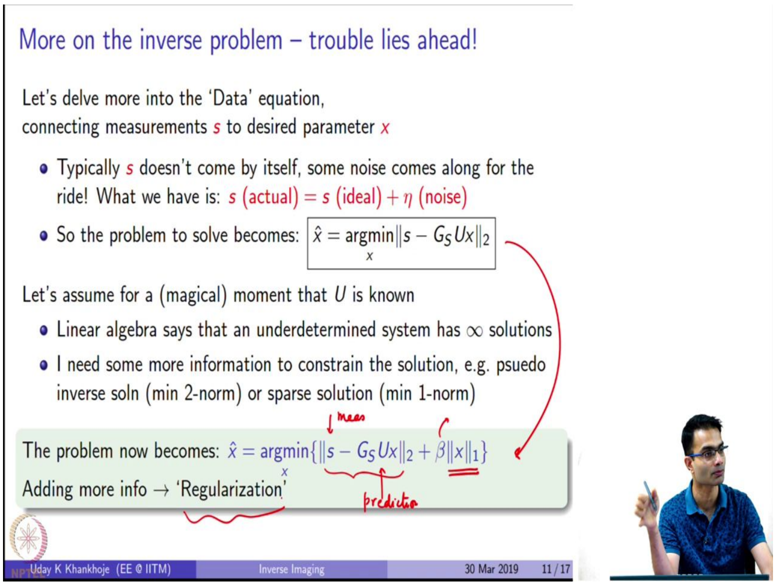
Computational Electromaganetics
Prof. Uday Khankhoje
Department of Electrical Engineering (EE)
Indian Institute of Technology, Madras

Applications of Computational Electromagnetics
Lecture - 14.03
Inverse Problems – Challenges

(Refer Slide Time: 00:14)



Let us look at this data equation. So, what all challenges are there in this inverse problem? So, this data equation s is connecting the measurements to the desired parameter we have seen that. Now s of course does not come by itself. What comes with it? Noise right. So, the $s(actual)$ contains the ideal measurements and noise. So, of course, I cannot I mean this is a real life problem this is an antenna there will be some noise on the receiver so you will get noise corrupted measurements.

But we expect this to be a challenge in any engineering problem. What else? So, what happens is that I want to solve this problem right. So, argmin means that x which gives the minimum of this expression $\|s - G_S U x\|_2$ and that is my solution. So, let us assume magically some oracle has come and given us this value of U. Let us assume it is not realistic, but let us assume even if we had perfect knowledge of view what is the problem.

So, we this is the part we have already discussed that the underdetermined system will have infinite solutions. So, there is some more information that needs to be provided to in some sense reduce the infinity right. So, what I could do is in linear algebra if I am presented with an underdetermined system of equations, do I just declare that I cannot do anything or do I still try to do something? I use you have you have all heard the word of a pseudo inverse.

Right, pseudo inverse exists for matrices that I am not squares right; what is the property of a pseudo inverse solution? For example, in an underdetermined system of equations, I can write a pseudo inverse and get some solution for x and that x has what property an underdetermined system? So, those of you remember your linear algebra the pseudo inverse was the solution which had minimum distance from the origin.

Student: Least non solution.

Right least.

Student: Non-solution.

Non-solution, but which norm was that? Euclidean norm; right, 2-norm.

Student: 2-norm.

2-norm; so it had minimum 2-norm. So, I can if I use the pseudo inverse I am basically saying give me the minimum 2-norm solutions, where one and the same statement. They may be some motivation to go for what is called as sparse solution, then I will give you a motivational why we would want a sparse solution and that is sparse solution comes by imposing a norm on the minimum 1-norm instead of 2-norm minimum 1-norm.

Student: Least energy is minimum.

Least energy is minimum 2-norm and sparsity; so, what is sparsity means? Most of the entries are zero, very few entries are non-zero ok. So, we may wonder why should this be a solution why should I want to promote this we will talk about that. So, if let us say just assume for a second that I want a minimum 1-norm, the next slides I tell you why 1-norm is a good idea, but let us say I wanted minimum 1-norm.

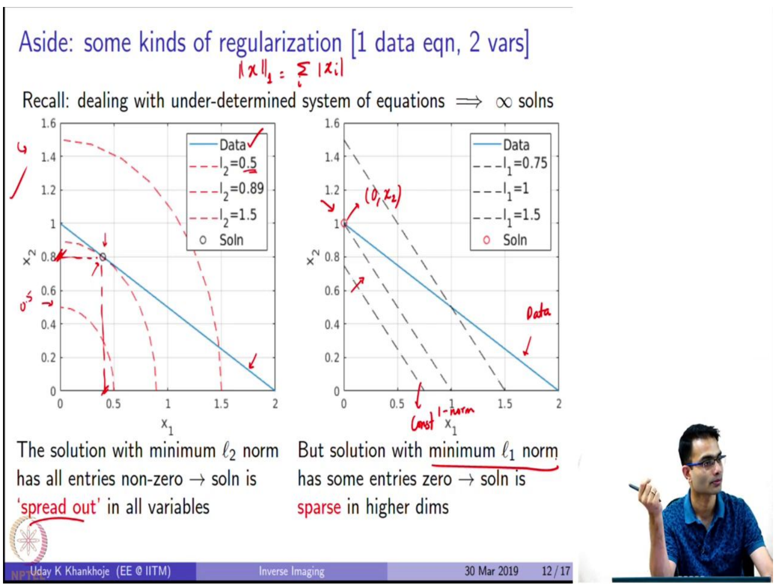
Then this optimization problem gets modified like this that not only do I want to minimize the distance or the difference between prediction and measurement right, this is my measurement and this is my prediction. Not only do I want this to be minimum, but I also wants the 1-norm to be a minimum and beta is some hyper parameter which I will determine numerically.

So, I my ideal solution is what? One which minimizes both of them, out of this infinity of solutions I want that solution which minimizes this. And in the language of inverse problems this or even statistics this adding more information is called regularization that is the technical word for it ok. So, I have regularize my solution by giving some more information and what is that more information? Minimum 1-norm, minimum 2-norm something else, it's up to me.

Student: Sir, how do you say this as (Refer Time: 04:27).

Yeah, so, why it should this correspond to reality is the question right. Why should my solution to this breast cancer imaging problem have minimum 1-norm?

(Refer Slide Time: 04:39)



So, now let us look at; so, what will do is we will just look at a property of 1-norm vs 2-norm and then motivate why 1-norm is a good idea. So, infinite solution that I found is undetermined, let us look at two separate cases ok. So, this first case over here is where I have this blue line over here; so, two variables one equation that is the simplest thing we can visualize, that one equation in two variables is what a line? That is your blue line; so, I am calling at the data equation ok and if I want minimum l_2 norm that is minimum energy then what I can do? So, these red dash lines over here these are all circles of constant 2-norm right. So, for example, this is 0.5 right, this one is 0.5 radius so I keep increasing this until I touch the data equation.

So, when I touch this what is happening? At this point the data equation is being satisfied because x_1 , x_2 lie on the data line; that means, I have got $s = G_S Ux$ very good and at the same time of all possible solutions can I say that this solution has the minimum energy? Yes, right because this is just a tangent over there. So, I say very good this is my minimum energy solution, but now when I look at the answer over here, it has some non-zero value of x_1 and some non-zero value of x_2 , so there in some sense spread out ok.

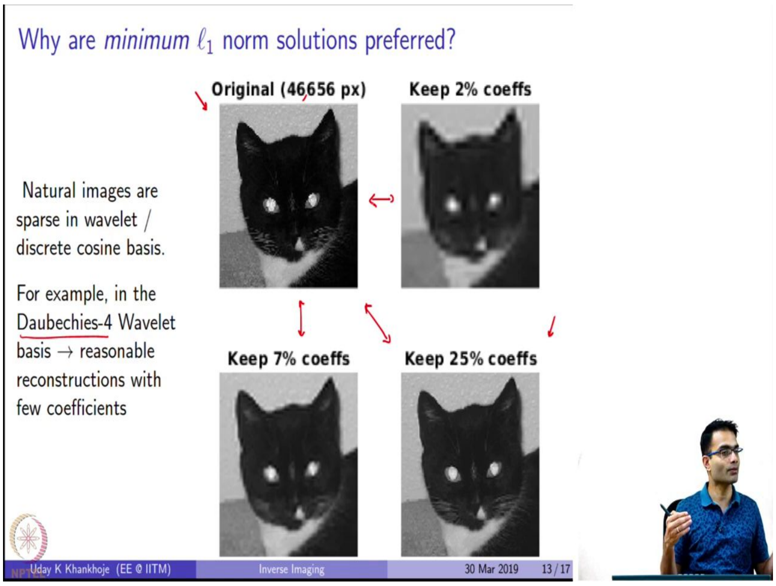
Now, when I look at minimum 1-norm solution, so this is my data equation as before, in this data equation what am I doing? I am plotting now these are constant 1-norm solution constant 1-norm. Because the locus for constant 1-norm is what; how do I define the l_1 -norm? Right it is a summation of $|x_i|$.

So, in the upper quadrant over here, what will the equation be? $x_1 + x_2 = const.$ right; so, that is just the line. So, these are all parallel lines that are expanding in the value of the 1-norm. Now, what is the earliest I can hit this blue line is at this point; so, this point over here it satisfies the data equation because it is on the blue line and it of all possible solutions it is that which has the minimum 1-norm right.

Now, just look at the nature of this solution. Over here, you can see that this has a non-zero value of x_2 and a value of x_1 is 0 right. So, in this very toy example it is a sparse solution because one entry is 0, if I expand this idea to multiple dimensions this hyper plane will touch at some axis only, it will not touch somewhere in between. So, my solution will be sparse. So, this putting a minimum l_1 -norm tends to promote a solution which has sparsity. So, your this is very clear like when I in the previous slide when I said that imposing 1-norm gives me a sparse solution you know why it does right and that is geometry.

Now let us look at the critical question, why should I why should my reality correspond to a sparse solution?

(Refer Slide Time: 08:06)



So, here is where there is a little bit of signal processing that we need to look at ok. So, we are going to look at natural images and in signal processing or DSP you have looked at different kinds of basis; so, there are things called wavelet basis, discrete cosine basis, Fourier transform basis, various such bases are there right. And I am going to look at a natural image right cat is a natural object, so it is also a cross section of breast tissue. So, this is the original picture over here top left this is the original which has some how many 46000 pixels.

So, in order to represent this picture correctly in the pixel domain, I need 46000 numbers right. If I keep I use what is called a Daubechies 4 wavelet basis we need not know what it is, but it is some kind of an orthogonal basis. In this orthogonal basis, if I keep only one-fourth of the coefficients only one-fourth I get this. Can you visual it any difference between these two? Visually it is almost, I mean, next to impossible with a naked eye, to tell any difference between these two images, but what I have done. From 46000 plus I have kept only one-fourth of the coefficient. So, only 10000 coefficients were needed, then I got a little bit greedy I did a little bit more optimization within this wavelet domain and I look at only 7 percent coefficients.

Now, I can see that it is a little furzy, but I can clearly make out it is a cat, I can tell you where the eyes are where the nose is where the ears are and so on, but how many coefficients that I have kept right, roughly 3000 coefficients. So, when I started from 40000, I am down to 3000 coefficients. Then I have got even greedier, I said let me keep only 2 percent coefficients right; When I keep only 2 percent coefficient this is very fuzzy. But, if I wanted to if my information that if my question was where are the eyes I can still answer the question if my question is how many whiskers are there, I cannot answer this question right.

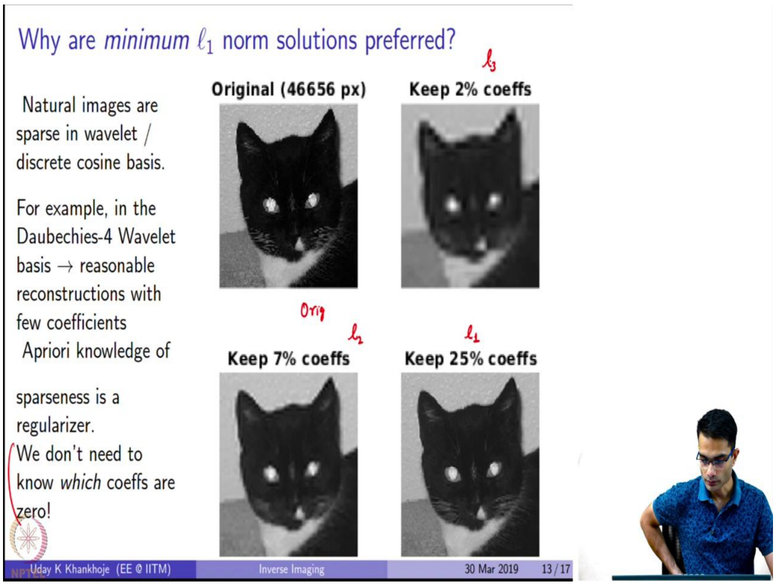
So, this is sort of a very good motivation why, if my imaging domain instead of looking at it in the pixel domain which needs 46000 numbers, if I transform it a linear transformation. So, linear transformation, a wavelet transformation is a linear transformation; means it can be characterized by a matrix simple. If I choose a correct matrix, I can with just 2 percent coefficients; so, 2 percent of this is going to be roughly how much?

Student: Less than a 1000.

Less than a 1000 right, 800 or whatever.

Yeah, it 900 or something like that with 900 numbers I can get a solution that is so good right, and what when I keep 2 percent coefficients what I have done to the remaining 98 percent? I have set them to 0. So, in the wavelet domain this image is sparse, very very sparse only 2 percent 2 out of 100 numbers are non-zero rests are all zero right. So, what, but I mean this is the place where I am using a priori information, I am using the apriori information that the cross section of breast tissue is a natural image; I know I am empirically it has been observed the natural images are sparse in wavelet domain in discrete cosine domain and so on. So, I am this is me using a priori information and solving the problem in the wavelet domain why because a number of variables there are lesser.

(Refer Slide Time: 11:45)



Yeah, the other great thing is that we do not need to know which coefficients are zero, I can when I solve the minimum 1-norm solution I do not have to tell the solver these coefficients are going to be non-zero. Tt can figure out which numbers are.

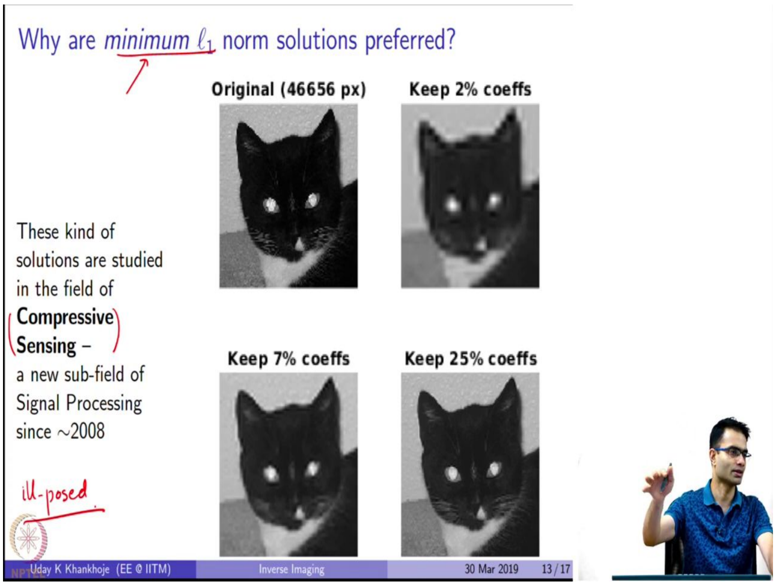
Student: (Refer Time: 12:01).

Yeah; so, that is actually what we have done. So, for example, this is the original image, this is what is called a level 1 wavelet Decomposition, this is a level 2 wavelet Decomposition

Student: (Refer Time: 12:12) level.

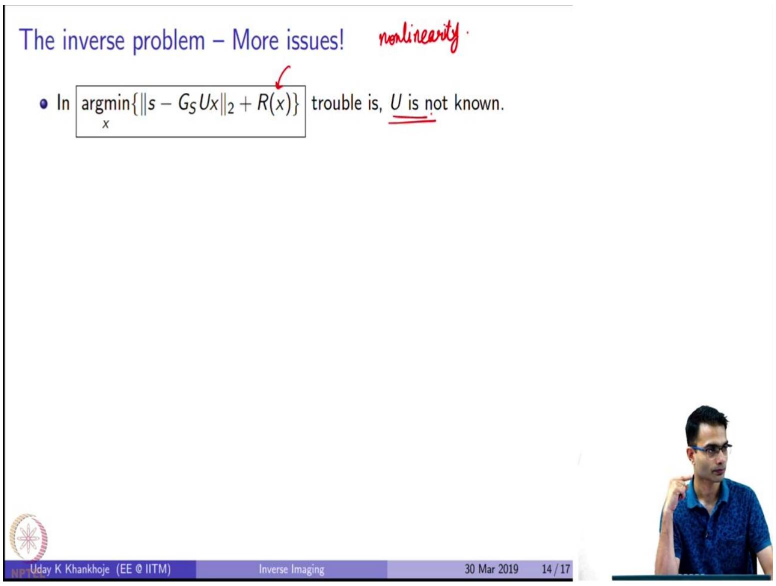
That is what has happened; I mean if these are not separate wavelet transforms, wavelet transforms to different levels right, so this is a level 1, level 2 and level 3 right. So, they are all related, they are not unrelated ok. So, if you study the theory of wavelet transforms you will know how they are very related, you can go even more, you do not have to stop at 3 right.

(Refer Slide Time: 12:32)



So, this is the field where these problems are studied is the field of compressive sensing. So, compressive sensing means I am trying to get sparse solutions from under sampled data. I have undersampled data and I want something that can recover the solution all right.

(Refer Slide Time: 12:53)



So, now let us look at a few more issues.

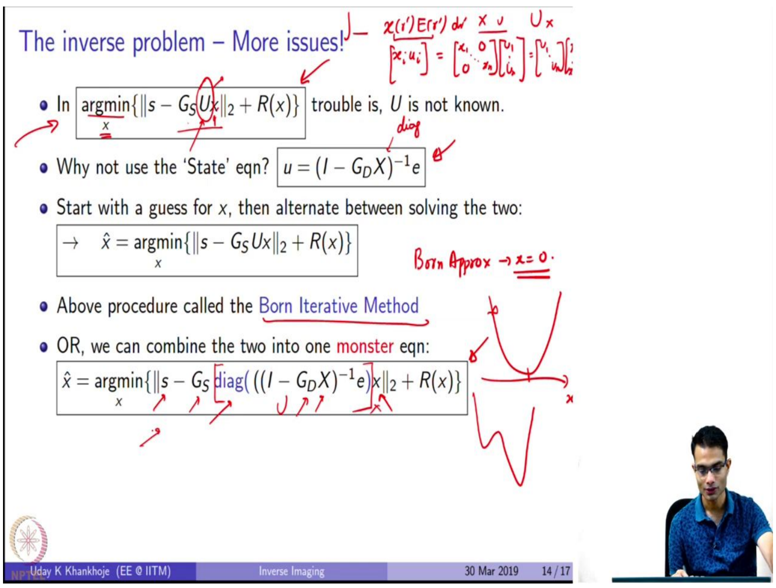
Student: (Refer Time: 12:57).

So, why; so, minimum l 1 means what? We have seen that minimum when we go back over here minimum l 1 meant sparse. I know beforehand that my object is going to be sparse in the wavelet domain right. So, that is why I am going to solve my problem and enforce minimum l 1-norm. So, that I know, I can get I will get a meaningful solution right, that is the logic.

So, this was one problem; I mean, what we have discussed so far was the problem of not having enough information right, not having enough information is what is a technical word for it is ill-posed. So, the inverse problem is ill-posed there is not enough information that I have to give additional a priori information, it is sparse, it is this band limited, it is this and that and all of that, that is a priori information.

The second problem that we are going to talk about is a problem of non-linearity. So, what we assume so far was that the field U is what? We had said magically someone has given it to me and we were only worrying about 1-norm or 2-norm. Now let us come back to real life where U is also not known right. So, in this problem this Rx is the regularization, this can be 2-norm, it can be 1-norm, it can be wavelet something whatever right let us just leave out leave the details out. But the main thing is that U is not known, we have not yet come to that problem; so, let us see what happens here.

(Refer Slide Time: 14:37)



So, you may say that U is not known no problem, I have a state equation that state equation told me that there is a relation between that use the matrix G_D. So, if you give me e and if you give me X I can estimate U and that U I can take and put in over here that is what I can do right.

So, what can be a strategy to solve this problem? Start with some guess for x ok, put it into this equation, get U from there, put that U into this equation and then I am done I can solve this. This becomes a problem only in one variable x right, sounds reasonable I could do that right. And if I do that that there is a name for it discovered late 80's early 90's called the Born Iterative Method. Such an iterative method starts with some guess for x, then calculates U form that and puts that U into the data equation and gives me the x ok.

Student: Which is one thing?

Yeah.

Student: What is this small x?

This x?

Student: Because argmin.

Ar; what is my variable of minimization is x that is written below the argmin.

Student: Permittivity.

x is contrast, permittivity minus 1, that permittivity minus 1 appeared everywhere. So, I called it one variable x instead of having to write something minus 1 and so x right so, right. So, this was the born iterative method, now strictly speaking when I go to minimize this problem those of you who have done a little bit of optimization, you should know how will I minimize this? Forget optimization supposing, I give you a function in 1 variable and I ask you find the minima of it what will you do?

Student: Differentiate it.

We will differentiate it. In fancier language in optimization, what would you do?

Student: Gradient.

You find the gradient. Basically idea is of differentiation. Now this object over here U over here, what are we doing? We are putting substituting some value based on an initial estimate of x that is what we are doing. But when I go to take the derivative are we keeping U to be constant or if I if it is a number then it is a constant I cannot take the derivative with respect to x for U, but actually it is U is a function of x and that function is here.

So, if you are very very strict what you should do? You should take this U and substituted into your original equation to get this one, this is your U right $G_S Ux$. Now, I have eliminated any intermediate variables it's only an equation in x right, because s is measurement, G_S is known, G_D is known, I have my x over here, I have my x over here. So, this is a very complicated looking equation, but it is an equation in one variable; one unknown not one variable one unknown x.

Student: Why did we do this?

Why did we do this?

Student: So, the diagonal.

Oh, why did I write this as diagonal? I was hoping to not get into it, but I will let us write this over here? So, this Ux over here that appears everywhere if you notice in the integral equation what was the actual thing, it was $\chi(r')E(r')$ right, and then the G over here and this whole thing was integrated.

So, this thing over here they always appeared in the language of linear algebra it was like writing $x_i u_i$ right. So, it was a column vector with $x_i u_i$ now, how do I write this $x_i u_i$? In terms of x and u, I can write this as and x like this x_1, x_n multiplied by u_1, u_n right. It gives me the same thing, this is 0 and this is 0 or I can also write this as $[u_1 \cdots u_n]^T$ and a column vector $[x_1 \cdots x_n]^T$.

Student: (Refer Time: 19:08).

Yeah, that is why you notice that they sometimes it is small x, sometimes it is X right. So, when I write it as capital X I mean it as the diagonal version; that means, so this is capital X into u and this is Ux ok. But since we are not solving this right now, I am just leaving it like that right. So, that is why this whole thing becomes diagonal of whatever I get, whatever I get is going to be a column vector I need to put into a diagonal matrix form right.

Student: Sir, how do we choose an initial value of x?

How do we choose an initial value for x. What is the most, what would you do if you are trying to solve it?

Student: 1

1; why 1 or you can do 1?

Student: Permittivity is like (Refer Time: 19:53).

This is x, not permittivity this is contrast.

Student: Contrast so, it going to be 0.

0.

Student: (Refer Time: 19:59).

0 contrast.

Student: 0 contrast.

So, 0 contrast is actually is the most popular starting guess and there is a name it is so popular there is a name for it which is common in not just in inverse imaging, but also in particle physics, it is called the Born approximation right to begin with. So, this integral equation that we got comes in quantum mechanics scattering equations also; so, say this is the same Born from quantum mechanics and the physics people call it Lippmann-Schwinger equation and electrical engineers call it Fredholm integral equation this is the same thing.

Now, what is the problem with this equation? Its exact, so when I go to take the gradient of this, it has everything inside it there is a x where it should be unlike this equation where U appears to be just some number some constant right. So, this is misleading; so, if I wanted to correctly find out the gradient, I should use this equation and finding this gradient is a lot of hard work ok.

Student: Why x 0 is then a.

Initial guess initial guess, you can choose anything you want I am just telling you what is common.

Student: Sir, the x is equal to epsilon minus 1 right.

Yeah; so, if you choose $x = 0$ means you are assuming vacuum everywhere and you are building up from there. You need not.

Student: Finally, the iteration will lead to the same guess?

Will the iterations lead to the same guess or not? Very good question.

Student: Like of a (Refer Time: 21:29).

In this born iterative method the question is will I arrive at the same point regardless or what guess I take?

Student: Right so it.

Yeah, will I get finally, the same x?

Student: I do not think so.

Answer is I do not think so, the answer is.

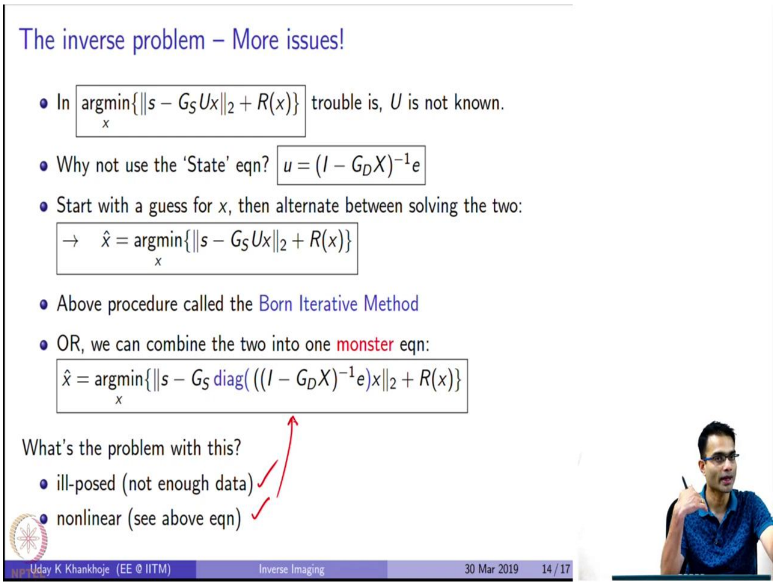
Student: Do not think so.

Do not think so, and that is correct do not think. So, why give me a precise reason why, we have enough information on the slide to tell me why.

Student: (Refer Time: 21:51) multiple.

Multiple right; so, if I if my cost function let us say I have a variable in one this thing and if my cost function was a nice parabola, regardless of where I start I always end up with the same point. But if my cost function were multiple minima right. Then of course, it matters where I started from, and look at this cost function over here, this objective function its anything, but a nice parabola right. So, where you start from matters right, actually that is one of the major issues in this problem. So, let us actually look at it, it will be a little bit more quantitatively ok.

(Refer Slide Time: 22:28)



So, they have already said that ill-posed not enough data and non-linear right, this equation over here is non-linear in x that is the problem. So, let us see the manifestation of this nonlinearity in how does it affect us, we know it is going to affect us, but how it is going to affect us.