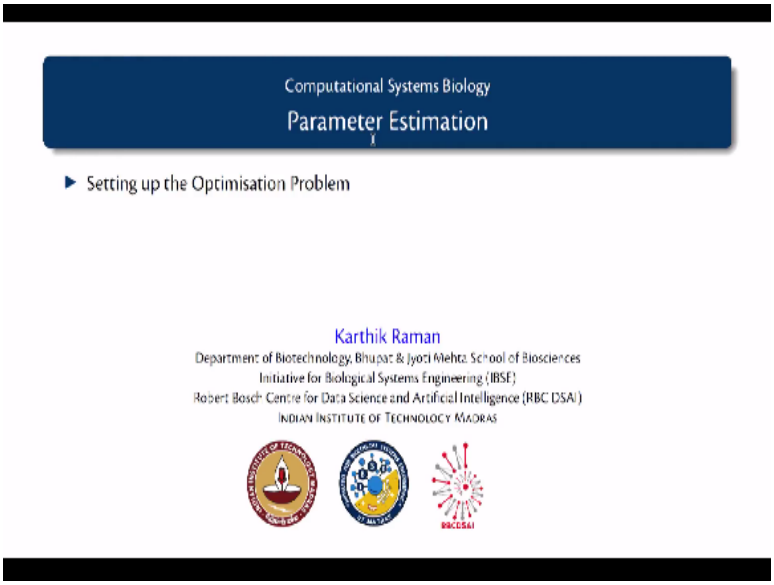
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology - Madras

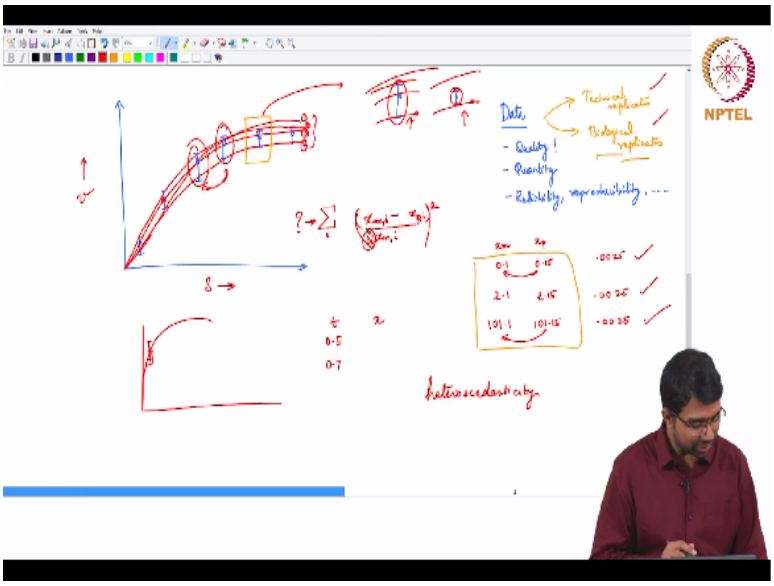
Lecture – 41
Parameter Estimation

(Refer Slide Time: 00:11)



In this video, we will focus a little more on how we go about setting up the optimization problem for parameter estimation. It is on the flavour of least squares but there are many interesting modifications that one can perform.

(Refer Slide Time: 00:24)



So what is it that people normally do? You have some time course. Let us just plot an arbitrary time course. You will not have the time course in fact. You would only have data points. Deliberately drawing the errors that are somewhat assymetric. May be let us just assume that a simple Michaelis-Menten. Let us start even with a simple Michaelis-Menten.

Even before we get here, you may have to answer some questions about the data. First quality, then quantity, reliability, reproducibility, so many more things. Assuming your data are good, you can then start with the fitting process. But if you look at the classic fitting process, can you tell me which of these is the best curve? **"Professor - student conversation starts"** One, the last one. **"Professor - student conversation ends."**

Let us, let us label them. 1, 2, 3, 4. I think most of you would say 3. In fact, that is what most of your algorithms would estimate. Let us give something. Why? Because you have this strange fascination to the mean, right. What is so special about the mean? It is not even a data point you observe. You never observe a mean by the way. You observe x1, x2 and x3. You did x1+x2+x3/3 and you computed the mean.

So the mean was not even observed by you in the experiment as such. So how would you, let us start with a classic least square fit. You will basically say x model measured-x predicted whole square, right. This would be your standard cos function, right. So in fact let us make, in fact you

would say… is that right. Across every data point, you minimise the sum squared error. Is that fine? Is that a good idea?

There is one issue here. Let us say your points are your xm and x predicted are 0.1, 0.13; 2.1, 2.13; 101.1, 101.15. What is the sum square error that you will get? Is that right? As in each case is just 0.05 but does it seem fine to penalize. This is well perfect, it seems. Whereas this is an error of, much higher error, right. This is an error of literally 50% right. Do you agree? Whereas what is the percentage error here? Clearly less than 1%, right but here it is 50% and your all these you have equal contribution towards your cos function. Does that make sense?

"Professor - student conversation starts" The dataset is our wide range of parameters. It could be a problem. **"Professor - student conversation ends."** So if your data scales weirdly, you may have to normalize it, right. So one easy way to that is, so this looks better, right. This is already a little better. The one other issue that still remains is, you are asking your model to be better than your data. Are you not doing that? Because you want every point, the curve to pass through the mean of every point whereas that is not even what you observe in reality.

So anything that falls between your bars is actually a good curve. That falls between your error bars like this, so all, I would say that these 4 curves are model or you know in mean accuracies in drawing are essentially equal. I would not say one is better than the other and let us, let us zoom in. Let us zoom in to a particular point here. No you have multiple, so in a good biological sample, you need to have technical replicates and biological replicates, right.

What is a biological replicate? What is a technical replicate? Technical replicate is just repeating your measurement try sort of. Biological replicate is running it in 2 different reactors, 2 different cells. You are doing a chemical experiment; it is not a big deal. You know, you assume that the compounds are going to behave the same more or less, right but 2 cells are not the same. Two cells will differ.

So if you have 2 different cells, 2 different populations of cells, you want to study them separately, right. So this, let us assume this as average across everything. This is, let us say,

biological duplicates across technical triplicates, right. So I am talking about an ideal scenario never the case and one more issue that you would have is, it is somewhat related this issue. In many biological systems, the initial response is very rapid.

So it is not unusual to find a curve that looks like this. So you are not only worried about this error, you are actually worried about how good your, is your stopwatch. Where did you exactly stop the time and when did you exactly make the sampling. You have an autosampler all that great but potentially a small variation, so let us say you measure t as 0.5 seconds and you got some x but may be your error in measurement was 0.2 seconds.

The delta in y might be much much higher. You have to worry about these things. These are all the things that you practically go wrong when you build these models which is why I am literally not focussed on the theory of dynamic modeling which is basically writing ODEs. We can easily do that. We saw a couple of examples yesterday. No big deal but then in practice, how do you handle these models?

How do you know if you, if that is good at all? So you are getting to what is the measure of goodness, right. So this is a measure of goodness. So if we zoomed into this point, you have something like this and you have a few curves that go, one curve that goes here, one curve that goes here, one curve that goes here, one curve that goes here. You will have to evaluate all these curves carefully, right.

So if you look at another point, let us say that point looks like this and your curves are like this. These 2 seem equivalent, right. They are reasonably similar; do you agree? The distance of the curve from the mean is not too bad but the problem is this was a very tightly distributed point. This was already a poorly measured point and you may want to give some more linear to the model here than here. You may want to penalize this curve a lot more than you penalize this curve.

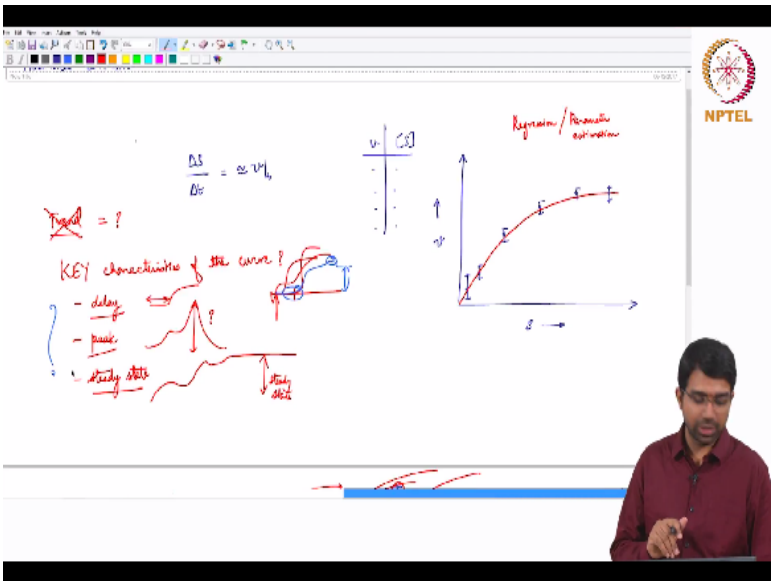
“Professor - student conversation starts” Weight least. **“Professor - student conversation ends.”** So you need a weighted least squares. Let us say we put a delta, right. Does that add up.

So what did we say? You want to do a classic least squares? You want to just take the difference between your point, your measured value and your predicted value but may be that is not good enough because as we looked at this example, the same error could, the same, you know, numerical difference could mean a very different percentage error, right. So in which case you may want to normalize it.

That was the first improvement that we did. But then that may be sufficient because you may have rapid variation in some point and you may have very poor measurements in one point and very good measurements in another point. You want to therefore weight different points differently. So this is what is called heteroscedasticity. You do not have the same standard deviation at every point, right. Each measurement has its own characteristics. So this points errors are different from this points errors.

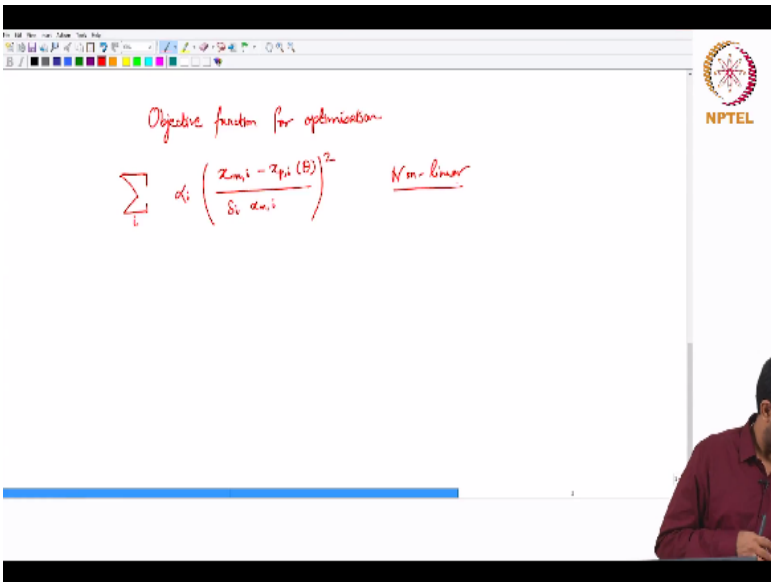
"Professor - student conversation starts" But how do we know standard deviation experimentally. You make these replicates, no. Biological. So you have a few, I am hoping you have a few measurements. Otherwise, you have no clue. Otherwise you have to then just have an approximate feeling of what is your reasonable error that you are willing to admit.

I am okay with an error of 10%. I am okay with an error of 15%, right. So that has to factor up in this delta. (()) (11:46) initial time I want less error. I want to gain… Exactly, yes. You do not really can avoid, if you are asked in the steady state prediction by 5%, it is not a big deal, right.
"Professor - student conversation ends."
(Refer Slide Time: 11:59)



So this is where it comes back to what we were talking about here, right. Is your delay more important than your peak or is your steady state more important than your delay or peak or whatever? So but potentially you can capture all of these in some sort of a weighted fashion. Some weighted least squares formulation should be able to capture all of these. So if you want let us add another alpha a here as well. So what is that that you are doing? May be I should, I will rewrite it more clearly.

(Refer Slide Time: 12:29)



So we said your cos function is x measured of i, in fact the theta does not apply here, -x predicted of i for a given parameter set theta/x measured of i, something to account for the heteroscedasticity of each point and may be something else to account for the relative importance

of each point. How desperate are you to achieve a fit at point i compared to point j.

"Professor - student conversation starts" (()) (13:18) could be some measure of standard deviation. How much error are you willing to accommodate in one point. We can combine phi and delta i. Potentially, but see alpha i, you can just leave it out, right. I am just proposing it as an additional improvement across i experimental points. This is essentially your cos function. This is clearly a non-linear cos function.

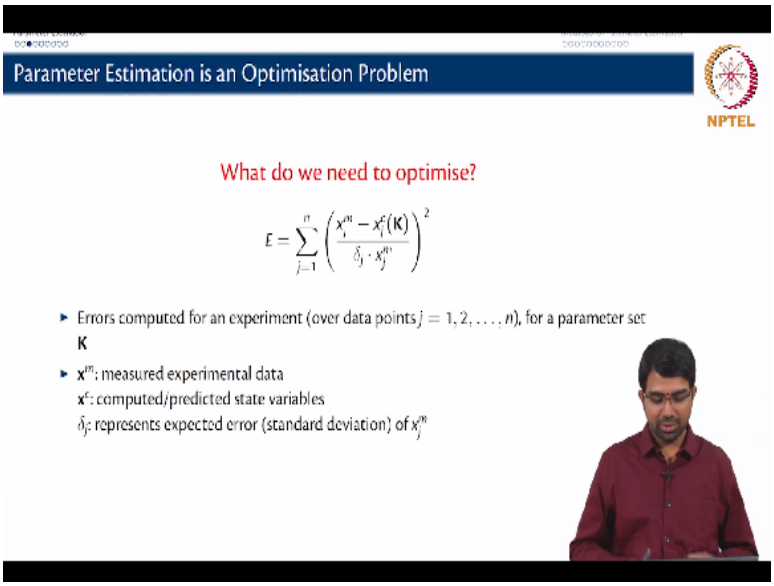
Any time you see the word non-linear, you have to be worried, right because everything gets a little tougher for non-linear systems, not necessarily. Because the problem is, how do we compute each of these points. Your problem actually comes here, right. xp of theta is even more complex. You are going to vary, see it is, you can say it is convex in x or something like that but you need to look at what is it in theta. It is very very non-linear in theta.

So that may be to non-convexity. So it is non-convex. I think there are (()) (14:43) may be convex but I think very rare. Sir what is alpha? Alpha is just some, you, okay. You just say that this point has an importance of 0.9, this point has an importance of 0.6, something. So something to basically say that I want a much better fit here than I am, I am going to accommodate poorer fits more here than here.

You have to give importance to some. Relative importance of the points but you know this is not commonly used. I am sort of making it up but basically it gives you a handle to have much better cos functions. See the problem is I do not think anybody use cos functions. In cos functions, you need to be tuned because modeling is a very subjective exercise. It depends upon what is it that you want to achieve at the end of the modeling exercise.

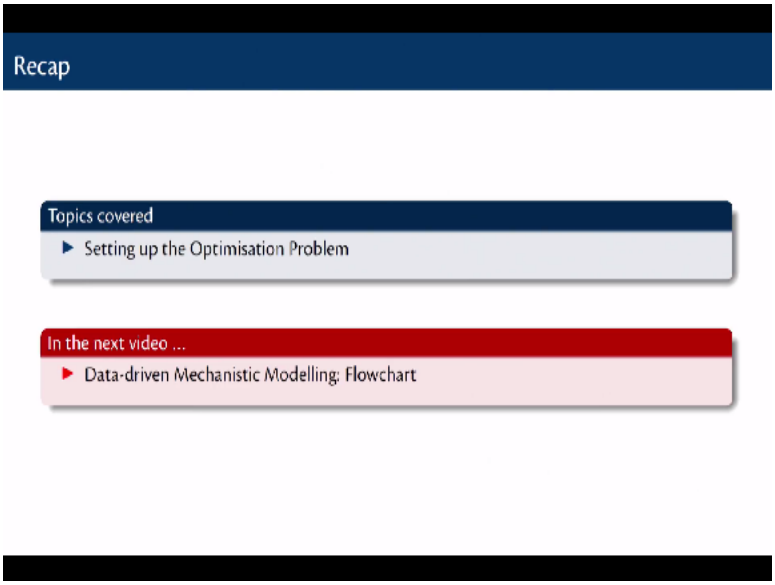
And you just cannot open Fn search in Matlab, take the regular cos function and run with it. You are not going to get a good model by doing that and this is something I really want to emphasise through this course, we want to be very practical, right. We want to do; we want to solve the real problems that one encounters when doing systems biology. **"Professor - student conversation ends."**

(Refer Slide Time: 15:52)



So what do we need to optimise? This is what we need to optimize, right. This is very similar to what I showed you using C instead of p, computed instead of predicted but it is the same thing. x measured-x of theta and using K here. Everything else is the same. This is what we need to optimize because essentially a complex non-linear optimization problem.

(Refer Slide Time: 16:17)



In this video, we had a good overview of how we set up the optimization problem for parameter estimation. The next video, we will look at a classic, a flow chart for data driven mechanistic modeling which is what we have been discussing for the last few classes. So you have some data on a biological system and you want to build a mechanistic model, how would you go about

doing it? What are all the various steps involved?