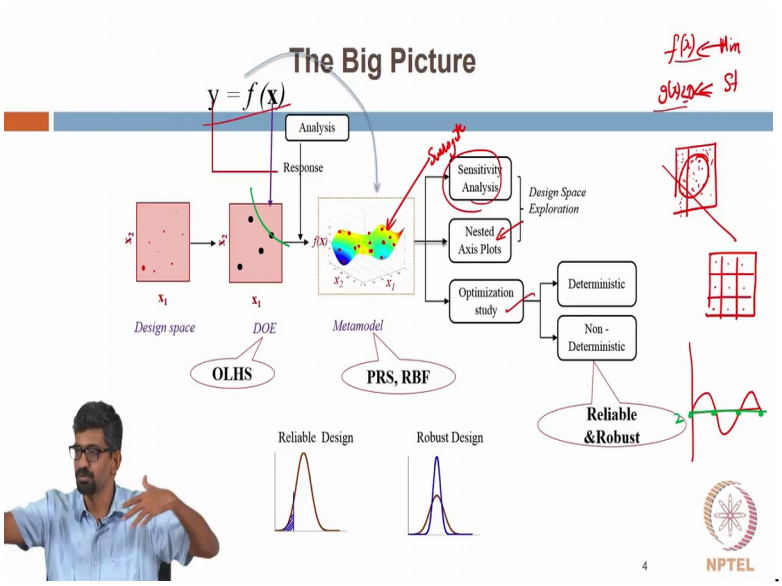
**Surrogates and Approximations in Engineering Design**
**Prof. Palaniappan Ramu**
**Department of Engineering Design**
**Indian Institute of Technology, Madras**

**Lecture - 05**
**The Big Picture - Overview**

(Refer Slide Time: 00:18)




This is just to give you a big picture and then we go back. Just like any other engineering; we are also interested in y equal to f of x. If you see, this equation is so powerful; that 90 percent of engineering is based on this equation that is all. Not only in optimization not only in what we are doing; what linear algebra, what you are doing you are only doing this; that is all. MATLAB 90 percent is based on this. You know how to solve a equal to b x you are half an engineer; then you need to only train yourself in the domain that you know but, you should know everything in solving a equal to b x. That is all is this, what is a coefficients x is x 1 x 2 x 3 that is all you need you know that you can solve all half of the problems.

Now, the question in terms of optimization is; first there is a design space; this is all the engineer knows; sometimes the engineer does not know even this. You know what the factors are but, you are not sure whether only those are the factors are there are also other factors. So, then we say that trivial many two vital few. So, you cannot say like efficiency of engine; it is dependent on infinite number of factors, but infinite number of

factors you cannot use for designing you will have to bring it down to few. Then, so I have design space x 1 and x 2 because I can show only two variables here, but it can be n dimensions also.

Now, the second question is; this is what you want; that is what we are talking about this whole time correct. I want a f I want a g I want a h. So, that I can write this; f of x is what I want to minimize, such that it is g of x is less than 0 less than or equal to 0. But for that I need my f and g that is what we are trying to see here. You are giving me a design space; you are asking me a function f of x. I do not have this function I have to construct this function. How will I construct you are also saying for any combination of x 1 and x 2 for me to get the response it is going to take 4 days for one combination. So, I cannot give you infinite combinations.

So, you need to give few combinations here so that I can construct this function; we will worry about constructing this function later. So, there is also a budget for doing this. They say we will give you about 30 days 1 month you take 4 weeks you take; you do whatever it is you give me the function. Then what do you do? You say each simulation in a cluster runs for 6 hours

So, if I am smart I will batch this into a single run. So, 6 hours means per day it is 4 runs you can do ok. So, for 30 days I can do 100 and 20 runs. So, I will start populating 100 and 20 points in my design space. Meaning: 100 and 20 points like this; each one is a different x 1 x 2 combination they are likely to give different y values.

Now, how can I place these points inside this design space? What should be my criteria? That is also a criteria, correct I can say I do not like right side so, we will keep all of them in the left side. Then this size is no I do not like us rectangle you know let us do a circle that too at the center and we will put all the points inside that.

Then some person, who is an engineer, comes and says; hey you know that design space are decomposed into good and bad. And always the left side is bad because, it is lower. So, we will always only sample in the right side right top. If possible all are alternatives right some alternative you can have. How will you choose; where will you put these points? Or without any of these assumptions also someone comes and says: I know that you can grid it like this and then one point in each grid. Imagine that you can grid it into 120, then you will put one point into each grid. This is called uniform sampling they are

all equiv spaced, but there is a problem with uniform sampling; what is that? Imagine that I have a curve like this sinusoidal curve ok. I give you four points. So, you do equidistance, what do you do? You put it here, you put it here, you put it here, you put it here; what is your function approximation now.

Student: (Refer Time: 05:30).

This will be your function approximation; correct right, you had 4 points everywhere it gave you I do not know let this be 2 it gave you 2. So, you said even by me changing my x it is always giving me 2. So, the function value is 2. So, there is a problem with, but then you should have some information on your response, no one will do this because you know the function is periodic ok. You do uniform sampling when the function is a periodic. But there are certain reasons; why people who do not want to do a periodic sampling which is uniform sampling, ok.

So, then they said ok, but you are doing this, but what should be your goal you believe. When I am putting these points; whether it is 4 or 40 or 400, what is it that you are looking for?

Student: (Refer Time: 06:25).

So, I should trend, pattern; I should be able to approximate that, approximate I am not using that word because, that approximation is one level more, because, you are going to fit and there could be a fitting error it is not the problem with the data, you are fitting it in a way that is not related to the data. So, the data itself I would like to extract as much as I can ok.

So, imagine that, someone is visiting your lab from what country do you like other than India.

Student: US.

America, so some collaborator from the US comes, he is visiting your lab ok. So, the person says I am going to stay here for 1 year and I am interested in knowing the, culture of Chennai let us say. Let us say that he knows the culture meaning like he knows where all to visit Chennai to expose this guy to the culture of India or Chennai whatever it is.

He says that I am going to be here 1 year and let us say every Saturday Sunday I am available ok. So, that is each point.

So, then he has a relaxed time; he says 365 days 52 weeks2 days in each week; so there is lot of days. So, I will take one by one you can do ok, but the same visitor comes and says I am staying here only 1 month. And then the last week I am going to travel, so I do not have time only 3 weeks 2 days in each week 6 days

Now, I want to experience Chennai. So, you know that you cannot go to every nook and corner and visit, but how will you schedule your visits? Do not worry about the timing and all this: I will go in metro do not worry about that; what I am saying is what will be your goal. I mean that also comes into picture, how far you travel and all that also becomes. But what will be your goal, if you want to expose. So, your idea is I will visit as many places, but that is again limited right; I will have to take, I will have to go that far I can go only to these many places, but I will choose the places in such a way that it will give the diversity of Chennai or India whatever you are worried about ok.

That is the goal here. There is only limited simulations; you cannot ask for a million simulations, I will tell you that you can run only 25 simulations. But, you will have to choose the places where you are going to simulate ok. If I take you to like a park hotel bar, then I will also take you to a local bar for you to understand; what is the diversity of Chennai ok. So, I take you to a high class restaurant, I will also take you to a low class restaurant ok. Then, if I take you to I do not know like a cinema theater then I also need to take you to a drama theater for you to understand what is the social expansion ok. So, I take you to a private beach then I will also take you to a Besant Nagar beach for you to understand; what is the diversity.

So, that is how; and then you know this is two extremes of the same stuff, but also culture wise, I will choose what will expose you which will give you a different stuff.

Now imagine that, in the morning, I took you to a dancer one Bharatanatyam and then in the evening, I took you to dancer two Bharatanatyam; does that enlarge your knowledge any much anything about Bharatanatyam? Maybe a little bit, but about Bharatanatyam in general it will not increase much right. But instead, if I took you to some folk dance in the morning and then took you to a Bharatanatyam in the evening: yes, you have a little bit more understanding of what different types of dances are ok.

So, I want to get as much information from this limited number of points; that should be your goal. But, obviously, these are all verbal you need to cast it as an optimization problem; that is what all your Ola and Ubera are doing ok. Because, they want to have as many pickups as they can, and that is why you do not get cars in the Adyar gate. If you are smart enough what you should do is you should book it from the Velachery gate, because that is where they have their storehouse in phoenix mall; you understand right. If they have to come to Adyar there is no parking spot for them in Adyar. So, they cannot park anywhere ok.

So, you see in the thing you will never see, any some passerby is there you will come, but you put in Velachery in the immediate neighborhood there are 50 cars there ok. So, that is how they are doing: minimize distance and maximize pickups. So, you should know that and run your algorithm accordingly. You maximize your chances by picking one from where correct right, if you want to pick someone, there are like 50 cars there in Adyar you should be lucky to have a car ok. So, always schedule it from the Velachery side. So, that is an algorithm right basically.

So, the deal is I need to locate these points in such a way that I will maximize my information, but boss you do not know the function how will you maximize your information. So, it should not there is it is called exploration and exploitation ok, we are only talking about exploration here. So, what it will do is; it is trying to use some relationship between x 1 and x 2 in terms of correlation and place the points in such a way that it will grab the maximum information ok. In case, there are some what you call like constrained design of experiments; someone came and told you know what this is the zone of interest: then yes, I will just put few points here and then the remaining all will go here. Just to get an idea of what this area is and then I will go and get it here. But that is usually not the case, it is only done in an adaptive sense, but people are doing such algorithms, this is generic.

So, once you have a design of experiment; that each of please understand generating design of experiment is not a problem; you give me 20 dimensions also I can generate it in mat lab or any python or anything ok. Generating, these points is not a problem ok. There is a function inside this room, I have told you right you do not know the function; there is a function inside this room ok.

So, now you are trying to approximate this function; I am telling you that you can use 20 points. So, you are just throwing these 20 points imagine that this mike are all 1 1 point that node is 1 1 point ok. Throwing these points in space is not a problem, there is no expense in this you just need to do some algorithm such that there is some min max algorithm that it needs to satisfy that is all ok. Throwing these points is not a problem, but, finding the function value at those points is what the challenges because, that is what is going to take you 4 days ok. So, you can throw as many points as you want, but evaluating those points is where the challenge comes in ok.

So, throwing these points is not a problem; sorry throwing these points is not a problem, but of course you should be wise enough to know that you will not be able to evaluate all the points. So, you put limited number of points and you run the analysis and it gives you the response at each of these point which is your f of x. And then, in terms of x 1;I do not know whether you are able to see, there are some dots here ok. So, there are some dots here, for the sake of completion I added a few dots ok. So, there are some dots here and I am fitting a surface of f in terms of x 1 and x 2; do not worry this is the surrogate model that you are going to talk about how to fit it is what the rest of the course is. So, do not worry

Now, once you have this, one application we have seen is optimization, but you can do lot of other things: one is you can do the sensitivity analysis, you can have nested access plot to look at what variable, what affects visualization, you can do an optimization; under optimization it could be deterministic or non deterministic. What I mean is non deterministic means, for instance your force is not a deterministic variable it is a random variable then your failure is also random. So, there are things called the robust design and reliable design it is out of scope for this discussion, so do not worry about that part.

See sometimes what happens is you are really not interested in an optimization. Let us say thickness was one of your variables sheet metal thickness, you run the optimization and it gave you 12.8 mm. It does not make any sense because, you go to Velachery to the hardware shop you get aluminum or steel sheets of either 12 mm or 14 mm ok. So, you say 12.8 it says then we should go to for 14, here are existing design and self is 14 only; what did he optimize ok. So, unless you are Tata you cannot generate 12.8. So, does it mean that the study was not useless its ender rendered useless no, it does not rendered useless.

So, I need to know what my sensitivity is; often times people just need to you go to we work with few car companies where they say that; you just tell me the direction in which it increases that is all; you just need to tell me, just give me the direction where I need to increase or reduce thickness, that is all I need. To know nothing else do not give me some value and all I am not expecting and it will not happen also, because today they are sourcing almost 60 percent of the components from outside; other than engine everything is else is being sourced. There are some manufacturing tolerances this that and all that, they say that you give me some x 1 to x 5 final design variable combination and all we cannot achieve, there is no way that it is possible.

So, you just tell me you should increase x 1 you should decrease x 2, you keep x 3 and x 4 the same, x 5 you increase; you are likely to reduce your weight that is all I want. I will decide what I should do. The direction is what I want to know. This becomes very important for them in aerodynamic studies not in structural. In aerodynamics, because they cannot go and change right this clay modeling and all that they try to do, to get the aerodynamic and then they put it in the wind tunnel for testing, they cannot do as many tests it becomes very very expensive for them. So, is the case with crash tests and all that, ok.

So, in such cases what they want to know is only the sensitivity. If you are going to change variable number 32 x 32 by 0.2 units: how much will your response change. The catch in this is there is also interaction, when you are changing x 32 did you change x 28 will there be a interaction; you need to worry about that one also ok. So, the sensitivity study becomes axis plots visualizations today with data in big data one of the major things is visualization; how would you visualize the data. For instance one of our students is looking at some methods to visualize not big data, but even the regular data, how would you visualize data. more than two dimensions I cannot show you even a contour plot here. I can only show you projections which it does not show you the relationship, that that is all limited it is ok

So, visualization techniques become very important, because if you can see a peak nothing like that, if you can see a valley nothing like that. It will not show you a peak our valley it will tell you that this is the direction, its increasing, its decreasing, some information do you give; and of course, optimization. So, this whole thing is called design space exploration. Often times people do not do optimization for finding the final

combination; they want to know the history which all points did it going the; so this point is near optima good enough this is something, that we already have; then they will look at their library and then they will figure out in 1982 they built a model, that is similar to this design. Then they want to know knowledge based a understanding or they go to the next best one, there is some relationship between these two why; they want that kind of an understanding often rather than giving one that is only for our class just giving one final value for x 1 x 2 and then finding what it is ok.

So, what I am trying to tell you is; to as the title suggests it is a big picture ok. So, you do not really hunting for one thing, you want to have an overall understanding of your design space. Hence, we also need to be able to approximate your function in an overall sense, not only in a local sense though you are interested in optimization is in a local sense, you want to understand how the function varies in an overall sense. This is from a design perspective.