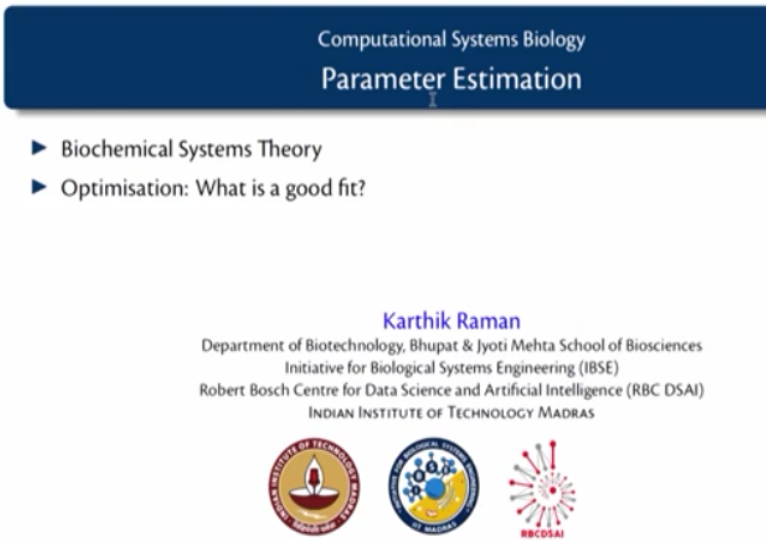
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 40
Parameter Estimation

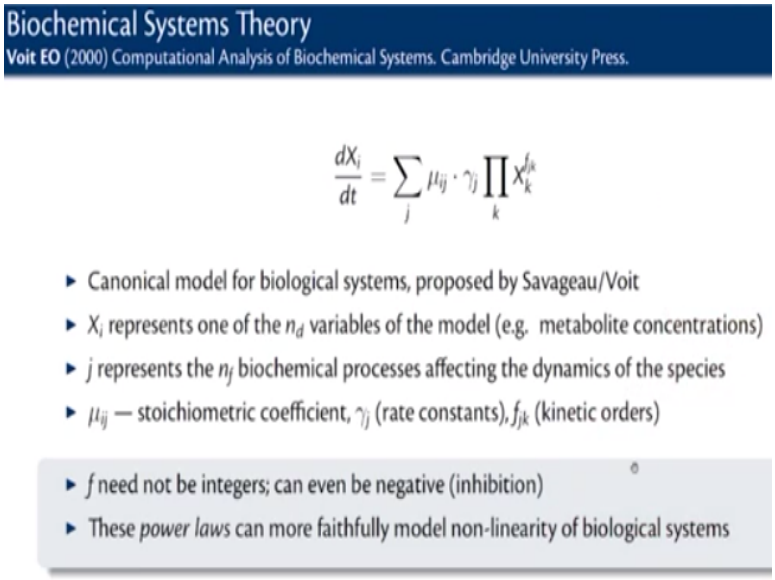
(Refer Slide Time: 00:12)



From this lecture onwards, we will look for several lectures we will focus on parameter estimation. We will first start with biochemical systems theory and start seeing how we pose parameter estimation as a nice optimisation problem. Welcome back, today we will look a little more on dynamic modelling, and particularly focus on a very important aspect of dynamic modelling, which is parameter estimation.

Typically, you will have a good data set, hopefully a good data set of quality, quantity, reproducibility and so on, which you need to employ to identify parameters for your dynamical system of interest. So how do you do that is something we will discuss and understand in the coming lectures.

(Refer Slide Time: 00:58)



Before we go there, there is also this classic field of study called biochemical systems theory, which essentially tries to cast all the different kinds of modules we studied yesterday. We briefly talked about Michaelis-Menten models, mass action models of course, and your Hill kinetics and so on. It tries to cast all of these modules with inhibition, without inhibition, with competitive inhibition and so on.

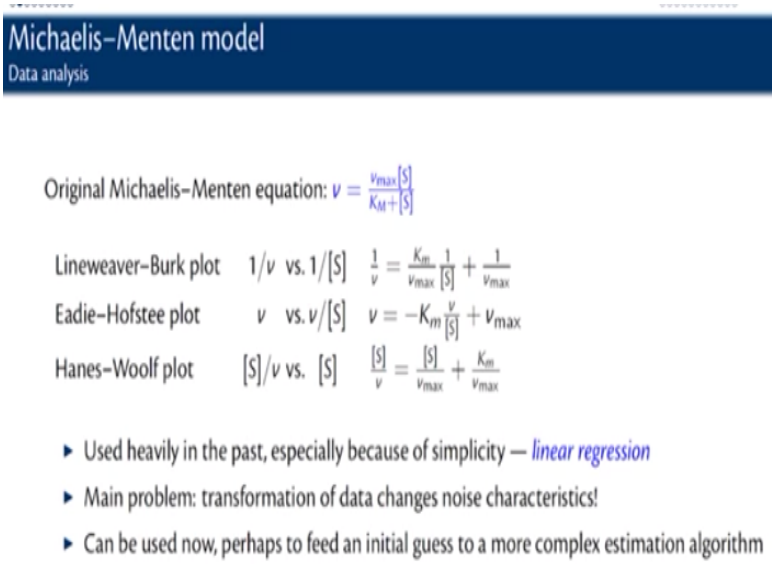
It tries to cast all of these into a sort of differential equations which involve power laws. You can see the power law here, so dx/dt is some product of xk raised to some power. What are your x size, they are your considerations of different species, typically? Your mu ij like your stoichiometric coefficients, you have rate constants and the kinetic orders. The orders can be negative and so on for inhibition and things like that.

In reality you will not find something as like purely a second order reaction, it will be like a 1.8 order reaction and things like that. So those are the kind of exponents that you will end up having in these systems and these power laws can also faithfully model the non-linearity of biological systems. They also give you usually saturating response. They do not give you like unbounded responses and things like that. They end up giving saturation responses and so on.

So given a system of this sort a typical challenge is you hopefully have generated some data or some biologist have come to you with some data on a particular system. So how do you then model this system, or in fact how do you model this system. The first step in modelling is you understand the system and write out the equations. The moment you write out the equations you still do not know the value of any of the parameters.

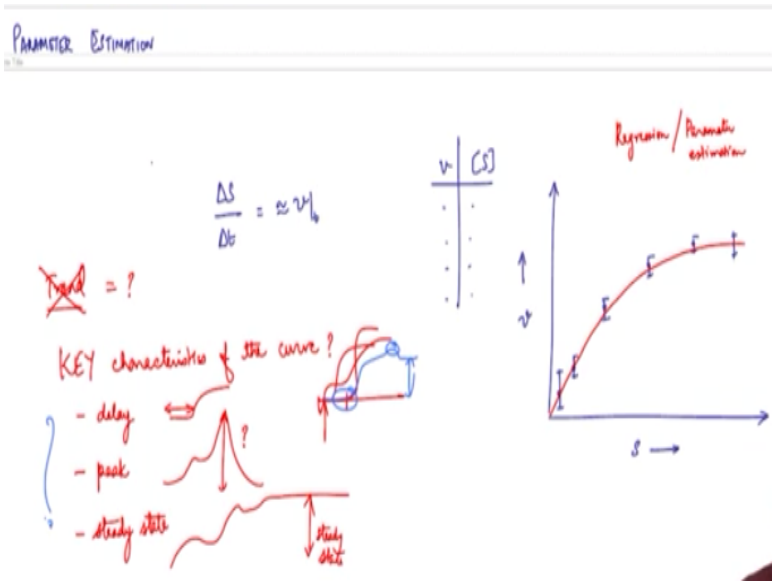
Until you know the value of the parameters you cannot make any worthwhile predictions, k is for every species, all species x1, x2, x3 over all the possible species. Many of them may have 0 exponents and things like that, it does not depend on them.

(Refer Slide Time: 03:16)



So let us look at the data analysis part, so I think most you must have carried out Michaelis-Menten kind of experiment in the lab. So what would you normally get from that experiment.

(Refer Slide Time: 03:26)



So what is the data that you normally get from a Michaelis-Menten experiment and how do you do that delta s/delta t, how do you know the exact t, t keeps ticking know, how do you stop t, yes remember, you basically literally stop the reaction by putting in an ice, and then you make the measurement and then you compute some delta s/delta t or something, this will approximately be equal to your v, v at a particular time.

You will essentially end up computing something like this is right, so you will have t and v, and of course s, v is the rate of the reaction, in fact you may not use t as well, in the experiments, you basically just compute the initial rate, so you compute v and s, and you fit it. How do you fit it? There are 3 different ways to do it, that is the Lineweaver-Burk plot, that is the Eadie-Hofstee plot and the Kanes-Woolf plot.

What are the advantages and disadvantages of each of these? I am not going into it; you would have studied it in the classic biochemistry course. The one thing to note from the point of view you of this course is these transformations are problematic, because what happens when you transform something. You are is transforming something right, you are taking v, taking reciprocal, you are taking s, you are taking reciprocal, v cannot be 0.

So you will have some behaviours close to that 0. Your errors also get transformed, usually it is fair to assume that the errors are normally distributed. You can assume different kinds of noise but basically central limit theorem. But what happens in these cases, the errors now get transformed, so all those assumptions no longer hold, which means that you have a, the math that you are using is going to be problematic.

So 1 thing that I would normally suggest is, it is time to forget these plots. I think classic biochemist will cringe at that. But basically how do you estimate the parameters now. Why did we use these linearized plots? It is just so that you can plot these points on a graph paper and then see what is the correct slope that will, if you want to do it systematically you will still have to use the computer.

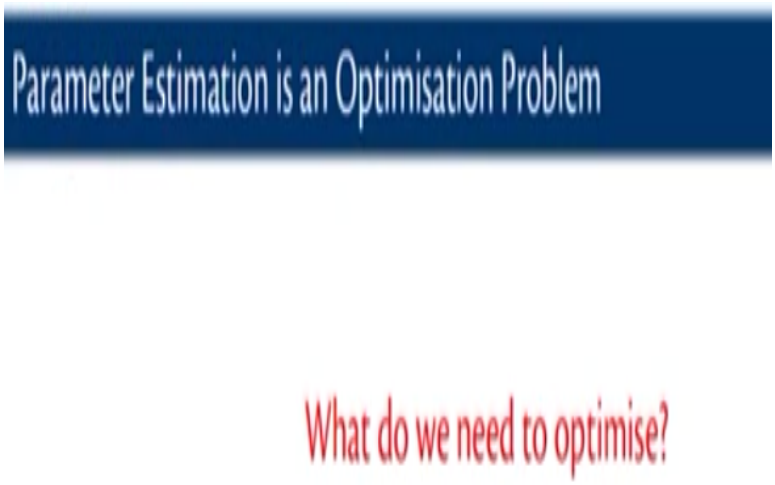
If you are any way going to use the computer, might as well do non-linear fitting, why do you want to do linear fitting. Linear fitting is good for graph papers. So these plots are still useful in terms of maybe getting a decent guess for v and s. But a 2-dimensional problem is not at all hard, so you do not even need a decent initial guess for v and s. So these have been used heavy in the past, because of simplicity with the main problem is that transformation of data changes noise characteristics.

But we can still use it now perhaps to feed an initial guess to more complex estimation algorithm. So this basically goes back to curve fitting. So let us say you have a few data

points here. You now want to fit a good curve through this. You may want to fit a curve that looks like this. These essentially are parameter estimation, how do you do this. This is a very interesting problem.

Of course the most important problem in modelling is trying to build the model in the first place. But that process itself is incomplete before estimating the parameters. I could come up with an arbitrary model if you can never find a parameter set that can connect to this model, to my real system. By the real system I mean the data that the real system through up. If you can never connect the 2, then you do not have the model in the first place. The model is not tenable for this purpose.

(Refer Slide Time: 08:26)



So parameter estimation, you can imagine is an optimisation problem. What is that you want to optimise. You want to basically improve the fit to the data. There are various ways to pose of data (()) (08:39) would say that can you maximise the probability of observing the data given a parameter set. So you want to find the parameter set that maximises the probability of observing the given data set. That translates to maximum likelihood and so on.

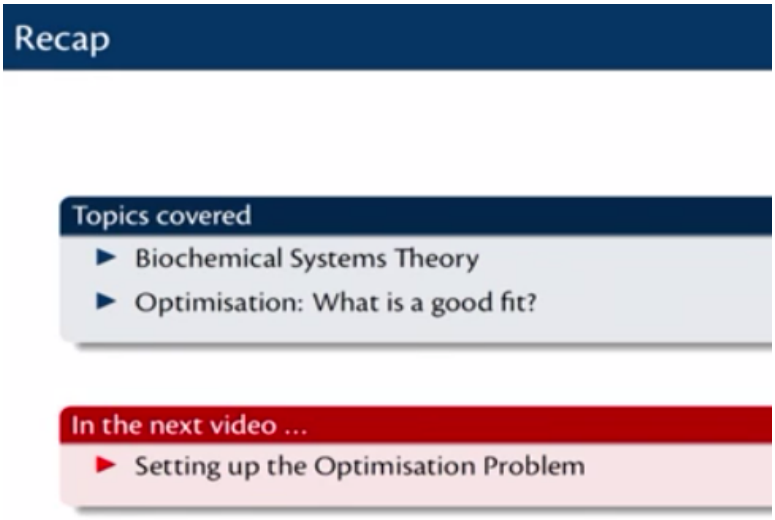
But what is that you really want to optimise here. So let us look at this example. What is a good curve? This can become a very interesting question. This is still an open question. It might sound sort of simple and trivial, but what really is a good curve. So let us say, I agree with trend, what does it mean by trend, or rather than trend can we now look at key characteristics of the curve, what could be the key characteristics of the curve.

You may have a delay, meaning like this, so you may be interested in capturing this delay in your model. You may have a peak, are you able to capture this. Very commonly, you may have a steady state, can you capture the steady state. You may have a delay in the response, there is some, let us say you administered a drug at time t, time 0, you could have different responses for this.

It could start like this, it could go like this, it could go like this. If this could be from you model, and your real data let us say looks somewhat like, let us just assume, which of these 3 is good, because if you were to compute some least squares or something like that you may not be able to truly say which of these is good. You may want to subjectively attach much more importance to 1 of the features.

It could be the steady state, it could be the peak, it could be the delay. What is more important? This is very subjective, this means that you need a very intelligent way to estimate parameters, and this is something people normally do not do. I will tell you what people normally do.

(Refer Slide Time: 12:00)



So in this video we had a brief overview of biochemical systems theory and we also studied what are all the various ways 1 can specify an optimisation problem, typically a good least square fit as a classic strategy. In the next video, we will continue to look at what are all the challenges in setting up a good optimisation problem which helps us get good parameter estimates.