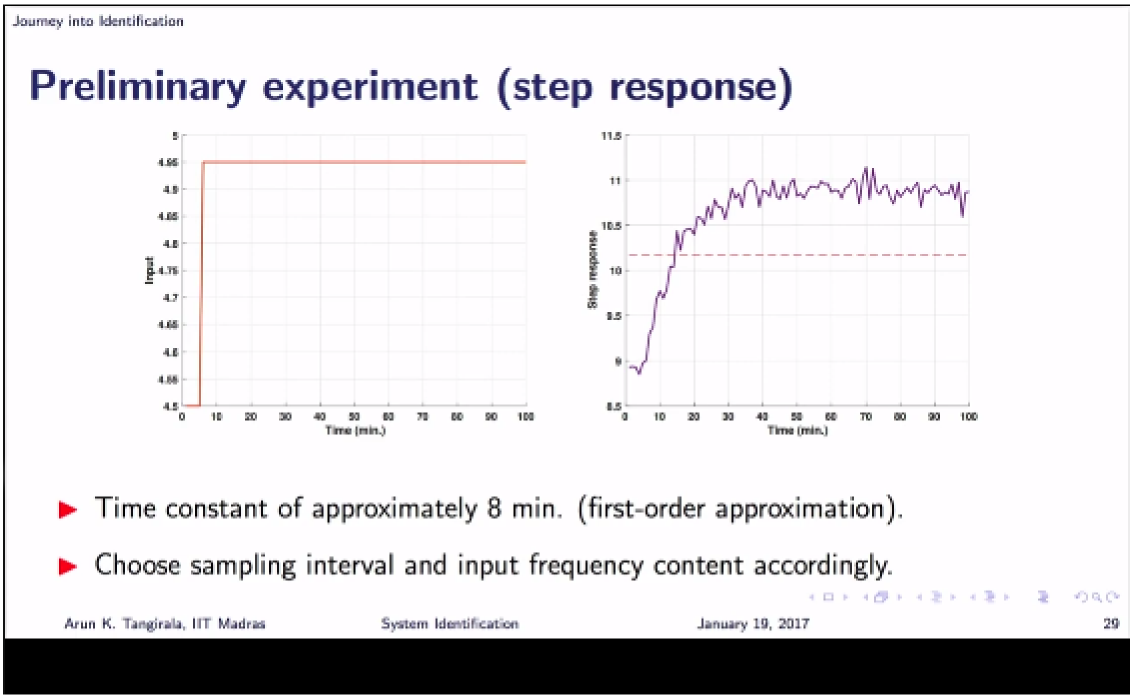
Ch5230: System Identification

Journey into Identification

(Case Studies) 9

Okay, good morning. What will do is we'll, of course, go through the liquid level case study in detail today. One of the messages that you should take from this case study, there are many things to take with you. But one of the highlights of this case study is that if you make a wrong assumption on the noise model, which we have not yet gone gotten into, but at the end of the case study you will realize this. If we make a wrong assumption on the noise model, it can have a serious influence on the deterministic model. That is the accuracy of the deterministic model and then, of course, a precision also, but that will come in due course of time.

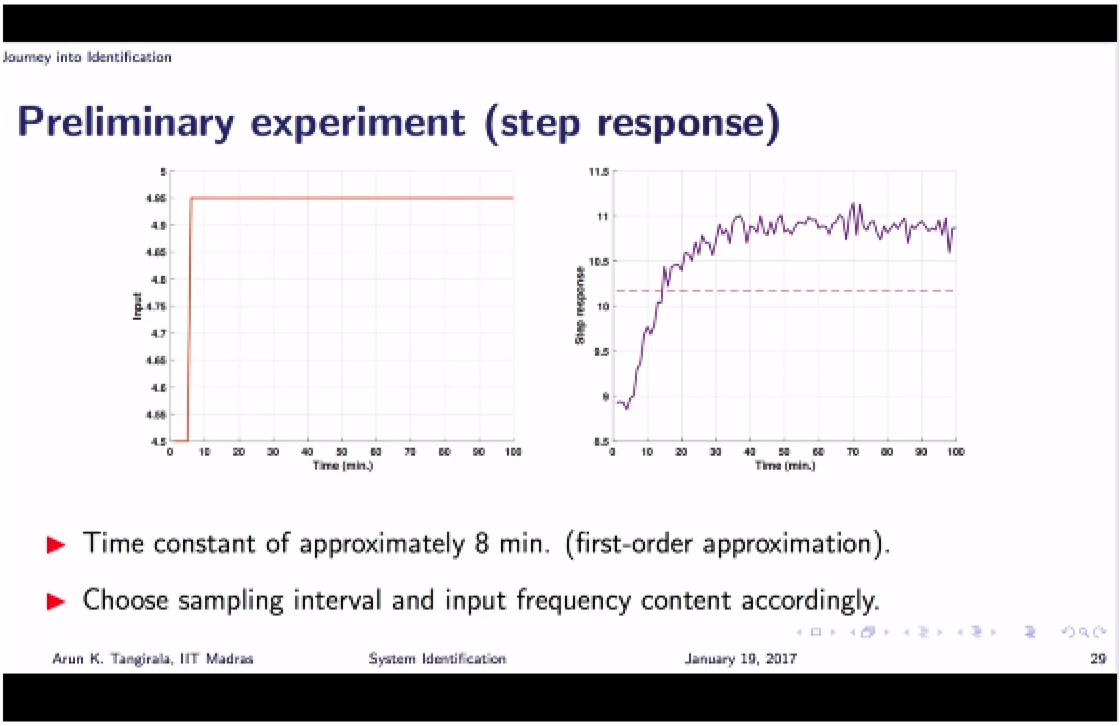
(Refer Slide Time: 1:00)



Let's get going here. The point where we ended our discussion yesterday is at the step response. We were, we said here now we have a step response that comes out of a preliminary experiment. And also explain in briefly as to how you go about deciding the magnitude of the step change that you would introduce in the input.

Now assume that the 10% change that I have introduced, still keeps the process in the linear regime that means a linear, a single linear approximation will still work. Let's not worry about that now. The more important point in discussion is the time constant that we identify from this type response. So as I said, if you look at the graph carefully, the step change was introduced at 6 minutes.

(Refer Slide Time: 1:55)



And all of us are familiar mostly, most of us that the time constant of a first order system.Why do I assume a first order system? When I look at the step response, one of the first inferences that I draw is that it seems to have a first order dynamics. Now, mind you, that over time systems can also have this kind of a step response. But nothing given to us we can assume a first order dynamics and with that first order approximation, we know that the time constant is a time at which the response rates are 63.2% of the final change, although we say final value, it's actually the final change and that has been indicated here by a red dashed line.

So if you were to zoom in and so on, and you will find that roughly the time constant is 14 minutes, sorry, 8 minutes, right. The change occurs at 14 but we introduce a change at the input change at 6 minutes. So the time constant is 8 minutes.

Now this time constant is a very valuable piece of information for two reasons: one to make a decision on the sampling interval. I'd ask you to think about it. The time constant is useful for making two important decisions, or it has at least two important roles to play. One, it allows us to make a decent decision on the choice of sampling interval, reasonable choice one can make and secondly it allows us to figure out what frequencies should be introduced in the input. So let's discuss the first aspect. Why is the sampling interval related to the time constant? Any idea? Why is a choice of sampling interval related to the time constant?

[3:57 inaudible]

Okay. But what is the connection between sampling interval and time constant?

Maximum response occurs during a [04:08 inaudible] during sample [4:08 inaudible] So, if that is the maximum frequency no other changes can happen so rapidly, the sampling period should be related to time constant, if it is less than that we will lose some of the information.

Okay. I think we need to explain in more layman's terms. Although, your answer has most of the message in it. Any other answer,what could be the connection between time constant and sampling interval?

It is the time constant is [4:36 inaudible]. Those are sampling intervals [4:40 inaudible] you're about to move and [4:43 inaudible] which people determine how the system change.

Right.

That is so.

Can we put this in a more simple way? If the process dynamics, if the process changes slowly, then obviously we can afford to observe it slowly, right? And time constant is a measure of how fast the processes changing in the sense process changing in the sense the response is changing the time to a change in the input. Obviously, if the time constants are low, a small then the process reaches a steady state very quickly. And obviously then, I have to observe fast. So sampling interval has to be small. We are not saying that the sampling interval equals a time constant. We have not said that we're just saying that there is a relation between the time constant and sampling interval and that's a direct proportion. If the time constant is large, then the process has slow dynamics, then obviously, we can afford to observe slowly. In general, the sampling interval that we choose for any process is dependent on the nature of the dynamic. So if I take, let's say atmospheric temperature. And I want to observe the outside temperature. So I take a sensor with me stand outside. And I have to decide I have a data position system. And I have to decide on the sampling interval.Would I choose to observe every second? No. Why? Because we don't expect significant temperature changes to be occurring on that scale. Maybe we choose to observe half a minute or a minute, sometimes even 5 minutes. It depends on the time of the day, how quickly the temperature is changing.

On the other hand, suppose I'm looking at a moving car and I want to sense the displacement, can I afford to actually observe every minute. Suppose, I'm standing here and I'm observing the cars moment I'm standing anyway, doesn't matter. So I started at time zero, now I have observed its displacement. From the time go, the car starts moving and I have to observe its position. Can I afford to actually observe this car position every minute? It would have actually travelled a lot and a lot of dynamics would be missed out. Some maybe critical pieces of information would be lost. There maybe I have to observe a second or even milliseconds, or even half a second.

Again, that takes us back to the same point that sampling intervals have to be chosen commensurate with the dynamics of the process. Now let me also tell you that as much as this sounds alien to some of you, it isn't. All of us have gone through courses on numerical integration, right? At some point or the other we have taken courses title computation techniques or something that tells basically that tells you how to numerically integrate differential equations.

What is the key idea in numerical integration? What is the basic principle?

Dividing into smaller intervals.

Dividing what in the small intervals?

In curve.

Which curve?

The curve of interest.

Thank you. What is the curve of interest? That's excellent. That will still keep you in the interview room.You have to be more clear. What is this curve that we're talking about? Solution, the response Right? The solution. So we divide into smaller segments in a sense in amplitude or in time? In time. And then what do we do?

Sorry. Okay. So we integrate over small steps, right? With divide, not the response, but actually the time, profile that the time over which the response is evolving that is what we divide into smaller steps. And then integrate, make some approximations over that small step size, small steps and then put together everything. So important decision that one has to make in numerical integration is the choice of step size, right? That's exactly the choice of sampling interval. Is the same thing, there's nothing. In signal processing, it's called sampling interval and numerical integration, its called step size. Again the story is same there, of course, the advantage there is you can use fix step size like and the analogy of fix step size and numerical integration is uniform sampling interval, in signal processing.

But in numerical integration, there is also another strategy, which is variable stepsize, used by [09:48 inaudible] methods and so on, right? Many methods can use that.What is the idea behind variable step size? What is the idea? Why can't they use a fix step size? Why don't I want to use a fix step size?

Okay. Good. Cost computation. So what is variable step size? What is the principle behind this variable step size?

[10:16 inaudible]

Yeah, you take larger steps. So as your response is approaching flatter regions, you can afford to take larger steps, because you wouldn't have missed out someone. But where there are very quick changes, you have to take very small step sizes. And that's the idea behind variables steps. Again, the principle there is the same, you're looking at the local dynamics, but here we are not looking at the local dynamics, we are seeing will for implementation reasons, we'll do the fix sampling internal, its lot easier for the instrumentation, right? There are also in signal processing variable sampling intervals and so, we don't get into that. But for implementation purposes, it's a lot easier to stick to a fixed sampling interval and obviously, the one that governs a fixed sampling interval is the dynamics that is a transition phase. And that's exactly the idea. So in general, there is a rule of thumb, of course, we'll revisit this concept but I'm just giving you upfront, a rule of thumb for choosing sampling interval which is one-fifth to one-tenth, don't think of this as 1 or 5 minus 1 over 10. Okay. Maybe, I'll write it in words. Of the time constanttau, if you were to denote the time constant by tau. And this is of course valid for first order approximations. There are other guidelines available for under damn systems, which is again based on rise time. For first order systems we don't speak of rise time, right. So here what we shall choose is one-eight, so that the number sampling interval is a sweet number one.

Okay, we don't run into it doesn't matter. I mean, don't think that decimal values of sampling interval is a problem and so on. It's just easy that's all. So that when we plot the discrete time signal as a function of the sampling index, you can as well treat it as time. That's the idea, okay. So the sampling interval that we set now is 1. What is 1 here 1 minute, because the time tau is has the time units of minutes. Okay, that kind of settles the first aspect of the point, 2 points that I made earlier. What is the second point that we said? The time constant helps us in input design, right, in choosing what frequencies to be pumped into the system into the input. Should the input, should the system be excited by rapid fluctuations or low fluctuations or lower or medium? What kind of frequencies should be injected into the input? This question can be answered largely by knowing the time constant. Now, there was another condition that we had said couple of lectures ago, on the input design that is, we said there is some condition that the input has to satisfy. And what was that? If you want to identify unique model, so it should be sufficiently or persistently exciting, right? And later on in this course, we will learn the theoretical definition of persistently exciting signals. But what that means essentially is all frequencies should be present in the input. That means you should have long answer questions, you should have a short answer questions. You should a medium answer questions, you should have rapid fire questions. Remember, whenever you are introducing a change in the input, you're asking a question to the process. And when you're changing the input very rapidly, you're not giving enough time for the process to settle down. It's like a rapid fire round and quizzes, right? In many of the quiz competitions that you see, there is a rapid fire round and then there is a no time given 10 seconds and so on. It's exactly the same in input design, you will see large, I mean, wide pulses or, no wide duration pulses, small duration pulses, and very rapidly changing fluctuations and so on.

So, the question is, should I include all of them? One theory says that the input should be persistently exciting, right? With the help of the example, we came to that qualitative conclusion. But now we are saying the time constant will help me decide what frequencies to be included in the input.Which means that we are hinting at not including all frequencies. On one hand, we are saying all frequency should be present that is theory. But then there is another concept that comes forward and says, look, you may not have to include all frequencies. Why are we saying this? Now think again, of an interview process, could be a PhD interview, a lot of you have gone through that, or some interview that you've gone through. You are appearing for some position, it could be PhD position, or it could be a job and so on. So there is a certain context and also as a candidate, you have certain calibre, I mean, you one cannot know everything in this world. It's not possible.

So there are certain domains where you have expertise and there are certain domains that you don't. The domain in which you have expertise and I asked you a question and you respond, we call that as a bandwidth. In technical terms for systems, we say that's a bandwidth of the system. Where I ask a question to the system that means I excite the system is an input it will respond. But like human, systems also have certain out of syllabus kind of questions, right? You remember out of syllabus questions that we use to face in school? Some question comes out of because the board has said the question paper and we have not been taught will say our syllabus merrily you would write it because then you will be credited. You will get some credit on that. It's not your fault. Likewise, systems also have bandwidths. What this means is that they may not respond to all frequencies, all kinds of changes in the input. For example, if you take a large tank, a big liquid level system, a huge cross section area and so on and in you introduce a rapid changes in the input. You will not see that in the outlet flow, the same kind of rapid changes that you have introduced in the input will not be visible in the outlet flow. Why? So system you can say has absorbed or ignored. Either way, basically, the technical term for that is system is filtered out. I'm going to be quiet. We human beings are also like that imagine that we being told repeatedly something. Solve the assignments, solve the assignments, solve the assignments and imagine me telling this. After the first instance itself, even before I begin you probably start ignoring me, but definitely having it listen to me rapidly saying this, very quickly saying this you would certainly ignore, there's no doubt about it, right? You will turn to do better things. That's because we just filtering out, where you can see we're absorbing or we're filtering out whatever it is there is no response.

With human beings it's a lot more, I mean, it is lot more strange and unpredictable things can turn violent also. But systems are kinder they don't turn violent necessarily. They just filter out and then keep quiet, they don't respond. So you may ask what is, then the problem if they don't respond, why are we bothered about it? I'm just going to introduce all kinds of fluctuations, if it doesn't respond, so be it. Not to worry at all. Is that, okay, is that the correct attitude? Is that the correct approach to input design? What do you think?

So I have the theory of persistent excitation says include all kinds of changes, rapid, shot, quickly changing one, slowly changing once, all kinds of changes can be introduced into the input. If the system is filtering out, so be it no problem. What is my problem? And I'm going to get that response. And my algorithm will figure out what is the bandwidth of the system and so on. Is that a reasonable approach to input design or not? What do you think? I say, it's not a reasonable approach. But now you have to think over it. Give me at least two reasons why it's not reasonable to have this kind of a stance, where I include all kinds of changes in input flow, you're presented with the liquid level system. Now, you have to decide what kind of inputs you have to design. First of all, you should be convinced that a step input, we have already spoken about it yesterday, a step input is a weekly exciting one, it gets you some vital statistics, but it may not get you the best model. And therefore, we now want to go to a richer kind of input or better input and this at this stage, now I have to make a decision and I'm saying that it's not a good idea to include all frequencies.

Physically we have to choose [19:49 inaudible].

It's okay, it's feasible.

Again, it [19:52 inaudible] support all times, so,

Okay, to the, up to the physical constraints, no problem. Let's not worry about actuator at the moment, let's say the actuator as ideal one.

Can you give me two reasons why I should not be thinking of injecting all kinds of frequencies into the input? From low to high, high frequencies would mean rapid fluctuations. What do you think could be the problem?

Cost.

Cost, okay. Cost of what?

High frequencies [20:28 inaudible].

Yeah. But if that is going to get you something, then it's okay. Right. I wouldn't know a priority. Suppose a new then yes, cost is an issue, because why would I spend energy on asking a question to which you will not answer, right? That's correct. So cost is one factor that is, this is called the identification effort. Why will I place effort on something that will not yield much? But suppose this is not a major criterion, it's okay. What happens? Why? What happens [21:00 inaudible]? Sorry. I'm sorry.

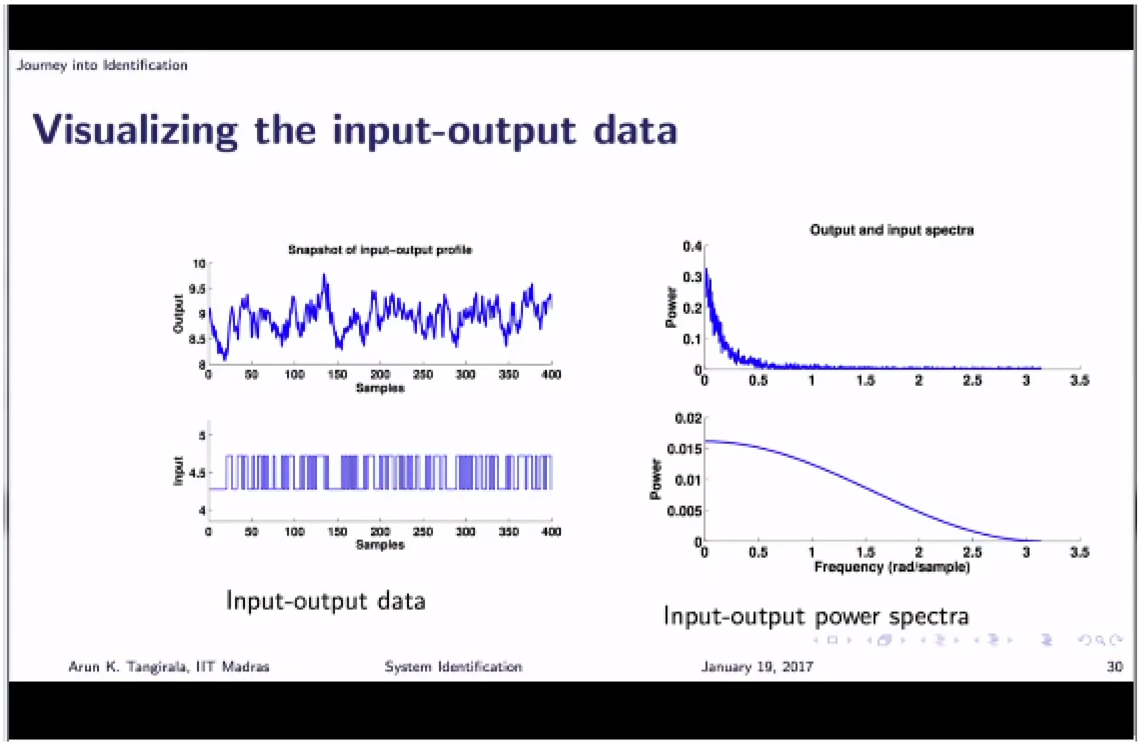
[21:05 inaudible]

So what happens that subsequently? So what he's trying to say is suppose I ask you, who was the father of Sant Tulsidas? Okay. There are smiles. And after that there is silence. No, it's a very good question. But it's out of syllabus for all of us. I also do not know the answer, honestly. So what we hear is noise, right? The AC noise which if lead to a longer period of time can put us to sleep. Now, in data driven modelling, when you are performing experiment, it's important that whatever you collect has mild amounts of noise. That is, remember we want to maintain a good signal to noise ratio, right? So there is that criterion as well. When I ask questions or when I introducing to which the system will not respond, I will end up having noise in my data. For those, corresponding to those input frequencies. As a result, the algorithm wouldn't know the algorithm [may think] this is the response of the process and end up confusing the noise for the dynamics. We don't want that. In other words, we want to maintain good signal to noise ratio at all frequencies as much as possible. And that is the reason why we should not try and, I mean, we should not excite the system at frequencies that are outside its bandwidth. You understand, basically what happens is the system will not respond and what you will end up with is a lot of noise for corresponding those frequencies and that can significantly alter your, the quality of your parameter estimates. We have already seen the effects of signal to noise ratio.

Now, let's put together everything fine. So, we have learned just now at least qualitatively that we should excite the system only within the bandwidth. We should ask questions only within the syllabus of the system. How is time constant going to tell me what is the bandwidth? Once we understand that connection, then we understand the connection between the time constant and the input design. What do you think? How do you think time concert is going to tell? Inverse relationship, so, we will, without going too much into the theory for a continuous time system, first order system, the bandwidth can be calculated by first calculating what is known as a corner frequency. When we talk of frequency response functions later on, I will elaborate on this point. For the first order continuous time system the corner frequencies one over tau radiant per second or by time unit, whatever the time unit is.

From this corner frequency, one can calculate the bandwidth. Roughly, the bandwidth is about the, in fact, depending on how you define bandwidth, but for identification we say that the bandwidth is roughly two and a half to three times this. That for the continuous time system, I can then translate that to the discrete time system and do my calculations, I have done all of that and I'm excited the system. Okay. So now I'm excited the system with what is known as a pseudo random binary sequence and the input profile is shown at the bottom panel of the left plot. Okay. So you see the input on the screen which is at the bottom, you can see wide pulses, narrow pulses, some medium width pulses and so on. That's like a question paper which has short answer questions, medium length, and then long answer questions, and so on. So the process has been excited over a certain frequency range. How do I know what frequency rain has been included? On the right hand side, you see what is known as a spectral plot.

(Refer Slide Time: 25: 10)



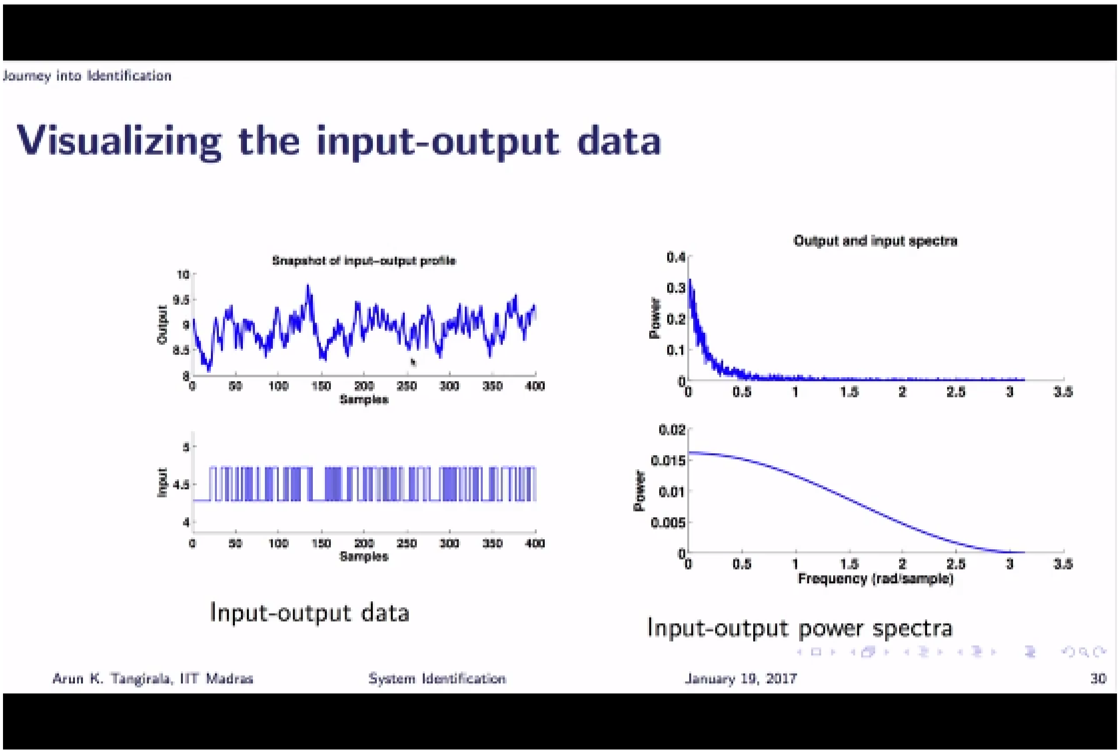
Again, the tech technical definitions of spectrum and so on, we can worry about them later on. Or you can also by the way refer, subscribe to the time series analysis course that's online. It's already live. You can enroll, there's a full semester time series analysis was that I taught, it's available online on the MOOC which talks about, which will then give you all the technical definitions. And, of course, we will review briefly the definition of a power spectrum later on in this course. For now, it's just sufficient to understand what this power spectral block tells us. On the x-axis, in a power spectral plot brought always you would have frequency what I'm showing you is a power spectral density plot.

The difference between power spectrum and power spectrum density, at least in terms of definition is power spectrum densities, power spectrum per unit frequency. Like any other density function. What does the power spectral density plot tell me? What frequencies have contributed to the overall power of this [26:13 inaudible]? So, if you look at the bottom plot on the right side, it tells us it tells you what frequencies have gone into the input. What is the content frequency content of the input and what do we infer that predominantly you have low, middle frequencies and as you go to higher frequencies, the power [26:37 inaudible] is actually low. That means we have designed input in such a way that the input predominantly contains low and medium range frequencies. The right extreme is the high frequency for a discrete time system. What have we done that, because we believe that the system is a low pass filter?

Okay. What do we mean by low pass filter? It preferentially allows or it prefers low frequencies. It responds to low frequency signals much better than high frequencies. And such systems are called low pass filters. Do I know this kind of thing information a priori about a physical system, usually you would end and if you don't, the step response will reveal that for you. Okay, any system of this kind, I mean, where you have first order dynamics and so on is a low pass filter. What this means is, if you introduce a rapid changes into the input, the system will not respond, it's attenuated those inputs changes those input changes are attenuated. Whereas, you will see significant response to small low frequency changes in input or maybe medium frequency changes and so on. So keeping in mind the low pass filter and characteristics of the system, we have designed and input whose frequency content is commensurate, right? It's like I have set a question paper that is within the bandwidth of the syllabus of the system. Now, let's look at the output. So the input is kind of known. Of course, as I said, this input is called a pseudo random binary sequence. It switches randomly between two levels minus 1 and 1.

Of course, I don't see my, I don't show minus 1 and 1 here, minus 1 and 1 is around the average. Essentially it's around the steady state that you are introducing. And you can adjust the change, the levels it may not be minus 1 and 1. It could be minus point 5 and point 5 and so on. But essentially the input switches between two levels randomly, and that is, and because there is an algorithm that generates a sequence it's a pseudo random binary sequence. And we will learn how to generate that later on in MATLAB system identification toolbox you can use id input to generate this sequence. Now on the top here you see the response of the system, right?

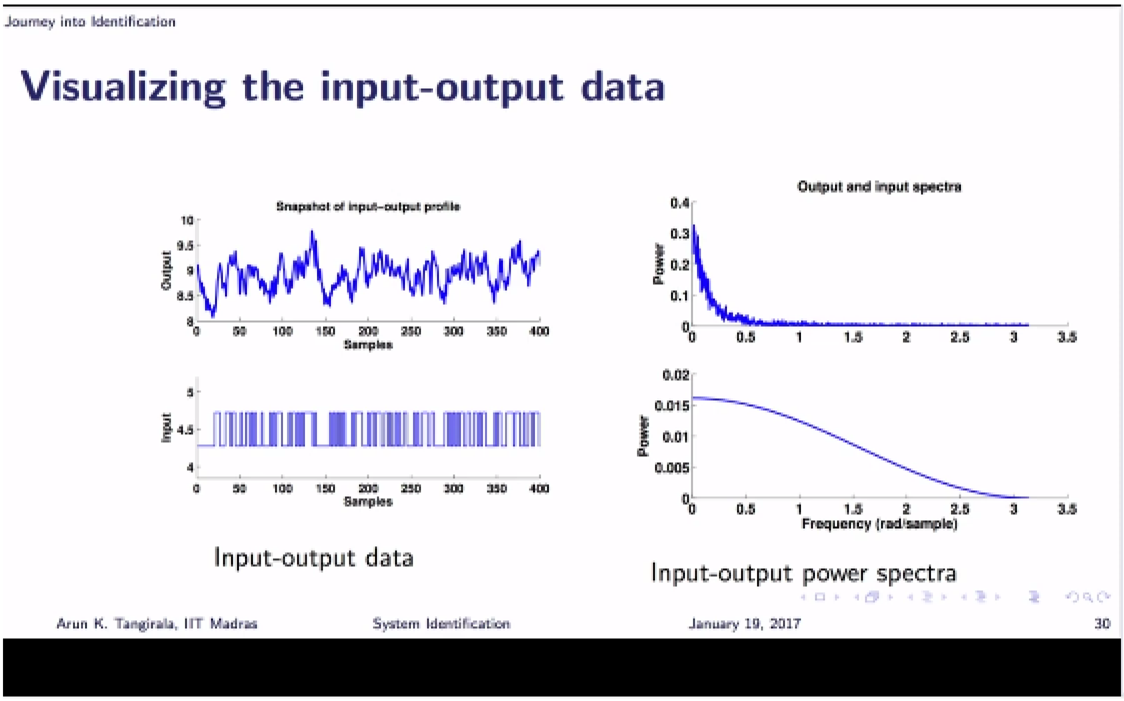
(Refer Slide Time: 29:05)



The response shows no crazy behavior or any peculiar behavior from the system. We, remember we said we want to spend some time with the data. We want to make friends with the data, we want to understand them and what are the kind of things that we try to search for here any presence of any outliers or anomalies in the process, or some time in missing data. A lot of times somebody does the experiment and you are retrieving data from a historian. And sensors can malfunction can miss out recording values you have to search for those. Of course, here we have done the experiment. Therefore we know very well that has no missing data. But we should also search now for trends if there are any linear trends in the process. For example, there are systems where the outputs not only change the way you see on the plot here, but also have an underlying trend, they can grow with time. When such trends are visible, it is important in identification to remove those trends prior to fitting a model. Because those trends are kind of deterministic trends as a function of time, it's always advisable to remove those trends because the theory that has been devised for parameter estimation, assumes so called Quasi-Stationarity. We will again talk about the technical definition of Quasi-Stationarity later on.

So to summarize, when we look at the response, we search for outliers, any anomalies, trends, any peculiar things and at the same time, also get a feel of how fast the system is responding. Occasionally, what I also do is I zoom into certain portions, where I have given wide pulses and see how long the system has taken to reach a steady state, and so on. So it's good time. It's a good idea to spend more time than what we're actually doing here. Okay. Now, the other thing that we want to discuss is the power spectral density plot for the output. What does it tell us which is shown at the right, top right graph. You have again x-axis frequency, y-axis being the power spectral density. What does it tell us with respect to the output?

(Refer Slide Time: 31:30)



Does or put together the power spectral density plots of the output and the input which you see on the right hand side and draw an inference about the system. Can you make an inference about the systems, "Characteristics".

Low pass filter.

Low pass filter. Why?

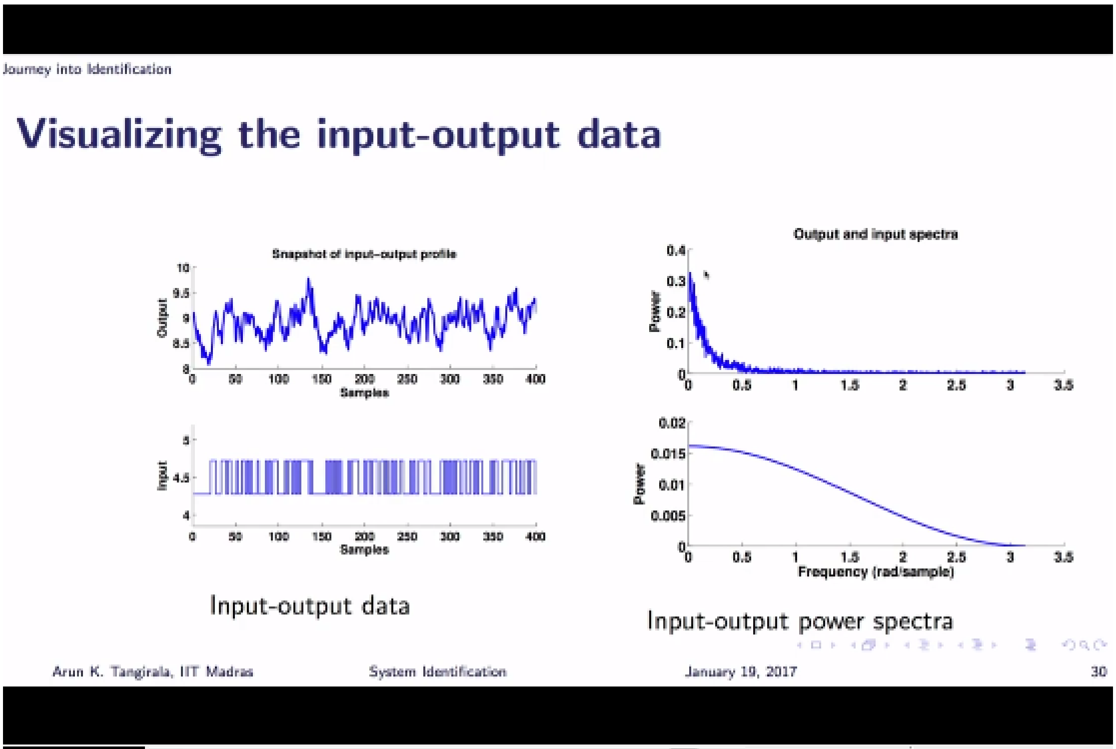
Power is estimated though, most of the problem.

Okay. Statement is kind of incomplete. Answer is correct. We can infer that it's a low pass filter, but the answer has to be complete. So he says that I can infer from the power spectral density plots that the system acts as a low pass filter. That means, we have already described what a low pass filter is. It preferentially allows low frequencies to go through. So now you have to tell me, you have to substantiate this statement by observing those two plots. Why? Whether you agree with the answer that it's a low pass filter? Do you have anything to say?

Yeah, it does filter out many of the frequencies in the input change.

So that you have to say that. So we, it has filtered out. The system has filtered out many of the frequencies that we have including the input which means what we thought was within the mandatory does actually further ignored those. That means that may be given another chance, I would come up with a better input. This input is not optimal one in terms of identification effort, the cost of identification. It tells me that I have unnecessarily excited the system over mid frequencies because systems as I don't respond to this kind of queries, right? There are many offices in this country with they don't respond to the queries. So after [33:28 inaudible] there's no point in asking those questions. Although we have RTI and all of that, but basically, if I were to be given a chance to redesign my input, I would ignore, I would not include these frequencies at all, because it's not just identification effort. Remember, in every experiment, actuator is moving and it's undergoing wear and tear. So I want to actually have a long-life for the actuator. What is the point? I mean, for us, if I'm asking a question, there is a vocal cord, there is some energy, the vocal cords undergo some strain, I don't want to really ask questions and strain my vocal cords to which the person, other person may not respond. Exactly the same thing here. So this clearly tells me that the system is a low pass filter.

(Refer Slide Time: 34: 15)



This is an exploratory data. So we're trying to understand it's a nonparametric analysis is what we call. Where we are not making any major assumptions about the system, of course, underneath we are assuming it's linear and timing variant, but beyond that, I'm not making any other assumptions. Only for LTA systems, we can speak of filters and so on. It's not that other systems are not filters but they have more complicated filtering characteristics. Okay. So we have understood quite a bit about the system, but it is not sufficient as far as production is concerned.