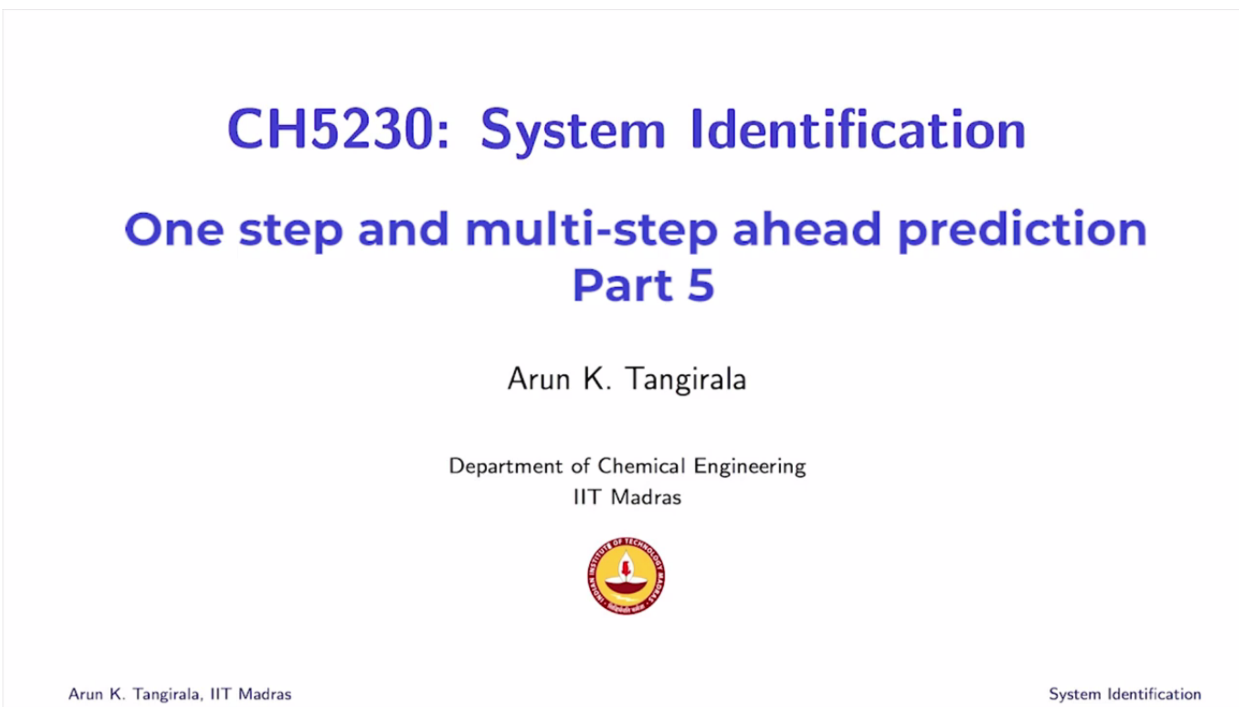
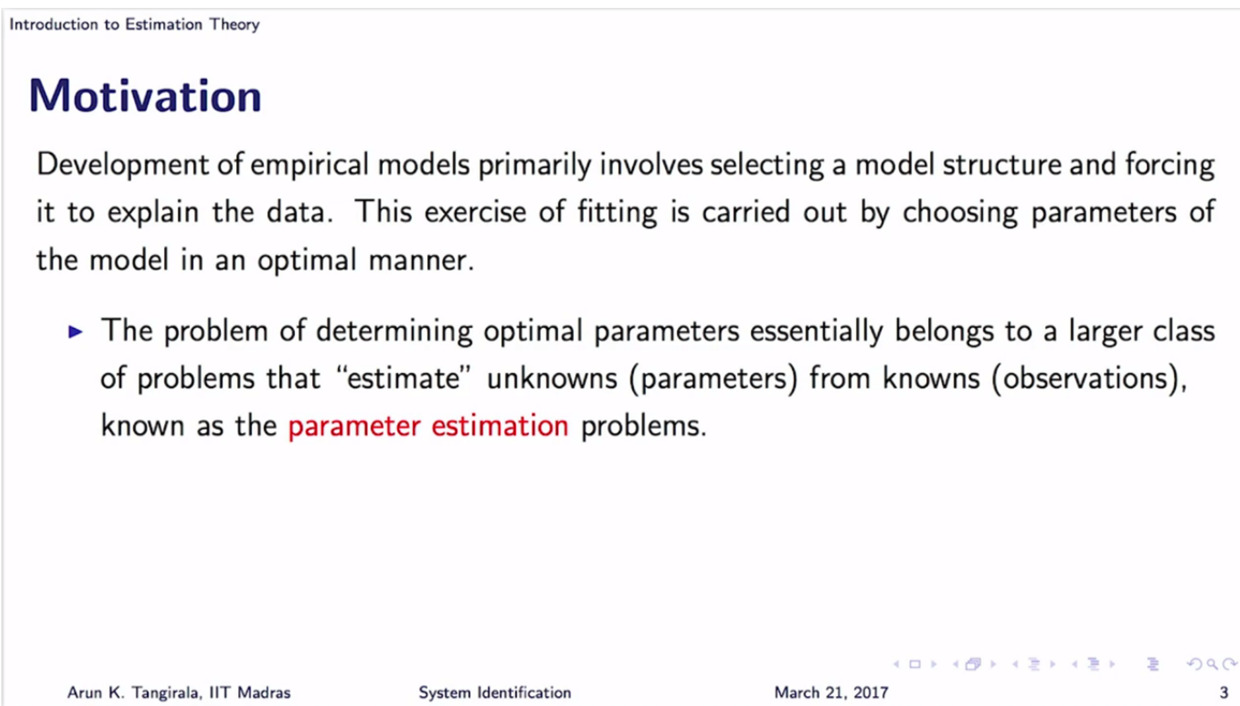


So we have come to a stage where we have chosen noise models, plant models we have learned how to represent LTI systems, deterministic place scholastic, what are the different assumptions involved quasi stationarity when I bring them together given model very importantly we have learned how to compute predictions. And when we are learning predictions we said predictions are useful for two reasons. One, given a model I will compute predictions. Two is given data I will estimate model by minimizing the prediction errors. So it is time for estimation now which is what we have been waiting for. So we are not transiting from theory to practice reality but for this reality there is a fantastic theory. Although we are going to estimate it doesn't mean that I can do a hotchpotch of it or some jugad about it and nothing, you can't do anything that you want. You can but you won't be able to infer anything meaningfulness necessarily. You can't make any conclusive statements. Estimation theory is this beautiful theory. It's very rich. It's probably at least as far as the literature is concerned, written literature is concerned it's about 200 to 300 years old with lot of contributions coming in from the last 150 years or so. Of course [00:01:48] is always there, it's about 200 years old.

So let's understand quickly what is estimation after all. We have been thinking that estimation is all about just estimating some parameters. But is a lot more. Alright. And we learn for the rest of the course the big chunk, the bulk of it that is remaining is estimation only.

So let's look at this estimation thing closely and understand what is it?



At the heart of any estimation problem is an optimization problem. That's the first thing you should remember. Any estimation problem is an optimization problem. And it is – and the so called goodness of the estimate, how good your estimate is. See suppose I ask you what is the width of this room. You are going to make a guess. That is also an estimate. Are you solving an optimization problem? We don't know. We have to ask the brain. It's a very difficult thing to understand the brain. But we are making a guess. We call that as guestimation. Okay. It's not really in the true sense estimate. But as an engineer if I ask you what is the width of this room you can't say 40 mm and so on. You have to, guess 40 mm maybe I don't know, two fingers can fit in.

So you have to give a reasonable guess. But believe me people have given weird answers in the past. I am not going to take a risk to guess the width of the room. But you are suppose to be good at that. That is also an estimate. But here the estimation that we are talking about is explicitly formulating a mathematical problem, specifically an optimization problem and solving it. The story doesn't end there. There is some more to it which is very important. And I will talk about that.

So predominately if you look at the estimation problems there are three types of estimation problems. Parameter estimation, signal estimation, and state estimation. So I am going to erase all of this because we are on a completely different ground now. So what is parameter estimation?

As the name says I am going to estimate the parameters of something. So parameters of what? Parameters of a model or parameters of a PDF. So if you look at historically, the parameter estimation problem began with the estimation of parameters of a PDF. What I mean by parameters of a PDF is if you have Gaussian PDF what are the parameters? Mu and sigma. If I take a uniform distribution, what are the parameters? A and B. The interval and so on like exponential, lambda and so on.

So historically it began with that. And then gradually models started to come in when people started to look at prediction problems until it's all about only random variables not about random signals. Now today parameter estimation is generally estimating any parameter. Then you have signal estimation problem. What is the signal estimation problem? I am given measurement of a signal. We know already that any measurement will consist of signal plus some noise. And the task is to recover that signal from measurement in an optimal. Guess the signal, estimate the signal. What is estimation? Anything that has got to do with recovering unknowns from knowns. As simple as it sounds. And I am always reminded of what my elder brothers used to tell me when I was going to school. He used to tell me I will tell you one simple thing, mantra to remember no how to write your answer sheet. If the question is short expect the answer to be long. If the question is long, expect the answer to be short because a lot of information has already been provided. And mostly it is correct.

In general itself, when you make very simple and [00:06:31] statements a lot of thing has not been said. That means it remains to be discovered and learned. But when a lot has been said then there is not much to be really learned in the sense on your own. So here I just said estimation is a problem of intelligently discovering the unknown or recovering the unknown given knowns. That's all. It's a very simple statement but beneath the simple statement is an ocean. And in statistics is also statistical inferencing, that is one field if you want to study.

So coming back to signal estimation that is what it is, very popular among signal processing. So historically parameter estimation preceded signal estimation.

Then you have state estimation. We all know what state estimation is. we have talked about state space models and I briefly mentioned what state estimation is. I am given model, I am given outputs, sometimes inputs if there is an input and I am suppose to estimate the states, those hidden ones. Again, historically, this is the latest. Although it is still about 50-60 years old it is still latest compared to the previous two.

Signal estimation

Goal: To estimate the signal(s) from the measurements.

Given measurements $\{\mathbf{Z}[0], \mathbf{Z}[1], \cdots, \mathbf{Z}[N-1]\}$, $\quad \mathbf{Z}[.] \in \mathbb{R}^m$ and a *dynamical* model,

$$\mathbf{Z}[k+1] = \Phi(\mathbf{x}[k], \mathbf{Z}[k], \mathbf{v}[k])$$

estimate the signal $\mathbf{x}[k] \in \mathbb{R}^p$ and the properties of the stochastic signal $\mathbf{v}[k] \in \mathbb{R}^m$. In addition, the random component $\mathbf{v}[k]$ is assumed to have a pdf:

$$\mathbf{v}[k] \sim f(\mathbf{v}; \xi)$$

where ξ is the vector of parameters characterizing the probability density $f(.)$.

The information set could also include possible input actions $\mathbf{u}[k]$.

Now having classified the estimation problems into these three there are two points that you have to remember. One that there are sub-classifications. So for example, in signal estimation you have for example, you have prediction problem. They are all just mathematics. So for example, this is the mathematical formulation of a signal estimation problem. I have just given this to you for information say. It's not really needed for this course. But just given you very generic formulation.

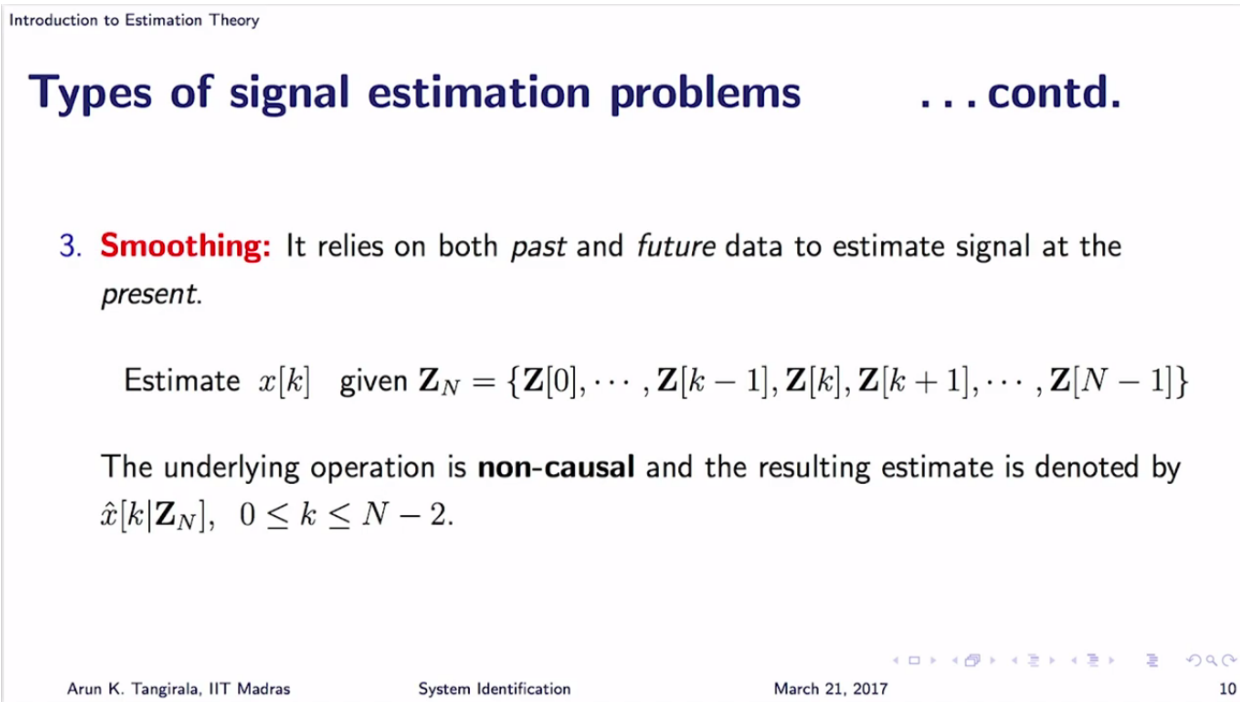
Types of signal estimation problems

It is useful to classify the large class of signal estimation problems into three categories based on the times of available information and when we wish to estimate:

1. **Prediction:** Information is available up to k and future values of the signal $x[k+1], x[k+2], \cdots$ are of interest.

$$\text{Estimate } x[k+1], x[k+2], \cdots \qquad \text{given } \{\mathbf{Z}[0], \mathbf{Z}[1], \cdots, \mathbf{Z}[k]\}$$

So within signal estimation you have a prediction problem which we have discussed at length, given information upto k, predict what happens at k+1, k+2. That's also an estimation problem.

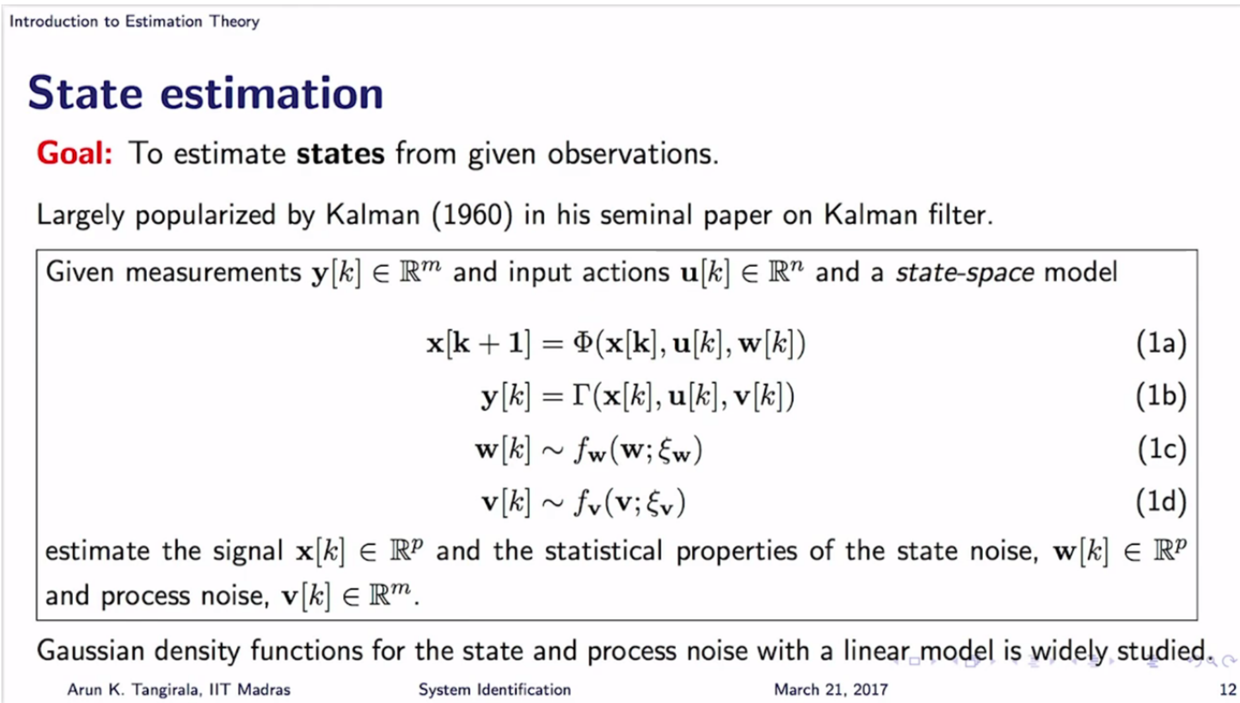


Then you have filtering. We have spoken about it. Given information upto k guess the truth at k. And then smoothing. So that's the third problem. Given information at k-1, k, k+1, k-2, k+2 and so on, get the estimate at k. In all of these it is assumed that the truth is not being given. The corrected form of truth is being given. And what makes estimation challenging is the scholastic nature of the process and the data. And that is uncertainty in the information and the uncertainty in the process knowledge makes estimation problem extremely challenging and very exciting and interesting.

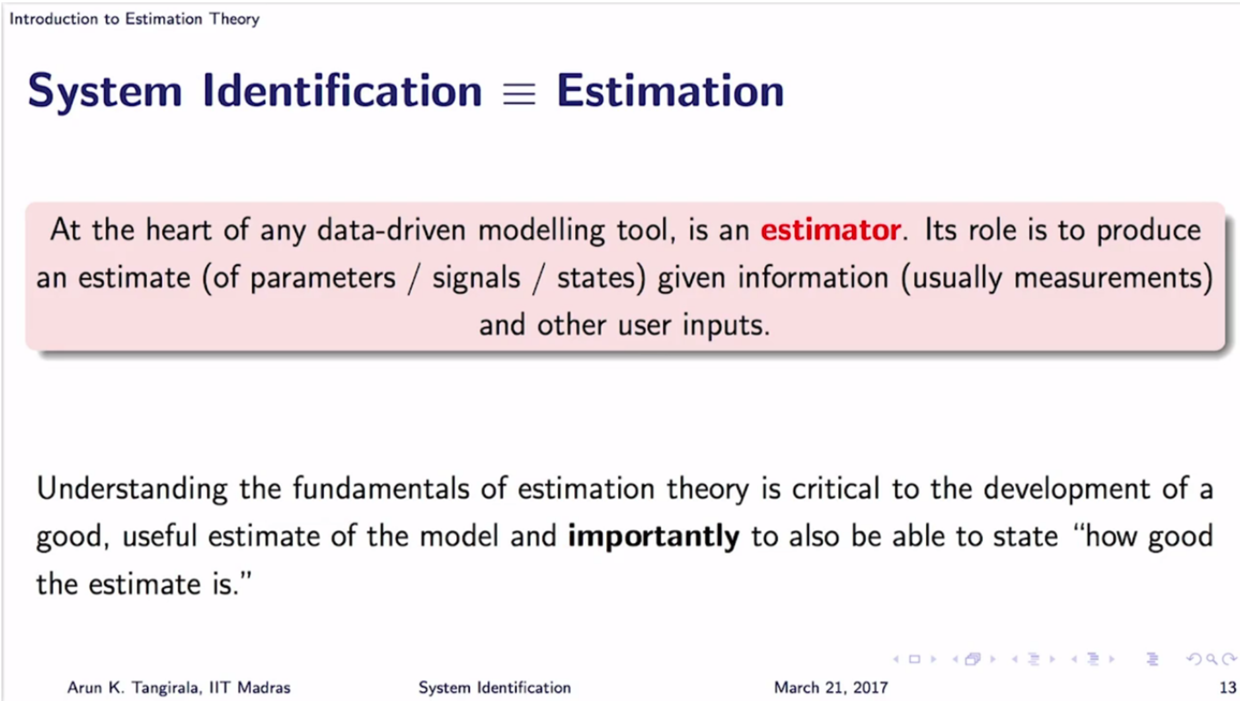
In the deterministic world, there is nothing to worry about because it is deterministic. Once I know I know it for all. There is nothing to really question what is – whether estimate is accurate and so on. There is nothing to worry about. It is only the scholastic nature, the random nature of the process and the data that makes estimation problem challenging. So these are sub-classifications of signal estimation.

Now what happens is although we have said there are these three signal parameter and state estimation, it turns out they are all equivalent. I will talk about that with an example. So the signal estimation problem can be posed as a parameter estimation problem which can also be posed as a state estimation problem. So then why this divisions? Because historically they are developed independently, number one. Number two, the applications are different. A statistician would not be bothered about signal estimation whereas the signal processing person maybe worried about all the three. And in some applications I maybe only estimating parameters, no signals, no states.

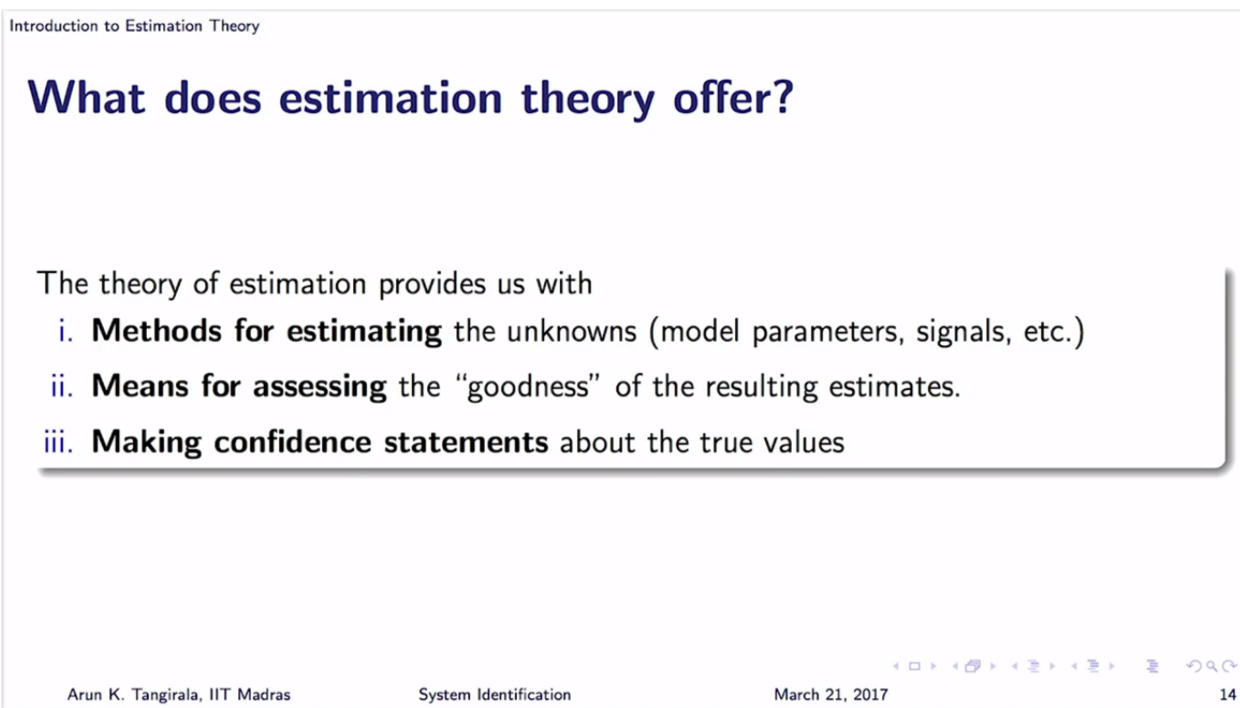
So the bank of methods that I have, the library of methods that I have for parameter estimation maybe very well tailored to parameter estimation. And not so much to signal estimation. So although theoretically they are all equivalent, it's like our models. We say difference equation form is the same as states, space, as convolution and so on. Yet they are different in some sense. So that is how it is.



I have already spoken about parameter estimation. There are two types of parameter estimation problems; model parameter estimation problem, PDF parameter estimation. Then you have the state estimation problem which became very popular after Kalman seminal work who showed Y state-space models are very useful and how to estimate states given measurements in an optimal way. We will talk about this briefly later on.



Now sys ID at it's heart is estimation. Without the estimation algorithm your sys ID kind of crumbles. Therefore, you should be very well versed with the basic theory of estimation.



What does estimation theory offer? We said already, estimation theory tells us how to estimate intelligently. Is that all to it? No. There are two more things. One, how good is my estimate? Very important. So if I ask you to guess the width of this room, what is your answer. I will take the risk? Six meters. Okay. Your guess. Eight meters. You have to tell me the units otherwise I can put in nanometers also. Eight meters. Any other guess? Sorry. 8 to 9 meters. Okay. Interesting. Political.

Okay. Want to be safe. But very interesting. So when I ask you you are giving me some different numbers. And he says 8 to 9, all of you have given single numbers. These single numbers that you have given me are called point estimates in estimation theory. And what he has given me is inter-estimate. Both kind of estimates exists in literature. One is a point estimator, other one is an interval estimator. But that point apart, if I ask you how good is your estimate? If I ask you again, what is the guess now? Suppose I give you another room of the same width, truth is the same. And I ask you to guess. Is there a chance that you will give me different answer? Likely right. Because there are all these things that go around Whatsapp, Facebook blah, blah, look at the magic you will be cheated all these. So I can do all of that.

So if I give you – if I present you with a room of the same width I know I have constructed it, and ask you to guess, you may give me a different guess. It's likely. So which means your answers can change. And in estimation we have already said randomness is going to be present. So as I am going to present different data records of the same process I am going to get different estimates. The process is the same at a statistical level but because of the randomness the data record is going to change numerically, estimate is also going to change numerically.

As a result I have to worry about the goodness of the estimate in the sense of variability, how is your estimate going to vary with the realization.

Not only that I know that my point estimate is not going to be accurate. That is one thing that we have to remember in estimation theory. Whatever estimate I construct from finite data, it's not going to be the truth. But if I am presented with all the realizations and I collect all the estimates corresponding to the realization maybe in some sense if I average I should get the truth. That is another feature I want and so on.

So these are called properties of the estimator. Beyond this, you also have to worry about the quality of data. See, you have to understand in estimation we are going from data space to unknown space. This is the space of parameters let us say. It's easy to study the parameter estimation problem to begin with. So I have to go from data which we will denote by Z to theta. They rhyme but that rhyming doesn't help much. Okay. estimation is all about going from this space to the other space in an optimal way. So the first thing I need is a bridge here. I need a flyover, not the one that I have in front of IIT, it doesn't serve too much of a purpose. But I need a bridge that takes me from the data space to the theta space. This is an unknown space and I don't even know where the truth is. Maybe the truth is here. I don't know where the truth is.

So I am always reminded of this where you see in Bollywood movies this dejected hero after the girlfriend has ditched him. He gets on to a train and he just TC comes and kaha jana hai. Jaha bhi le jao. Mera manzil mujhe malum nahi hai. So that is the problem in estimation. That's exactly the problem but it is much better than that. you don't have to grow a beard. You don't have to wear a shawl and so on.

So seriously. The beauty in estimation theory is without knowing this theta naught you can still get as close as possible to theta naught. You will never be – then you can ask questions how close I am, if I give you infinite samples will you reach the truth; all those questions can be asked. But still with finite samples I can get pretty close to theta naught. Of course there is a limitation. As

N increases you can get better, as the data size increases. This bridge is what we call as model. This model tells me how the parameters are entering the data. That is just a postulate. I am going to postulate it. See, ultimately if I have to recover theta from data I need a relation between them. That model is a mapping. And the mode accurate your model is the more – the better your estimate is going to be. That is closer you are going to be to the truth. In sys ID for example I have G and H which will tell me how the parameters enter. How the parameters and the data are related. The data itself is being generated independent of theta. That is the process. But I suppose that this is how the parameters and data are related. So this model is a key.

Then once you have this model you enter the space of theta and then you start your search for theta naught. So you need a guide. This guide has to – this guide also doesn't know where the theta naught is. But the guide will take you very close to theta naught if chosen carefully. That guide we call as the objective function. That's the criterion of it depending on what I say because there are many solutions. I have to pick the one that is optimal. Without knowing the truth I have to pick and we can still do it. So when we look at the example tomorrow we will understand.

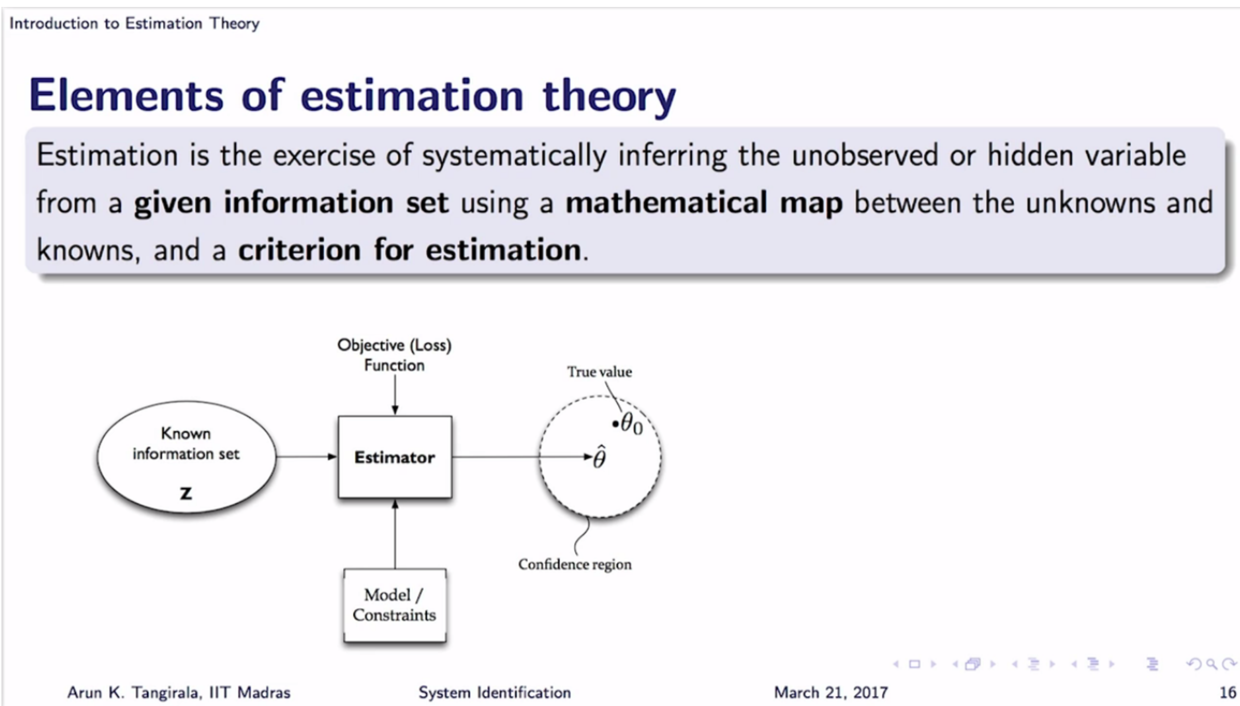
Anyway, so coming back to the problem, point here, not only does estimation theory tell me how to estimate but also tells me how to assist the goodness of the estimate. Thirdly, it will allow me to make some confident statements about the truth. What I mean by this is earlier we talked about the width of this room. So you said eight meters. Correct. Is that correct. Six meters. Okay. Now that's a point estimate. So I were to ask you what is an interval in which the truth could be? What would be the interval? Six to eight. Okay. Are you 100% confident or there is some uncertainty there? There is chance that the truth can be outside this. Correct.

So I cannot make 100% confident statement. Of course, you can say that it can be anywhere between one meter and 100 meters just to make sure the truth is within that in that interval. But what have you sacrificed? The width. The interval is too big. So if I want to make a very confident statement then I start increase the width of the interval which is not good. What I want is I want a narrow interval with high degree of confidence. So for example if you were to say I am 99% confident that the width of this room is between 6 and 6.1 meters. Kya baat hai. Very good. Okay. But to achieve that is not easy. So in a given estimation problem how do I first even construct a confident interval like this? Ultimately, we have to report that. There is no point in just giving a point estimate. The estimation exercise ends only when you have made some confident statement about the truth which means many at times we start with the point estimation problem and end with an interval estimation problem. This has been the classical approach for decades. In fact for maybe even more than one century. You start with the point estimate and then you end with an interval estimation. Base in estimation differs from this classical approach it directly gives you the interval everything whatever you want. It is not considered a classical estimator. It's a modern estimator. It has it's own advantages. But it has its own challenges. Anyway, so the points that you want to remember is estimation theory tells you number one, methods for estimating. Two, metrics for accessing the goodness of estimator as well as data. Your data has to be informative.

So what I was trying to say earlier is here if you were to just schematically draw data goes through an estimator and outcomes to theta hat or maybe the confidence interval. So what governs how good this theta is this data and this estimator. So you can think of data as the food, and estimator as the digester. Without data there is no estimation. So the food is important. Then this estimator processes that information, digest that information, and gives you some useful theta hat.

So when you think of the human body or any living being when you consume food, you are consuming food for generating energy. But suppose after consuming food my energy levels have gone down, it's worse. Then the problem is either the food and/ or the digester. Maybe the food was good, my digestive system has some problem. Correct.

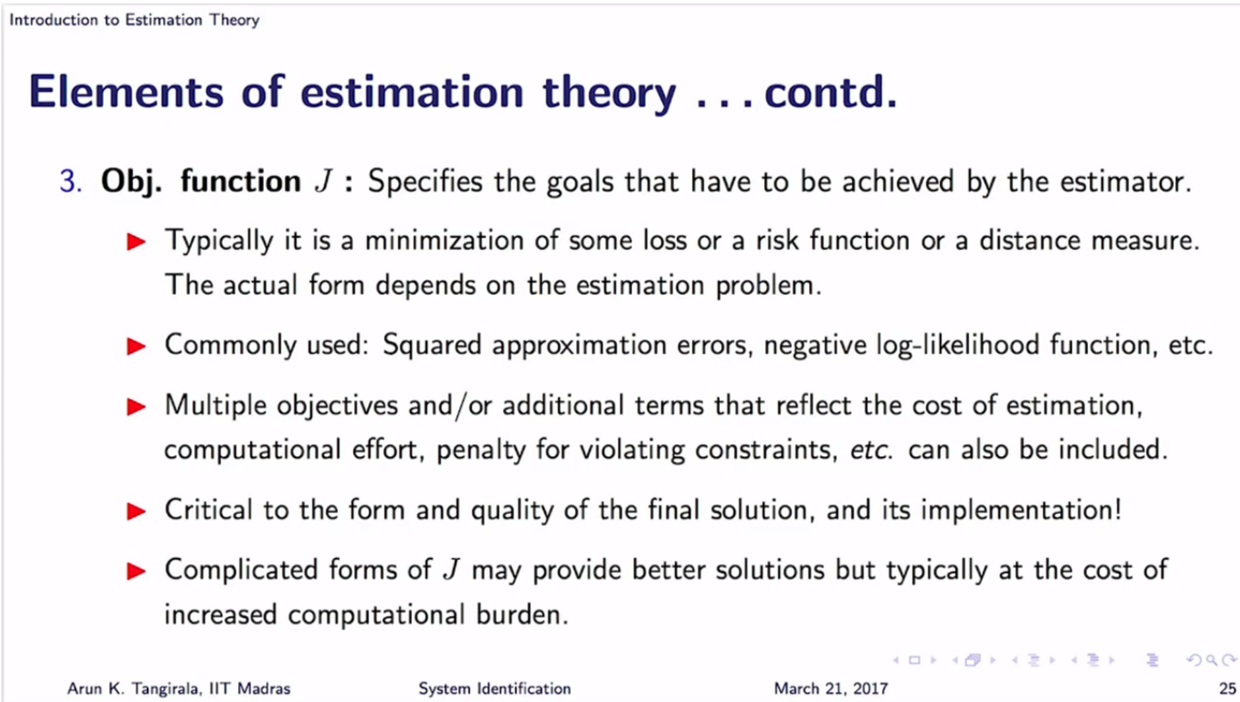
So estimation theory allows you, in fact gives you the theory of accessing the goodness of this estimator qualifying the goodness of this as well as telling you how informative the data is. More the information in the data, information is vague at this moment, but qualitatively more rich the data is with respect to the information. Suppose I give you the duster and I say I will give you the dimensions of this duster and ask you to guess the width of this room. Will that help? Is that data useful? So there is no information there. But suppose I give you this and I ask you for example to estimate how long it takes to erase the board of a particular size, maybe that is useful. So depending on the information contained in the data they are all relative. Information and the data and the theta they are all relative. The goodness of theta hat is affected. So this information contained in the data how informative it is was quantified at least literature initially by Fisher where we come across Fisher's information. And we will talk about that tomorrow.



So let me conclude by this diagram here today's lecture. Duster is there. Entire world is made up panchabhutas. You have here five elements. One is the data. In any estimation problem there are five elements. One is the data. Two is your model. You can think of this model as a model or additional constraints for example I know that theta has to be greater than zero and so on. So all those constitute the model. Then you have the objective function as I said which will it will act like a guide and take you close to theta naught. It's like optimal estimate. It will give you an optimal estimate.

Then you have the estimator itself. Well actually the estimator is not a separate entity. A estimator is a consolidation of your model, objective function put together but then sometimes there can be computational sites that means how you implemented in a computer. Those things also go into the estimator.

Somethings can take much longer to implement. So you may have to write a computation efficient algorithm. That goes into estimator and then finally you have the estimate itself. It's the fifth element. Once these four come together the child is born, theta hat.



Elements of estimation theory . . . contd.

4. **Estimator:** Essentially the mathematical device or expression that computes the estimate using the information $\mathbf{Z}$, the model $\mathcal{M}$ and the objective function J.
   - An estimator is also a *filter*. It "filters" the true solution from the given information. **Example:** Wiener and Kalman filters (estimators of signals / states)
   - As in the case of filters, estimators can be of various types - causal / non-causal, linear / non-linear, time-invariant / adaptive and so on.
   - Linear estimators are preferred to non-linear ones because of ease of implementation. However, the price paid may be the inefficiency and / or lack of *robustness*.
   - **Form of the estimator is dictated by the optimization problem.** An *a priori* form along with a criterion of estimation can be employed as well.

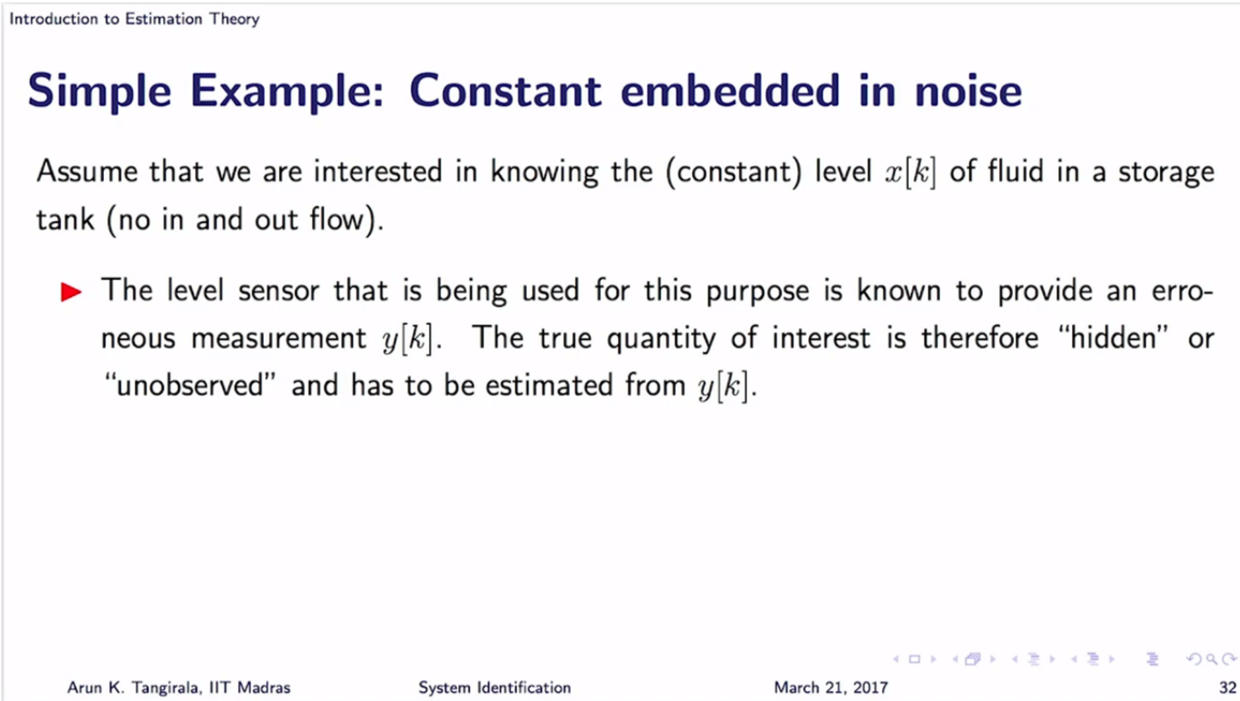
Introduction to Estimation Theory

Elements of estimation theory . . . contd.

5. **Estimate:** This is the final quantity of interest, that is produced by the estimator. Conventionally, the parameter or the unknown to be estimated is denoted by $\boldsymbol{\theta}$ (could be a scalar or a vector), while the estimate itself is denoted by a hat, $\hat{\boldsymbol{\theta}}$.

$$\hat{\boldsymbol{\theta}} = g(\mathbf{Z})$$

where the function $g(\mathbf{Z})$ is known as the *estimator function*.
Needless to state, $g(\mathbf{Z})$ implicitly depends on J and the model $\mathcal{M}$.
Note: It is also conventional to have the same notation for the estimate and the estimator, *i.e.*, $\hat{\theta}$. The actual reference is understood based on the context.



So these are the five ingredients. What we will do tomorrow is we have spoken about all of these. In fact, tomorrow I will take an example and then discuss all this elements, example that we will take up is a constant signal embedded in noise, how do you discover this constant signal. It's a signal estimation problem. You can also think of it as a parameter estimation problem. Later on we will show you can think of it as a state estimation problem and so on.

So it's a very simple example which actually serves all estimation problems. Okay. Thank you.