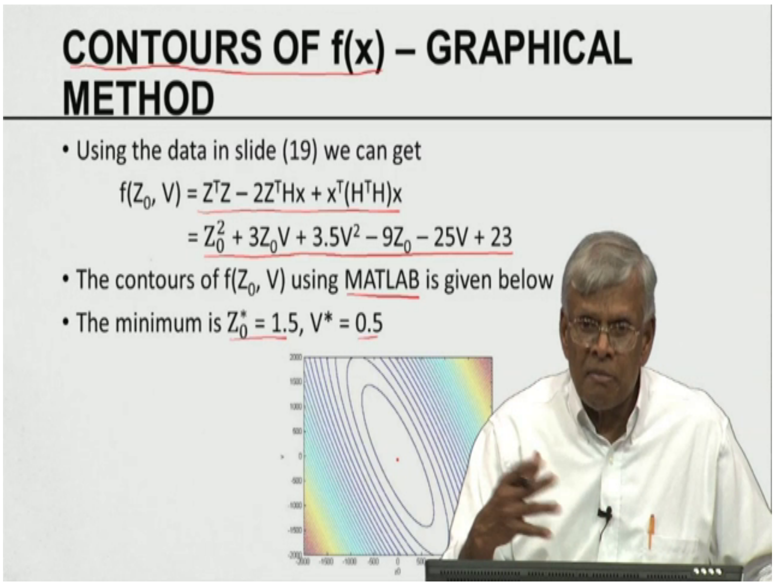
(Refer Slide Time: 44:38)



So, in numerical example, H is the matrix that is given here. Z is the vector that is given here. I compute t bar all the quantities in here. The 2 by 2 system takes this following form. If I solve these 2 by 2 systems I get V star, I get Z star, Z naught star. So, the fitted assimilated model is given by this equation.

In this case I would like you to verify the sum of squared error is 1.5, the square root of the sum of square error is 0.6124. This is the claim I would like you to call you to verify. I think it is better to do these calculations and verify the characters of these things to get a feel for how to do the least square computations.

(Refer Slide Time: 45:32)

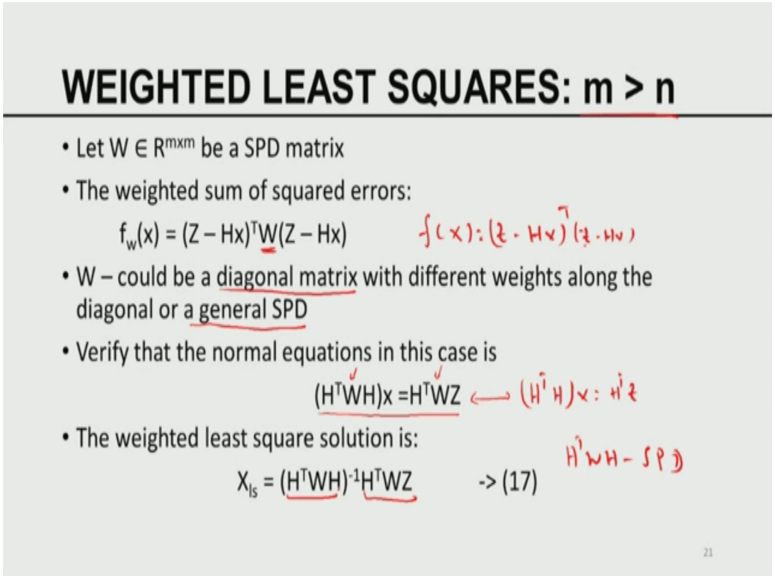


Now, I can I am going to graphically define the solution because it is a simple case 2 by 2. So, I can define what are called contours of f of x. What are contours? Contours are locus of points of the constant value. Now, f of x is Z naught and V. So, f of x has this particular form. For the example, numerical example, that we talked about in this particular case I have actually computed the f of x, the quadratic function takes this particular form in this case.

This is the particular quadratic function. So, what is that we are looking for? This quadratic function is like a bowl sitting and if you took cross sections of that and project them onto the plane they are called contours, using MATLAB I have drawn the contours. You can see that the minimum lies at the centre and if you look at the coordinates of this that happens to be Z naught is equals to 0.5, V naught is equal to 0.5.

So, this way for a small size problem of 2 unknowns you can actually graphically solve the problem by computing f of x and drawing the contours and looking at the centre of that contour. So, this is a graphical method. The previous one is the analytical method, we can solve simple problems by both graphical analytical methods it is fundamental that we do all these things when we are in the learning process.

(Refer Slide Time: 47:13)



So far, we talked about weighted least squares; I am sorry un-weighted least squares. Now, I am going to talk about weighted version. I am still going to be concerned with the over determined case.

So, over determined case, un-weighted least squares is what we saw. Now, over determined case weighted least square is what we are going to see. So, let W be a symmetric positive definite matrix of size m by m. So, instead of, earlier we had considered f of x is equal to Z minus H of x transpose Z minus H of x, in here I am interposing a matrix in between W. So, when W is equal to I, the weighted becomes un-weighted. I hope you see the difference between the weighted and unweighted.

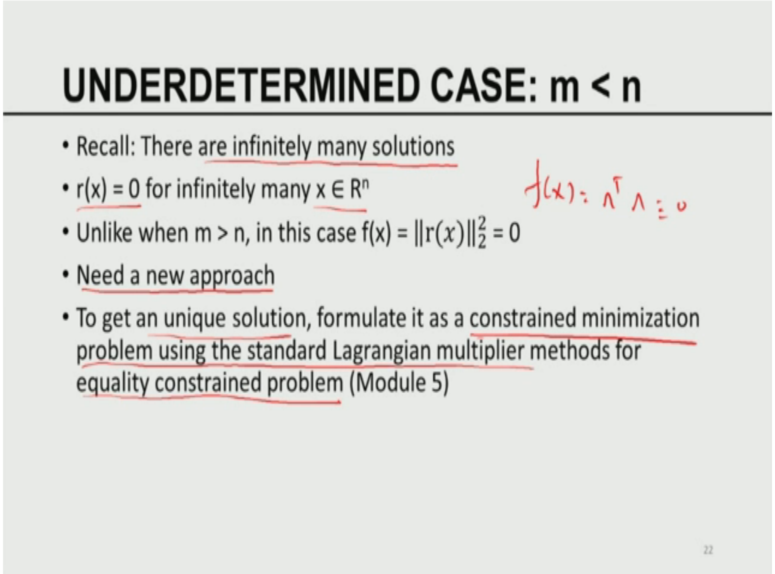
So, in order to emphasize the notion of the weight I am now putting a subscript f of w of x. So, f of W f x is the weighted sum of squares of the residuals. In the special case W could be a diagonal matrix with different weights along the diagonal or in general, it could be a general symmetric positive definite matrix. The difference in weight essentially tells you I am going to give different weights to different components of the squares in the residual error that is all what it means.

In the unweighted case, I am controlling all the sum of the residuals squares have the same value, total democracy that is what unweighted case is all about. In the case of weighted linear least squares some components have greater weight, some components have lesser weight; that means, I am going to keep more important to certain components and less important to another components. The question will arise how do I decide which one should be more important, which one should be less important, that is, outside at the scope of this discussion; that is something that the designer or the person who is interested in solving the problem has to bring to bear those arguments and make sense out of it. But, here we are interested in them mathematical setup. If you are interest in trying to weight the solutions once not understanding how the weights are obtained I am going to tell how to handle the weighted case.

So, W could in the simplest case, W is a diagonal matrix with all 1, which is identity or diagonal elements with all different elements or it is general symmetric positive definite matrix. Again, I can multiply the whole sides, we can try to minimize this as a function of a x. This is also a quadratic function of x. This quadratic function of x I can compute the Hessian, the gradient and equate the gradient to 0, I get a new version of the normal equation. You can really see they are in the unweighted case I simply got H transpose H of x is equal to H transpose Z, here I have a W factor interposed both in the left hand side the right hand side.

So, you can see these 2 equations have very similar structure. So, the least square solution, it can be shown H transpose WH is symmetric, it is also positive definite when H is of full rank. So, in that case I can take the inverse of this. So, x LS is equal to H transpose WH inverse H transpose WZ. So, this is the solution for the linear, static, deterministic weighted least squares. Equation 17 is the analog of the weighted least square compared to the unweighted least squares.

(Refer Slide Time: 51:19)



So far, we have talked about the solution of over determined systems, now we have to proceed to discussing the under determined case. The under determined case we have less observations compared to the unknowns. So far, we considered linear least square problems over determined case.

We talked about the case where it is weighted, it is unweighted. We talked about the generalized inverse; we talked about the notion of a least square solution different from that of the ordinary solution. We also talked about the notion of generalized inverse all that in the context of over determined system. It turns out the theory of over determined system and under determinant system are related, yet different.

Now, I am going to bring out the primary difference between the under determined estimation problem and the over determined estimation problem that we have already seen. So, consider the undetermined case m is less than n, m is the number of observation, n is the number of unknowns. Recall in the case of undetermined problem. There are infinitely many solutions. So, we have headache of one kind in the over determined problem namely, there is no solution.

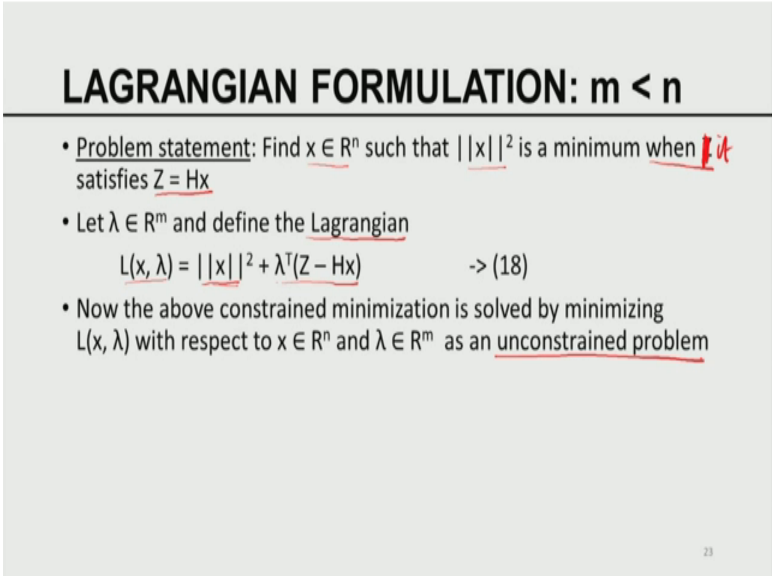
Here, headache is of another kind; there is not one solution, but there are infinitely many solutions. The challenge is, how we pick one among the many infinitely many solutions that make sense for us and why we are interested in uniqueness? When you want to be able to compute the solution using an algorithm, if you want to be able to calculate every calculation must have target. I want to be able to calculate this quantity, that quantity. So, since very algorithm always seeks to find at targeted solution, a targeted unique solution we need to be able to build in the notion of uniqueness before we start talking about computing the solution.

So, the computational process has to wait until we define, what an appropriate solution is, what is an appropriate unique solution among infinitely many possible solutions. So, in this case, look at this now, there are infinitely many solutions. Solutions mean what? The residual 0, so, there are infinitely many x for which r x is 0. If r x is 0, the f of x which is equal to r transpose r is identically 0. If r transpose r is identically 0, there is no x, there is no minimization. So, there is no possibility of doing anything similar to what we did in the over determined case, for the under determined counterpart. Therefore, we need a new approach.

We need a new approach to get an unique solution. In order to do that we are going to formulate this as a constrained minimization problem and this constrained minimization problem is going to be solved by Lagrangian multiplier technique. This constrained minimization problem is going to be an equality constraint minimization problem. So, you can see everything that we have seen in the module on optimization gets to be applied here.

So, the pathway to the solution in the under determined case is to formulate the problem as a Lagrangian multiplier problem using equality constraint and this equality constraint minimization problem is going to help us to pick that optimal solution among the infinitely many possible solutions that is the pathway.

(Refer Slide Time: 55:34)

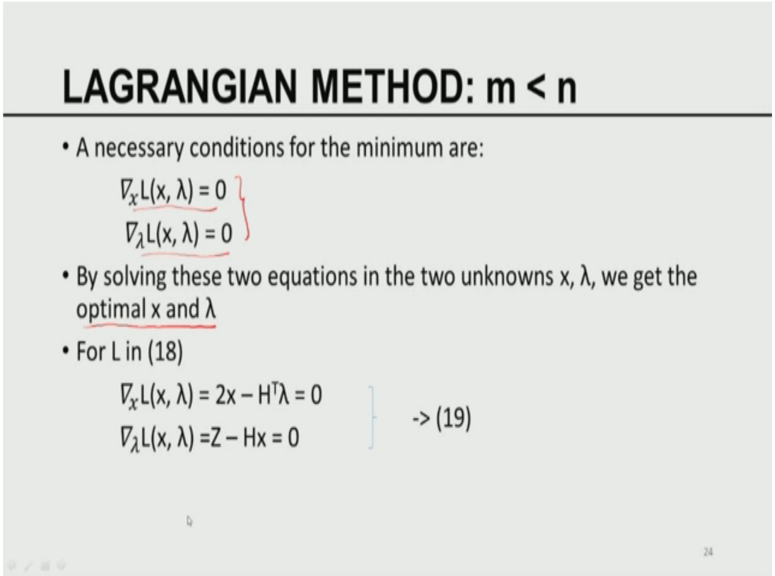


So, what is the problem statement? How do I state the new version of the problem? Find the vector x belonging to R such that it is norm is the minimum. Look at that now. I am not interested in any vector, I am interested in picking a solution with the minimum norm, but that x not only must have a minimum norm, but it also must satisfy Z is equal to H of x.

So, the problem statement is find x such that the square of the norm is the minimum when it satisfies. So, this must be when it satisfies Z is equal to H of x. So, x must satisfy this H of x, that is the constraint and the norm of x must be minimum. We formulate this as a Lagrangian multiplier problem. So, let lambda be R of m, because Z is the vector in the m dimensional space, H of x the m dimensional space. So, let lambda be a m vector, define the Lagrangian. The Lagrangian x, lambda is given by, this is the function to be minimized, this is the constraint. We are following the same formulation that we described in the module on optimization.

So, equation 18 becomes a Lagrangian. There are 2 independent variables x and lambda. So, the above constrained minimization problem is now replaced by an unconstrained Lagrangian minimization problem.

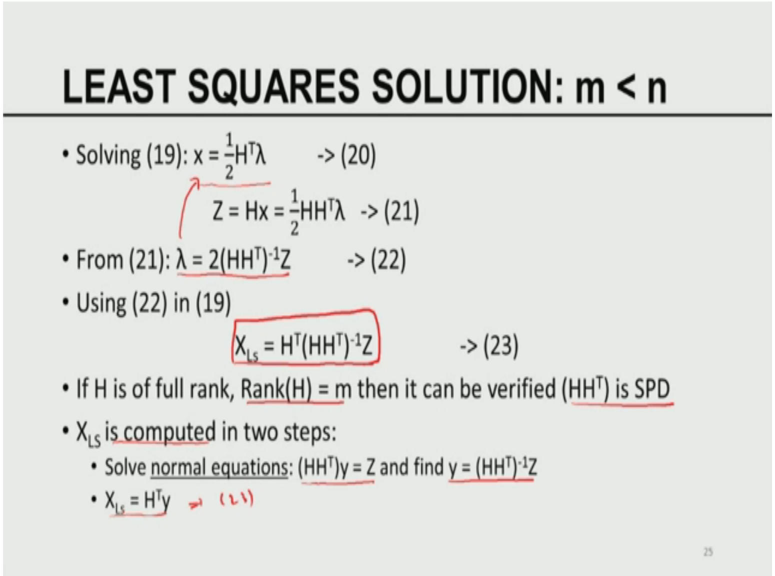
(Refer Slide Time: 57:20)



This problem is solved by standard techniques. I want to be able to compute the gradient with respect to x and minimize with respect to x. I would like to be able to compute the gradient with respect to L and minimize with respect to m. So, these 2 equations must be simultaneously satisfied to find the optimal x and the optimal lambda.

So, there are 2 unknowns, x and lambda, the lambda and x that satisfy these 2 equations are called the optimal x and the optimal lambda. Now, for the x, I am sorry, for the Lagrangian given an equation 18, if you compute the gradient of x, gradient of L with respect to x and lambda, there are 2 equations; 2x equals to H transpose lambda, Z minus H of x is equal to 0. We have to solve these 2 equations simultaneously.

(Refer Slide Time: 58:18)

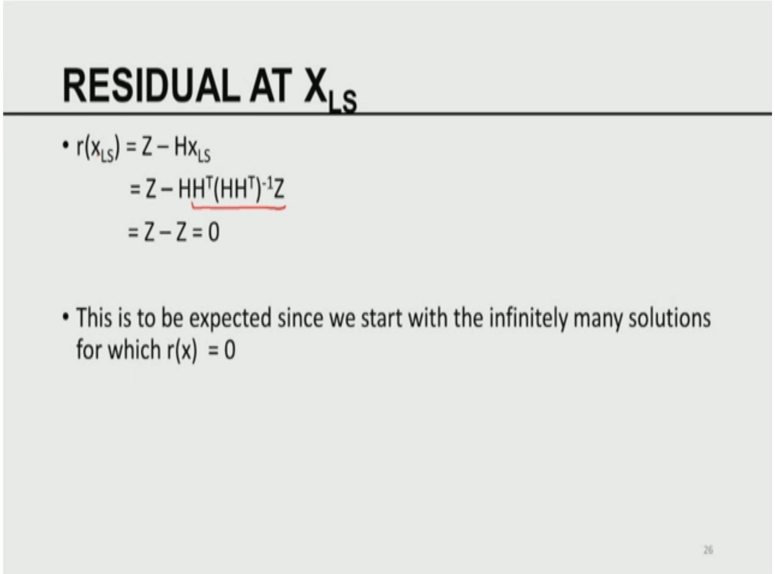


If I solve these 2 equations simultaneously, I get the solution to be lambda, the optimal lambda is given by this, the optimal x is given by this. So, if I substitute this lambda from here to this equation, I get the optimum least square solution X L of S to be H transpose HH transpose inverse Z. So, this is the unique solution in the case of underdetermined problem. Then H is a full rank, the rank of H is m, it can be verified that HH transposes symmetric positive definite. So, it is inverse exists. Therefore, X LS, the least square solution can be computed in 1 of 2 steps.

So, solve HH transpose y is equal to Z and find the solution y is equal to HH transpose inverse Z and then we can compute X LS, H transpose y using this I can this implies 23. Therefore, the computation of the least square solution is done in 2 steps; one, by solving a linear symmetric system and another using the solution substituting this to get the least square solution.

So, we have by invoking to the Lagrangian multiplier technique for equality constraint problem, we have obtained the solution for the underdetermined case.

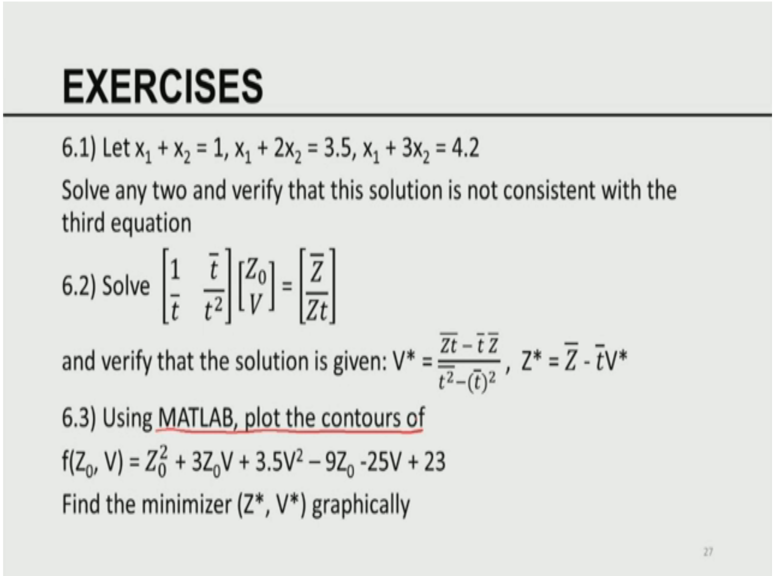
(Refer Slide Time: 59:57)



In this case, I know the formula for X LS from equation 23. So, r of the residual of the minimum is Z minus H of X L of S. X L of S is given by this expression. So, if you think of this and multiply by H, you have HH transpose, you have HH transpose inverse. So, the one is the inverse the other, so, they get cancelled it be. So, Z minus Z is 0. So, in this case the residual is 0.

So, the optimal solution is one such, where the residual is also 0. So, that means, it satisfies the constraint as to be expected since we start with the infinitely many solution for which r of x is 0, this residual at the minimum must also be 0. So, that is verified.

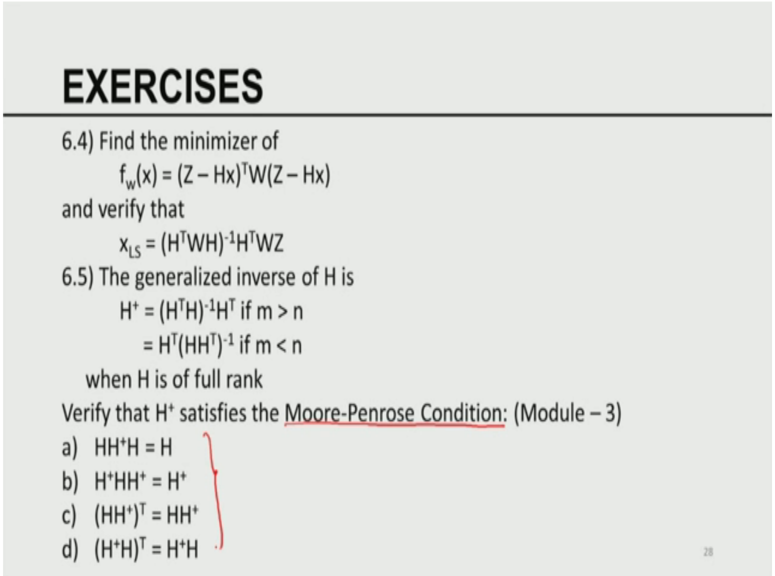
(Refer Slide Time: 61:00)



So, with this we come to the end of the discussion of the linear deterministic, static, inverse problem both undetermined and over determined. We solved the over determined problem in inconsistent case where they didn't have a solution, we tried to bring the right hand side on the left hand side together as close as possible. In the second case, there are infinitely many solutions. Among the infinitely many solutions, we have tried to find the one that is of least length, the norm of the solution is the least.

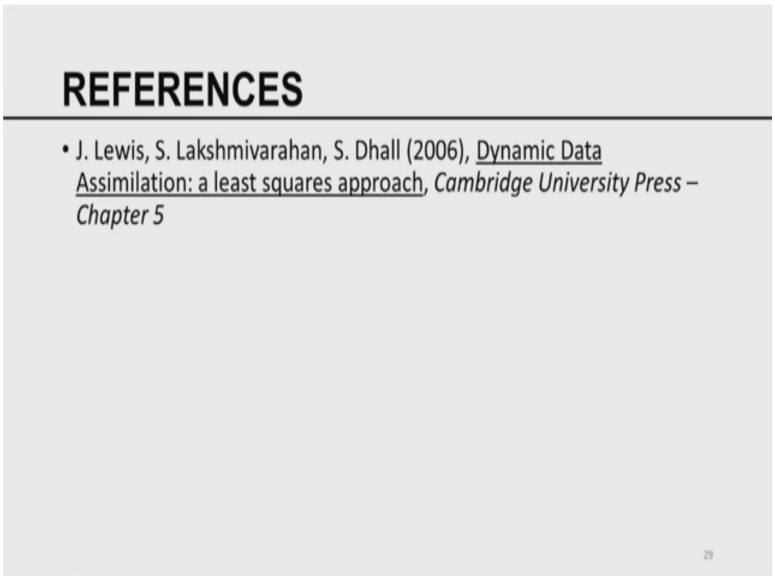
So, that is how we induce uniqueness into the least square solutions. With this I would like to encourage you to solve a couple of different problems. The problems are directly related to the development in the text. You see, in particular, I am going to emphasize that you must do the MATLAB related computer problem by plotting the contours. Once you plot the contours you can get rid off the minimum by graphical approximation by approximating the centre of the contour.

(Refer Slide Time: 62:21)



I am also giving exercises with respect to expressions for the generalized inverse finding the Hessian, the gradient of different functions. I am also trying to define properties of the Moore-Penrose inverse, which we have already discussed when we discussed matrices and the properties of generalized inverse are given by these 4 equations as the Moore-Penrose condition demands.

(Refer Slide Time: 62:58)



This development is taken from our book Lewis, Lakshmivarahan and Dhall published in 2006, Dynamic Data Assimilation: a least squares approach, published by the Cambridge University Press, it largely follows the development in Chapter 5.

Thank you.