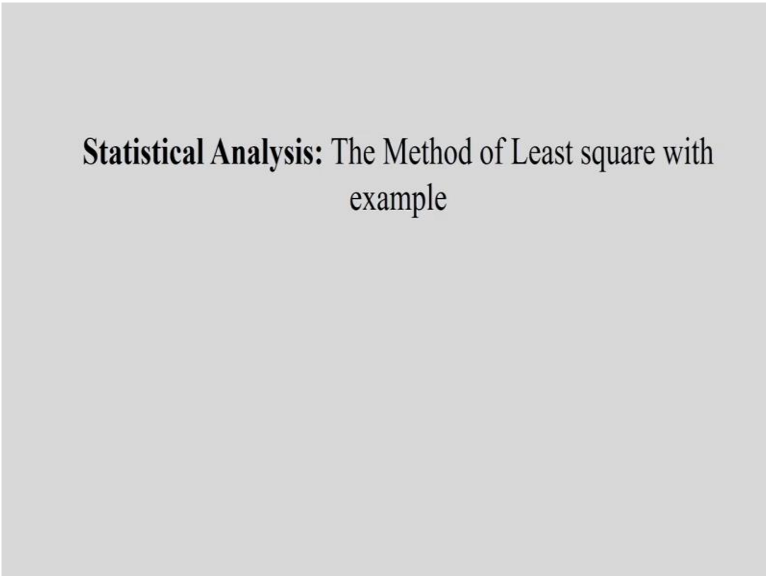
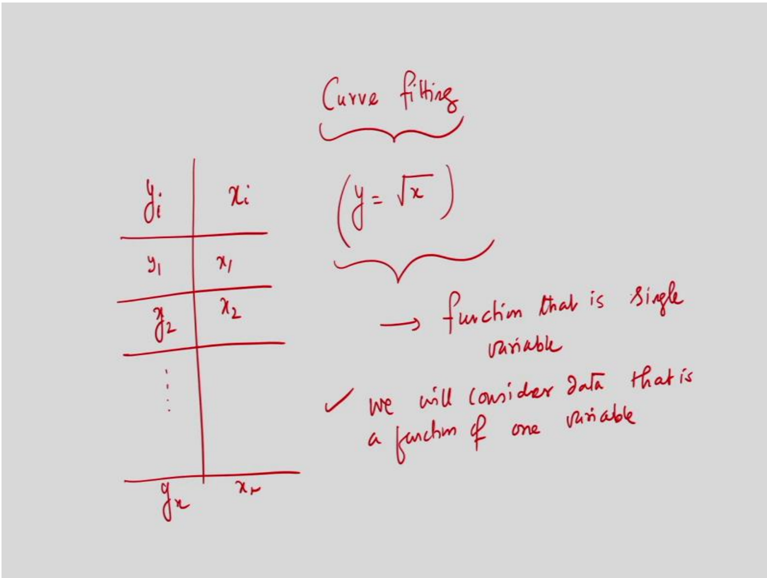
Experimental Methods in Fluid Mechanics
Professor Dr. Pranab Kumar Mondal
Department of Mechanical Engineering
Indian Institute of Technology, Guwahati
Lecture 43
The Method of Least square with example

(Refer Slide Time: 00:32)



Good morning, we will continue our discussion on experimental methods in fluid mechanics and today we will discuss about the, you know curve fitting. In fact, we will discuss the curve fitting analysis using a method which is known as least square. Before I go to discuss about the, you know curve fitting analysis, it is very important to know why you need to go for the curve fitting.

(Refer Slide Time: 00:50)



So, first question that I would like to discuss that why we need to go for curve fitting. So, you know that we have discussed throughout this course that in experiments, what we do, we need to you know, record we need to capture data. So, we have discussed that we have, we can measure temperature using you know that is called resistance temperature detector.

So, we can predict temperature knowing the resistance. Now, similarly, if we now consider the, that is what we have discussed in the last class that by knowing the pressure, we can calculate velocity. So, if we can measure pressure using any measuring equipment that we have discussed in the, in this course. So, we can calculate velocity.

So, if we consider 10 different pressure measurements and based on that 10 different pressure measurements we can calculate what will be the flow velocity. Now, question is, say I am writing that xi and yi. So, knowing the value of xi we can predict the value yi that is what we have discussed in the last class that is y equal to root of x.

Now, so if I write that y1, x1, y2, x2 like this, and finally, we will get yn and xn. So, if we know x1, x2, upto xn then we can predict y1, y2, yn. Now, there might be a case, where we cannot, so, if we conduct experiments, we have data set and based on that data set, we can you know, predict the curve, nature of the curve, why you need to go for that, because see x1 and x2, so, we can have value x1 or x2 and based on the values of x1 and x2 we can predict what will be y1 and y2.

Now, there might be a case that we need to know the velocity corresponding, corresponding to a pressure which is in between x1 and x2. So, if we are, try to know the velocity, the magnitude of velocity for that pressure which is in between x1, x2 then we have to again repeat the experiments.

But, doing experiments not always very easy, rather there are you know, experiment itself a challenging task and if we need to go for some sophisticated experiments then it is very challenging. So, instead of going for further experiments to calculate velocity for an pressure, for a pressure which is you know in between the measured quantity x1 and x2, then what we can do, we can instead of going for the experiments, we can directly get it from the curve.

If we have a functional agency between y and x. So, that means, our objective is to, to develop a functional agency between data points that we have captured from the experimental process, experimental method. Now, if we have a curve, then we do not you know require for further experiments to calculate velocity for any pressure and that we can obtain from the curve.

So, the objective of the curve fitting is to provide you know a generic platform, I can say that we can, you know estimate any quantity from the, you know which is function of maybe a single variable or multivariable, but today we will restrict our analysis for a case, where we will consider only single variable function.

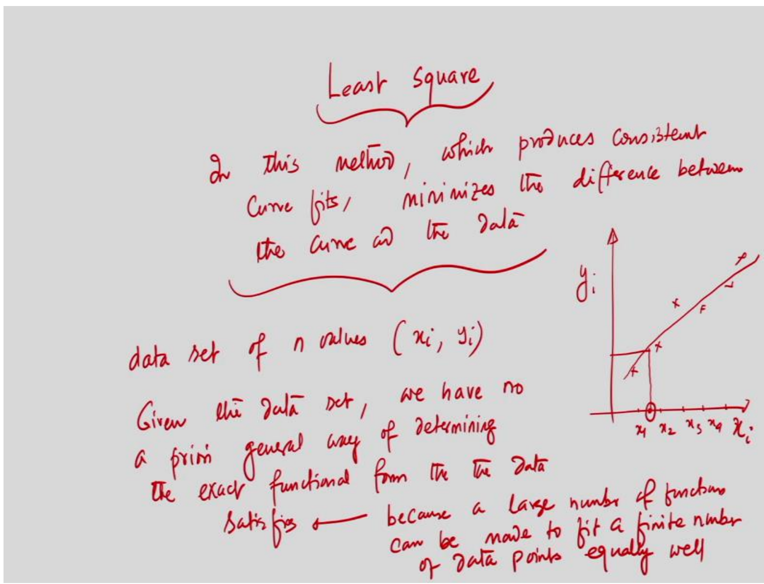
So, basically what I would like to say that we may have data set, but if you would like to calculate any output vary you know output parameter for a given input variable, and that input variable is not listed within the, you know measured quantities that we have, you know considered. But if we have a you know, curve then directly can get it from the curve any value between these measured quantities.

So, the objective of the curve fitting is to provide you know, you know, solution where we can obtain output parameter for any given input provided that inputs should be within this window. Now, today we will discuss that this you know module curve fitting, but as I said our analysis will be you know focusing which is a function.

So, our analysis will be on the curve fitting considering a function that is single variable. So, I can say, we will consider data that is a function of one variable. So, in this case, whether in this class, we will be, you know, considering data that is a function of one variable. So, now, today we will be, there are many methods which are used to, you know, obtain the fit, you know curve, to obtain a fit curve, I mean, the I can say, which are used to have consistent curve fits.

Now, today we will be considering our, we will be considering one of the methods which is also you know, very popular in this paradigm that is in the paradigm of experimental analysis, that is the method of least square.

(Refer Slide Time: 08:01)



In this method, which produce, which produces consistent curve fits. In this method which produces consistent curve fits. Minimizes the difference between the curve and the data. We will discuss the significance rather I can say the name least square from where it is coming. So, at least, we can understand from this that in this method which produces consistent curve fits minimize the difference between curve and the data.

So, our objective is to produce a curve to fit a curve based on the available data, but the curve should be you know, I can say the fittest curve, should be such that. The difference between curve and data point should be minimum. So, that is called least, so the minimum will be the difference between the curve and the data that is very important to know, note.

Say, if we consider that is what the example I have taken in the last slide that yi and xi, we can consider say, 5, 7, or 5 or 6 different values of x. So, this is xi, this is xi, this is xi, this is x like this. So, we can have discrete points, that is what we have considering the last class that method of rejection that is, you know, we are having discrete different input, so xi. Now, if we, as I said that is x1,x2, x3, x4 like this. Now, we know the value of yi for x1, x2, x3, x4, but there might be a case that we need to calculate the value of yi for a value of xi, which is in between x1 and x2, then we need to go for further experiments.

There are two different ways. Either you need to go for a further experiment, otherwise what we can do, we can directly based on the available data we can produce a curve and we can, just we can see if our deserve point is over here, we can just calculate from the curve what will be yi. So, that is the objective of having curve from the experimental data.

Now, what I can say, what I would like to say here that the curve which we have plotted should have minimum difference between the data and the curve. So, we can see that maybe a few points are lying on the curve while remaining other points are outside the curve, but the difference between the curve and the data point should be minimum and that is the objective.

So, this is the least square, so the, it will be least. Now, question, second question is that if we have say yi or data set, we have data set of n values, xi, yi, so yi, now question is if we would like to fit a curve from the data which we have recorded from the experiments, we really do not have low guess of priori, what kind of curve we are expecting to have.

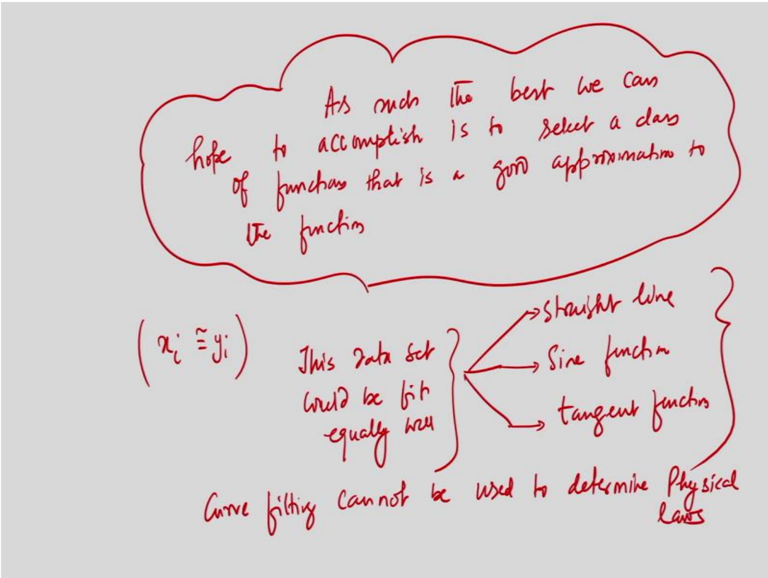
So, that means we have a, we have a data set of n values xi and yi, but from the data set, we do not have a guess, no idea, we do not have idea, a priori, the general way of determining exact functional relationship. So, I am writing that, that will help you. So, given the data set, we have no a priori. This is the case. So, we have data set of the n values, but from the given data set, we do not have idea, a priori, the functional relationship between the data points rather the exact functional form the data set will satisfy.

So, what we need to do, why? Because I am writing, why we do not have a priori you know guess, or information about the functional relationship between the data set because see xi and yi

I will discuss, because a large number of functions can be made to fit a finite number of data points, number of data points equally well.

So, from the given say, xi is equal to yi. So, I will write that, okay I will write. So, if we have a data set xi and yi, we and we, we have no of a priori, general way of determining the exact, exact functional form that, the data set you know will satisfy the reason is that large number of function can be made to fit a curve of the given data set, rather I can say, in a equally or fairly well manner.

(Refer Slide Time: 15:40)



So, what we need to do that is important that as such you know the best we can hope, hope to accomplish is to select a class of functions that is a good approximation to the function. So, this is what we can do that we can hope to accomplish, is to, is to select a class of function that is a good approximation to the function.

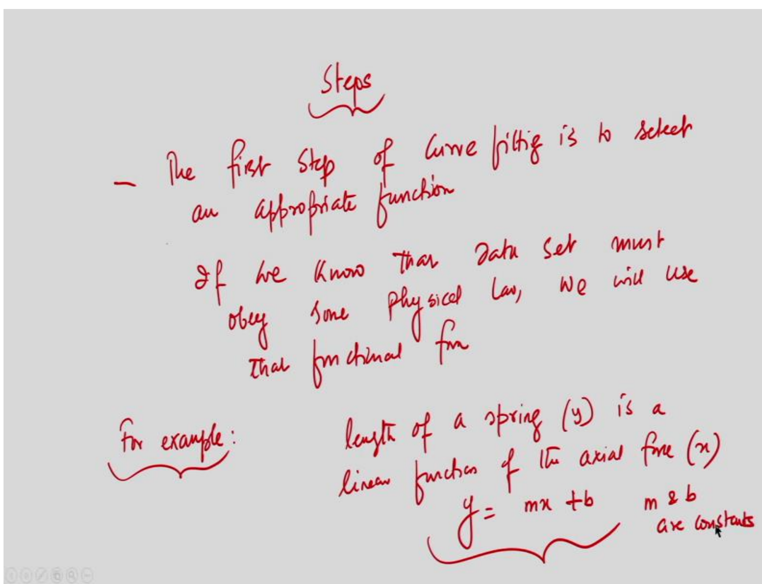
So, now, xi equal to yi. So, if we have a data set, xi almost equal to yi. Now, as I said that large number of, if I go back to the previous slide, large number of function can be made fit of finding number of data points equally well. If the data set is like this, we can see that a straight line can be considered. So, that this data set, this data set could be fit, could be fit equally well by a straight line number sine function or it you know, tangent function.

So, we can see this, the given data set say, we have xi is almost equal to yi, this data set could be fit equally well by, by a straight line, by a sine function or by the tangent function. So, we can see that we really, if we do not have a priori guess, or information about the functional relationship, that the data point should be satisfied, it will be, it is be, it is very difficult.

So, we should have a you know I can say that idea or if we can predict, if we can say that the data set must obey the physical law, then we, it is very easy. So, if this is not the case, then we really do not know whether this data set will be fitted by a straight line or by a sine curve or by a tangent curve.

So, what I can write the curve fitting cannot be used to determine physical laws. So, curve fitting cannot be you know used to determine the physical laws, fine. Now, so what are the different steps if we have a data set, what are the different steps we need to follow to fit a curve that we need to know.

(Refer Slide Time: 19:19)



So I am writing that steps which you need to follow to fit a curve using a given data set. The first step of curve fitting is to select an appropriate function. Rather, I can say this is the first problem, the first problem of the curve fitting is to select an appropriate function, if we know the appropriate function, then the analysis will be relatively easy, but if we do not know then it is very difficult.

So, if we know that, if we know that the data must obey the physical law, then we can use that functional relationship. So, we have not given data set, if we do not have a priori guess or idea about the functional relationship, then it is very difficult and that is why what I have written in the last slide is that the curve fitting cannot be used to determine the physical laws.

But if the case is something else, that something different that if we know that the data set, the data set must obey some physical law, we will use that functional form. So, this is very important that means, if we know that the data you know obeys some physical law then we can use that functional form.

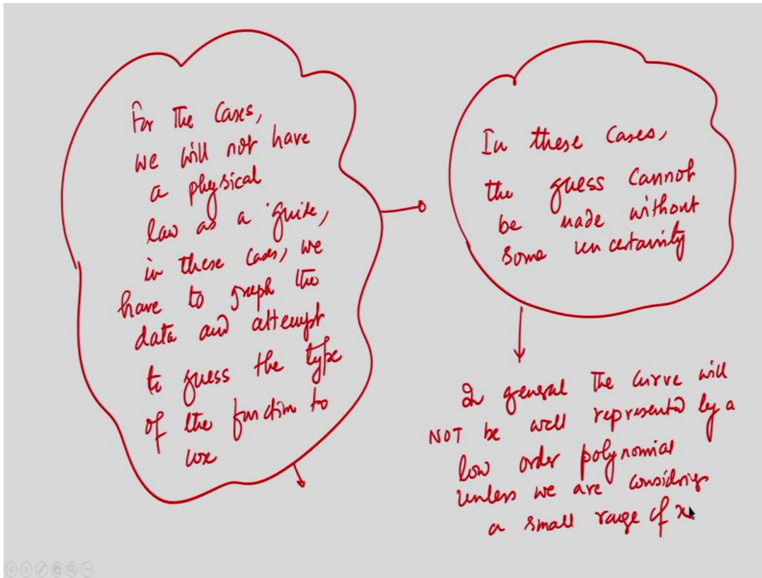
For example, that is, what is very important for today's class is that I have discussed that our analysis will be on a single variable function. So, fitting at curve from a given data set where function is a single variable function. For example, you know if we may, if you know that length of a spring which is output y is a linear function of the axial distance or axial force x.

So, we know the length of a spring is a linear function of the axial force that is y is equal to mx plus b. So, this physical law, so if we have a data set xi and yi, that means if we keep on changing the axial force, and we would like to calculate rather we will measure the axial length of the spring, you know expansion of the spring length that we will calculate.

So, maybe we have 10 different force, 10 different values of axial force that we have considered for the measurement of the spring coefficient that is what is done in the experiment. Now, if we have 10 different values of axial force, we have considered to measure the length of the spring, then I can have the straight relationship, because I know a priori that the length of the spring is a linear function of the axial force, that is a physical law.

If we know a priori that, that will be satisfied by a straight line, then we can consider y is equal to mx plus b where m and b we need to calculate from the data points and that is the curve fitting analysis. So, now we will be discussing this case and we will see how we can obtain the value of m and b essentially to fit a curve where the difference between curve and data points should be minimum using the method which is known as the least square. So, if we now go to, so if I write it, right here that m and b are constants. Now, this is the case.

(Refer Slide Time: 24:06)

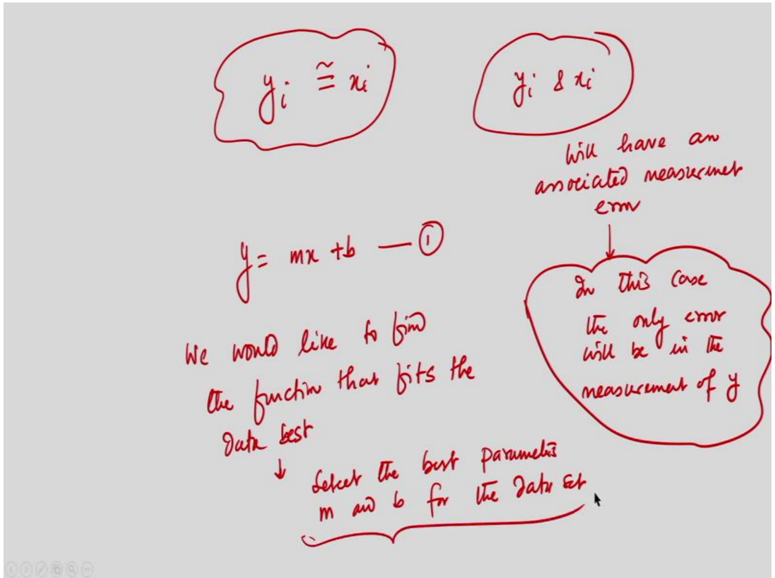


So, second question is, you might ask me a question that if we do not have a priori, you know, guess about the functional relationship then what, what should be done? So the, for the cases I am writing for the cases we will not have a physical law as a guide, in these cases we have to graph the data and attempt to guess the type of the function to use, type of the function to use.

So, this is what is very important for the cases, where we, we do not have a physical law as a guide. Now, what is important in these cases, rather I am writing here in these cases, you know the guess cannot be made without some uncertainty. So, in this case, the guess cannot be made without some uncertainty, so that means some uncertainty is inherent with this case, with these cases.

So, in general the curve, this is a guideline I am writing, in general the curve will not be well represented by a law order polynomial. Unless, we are considering a small range of x, that means, the guideline is that we need to consider high order polynomial, we can consider low order polynomial, if we have small range of x, so, this is the guideline. Next what we will do, we will consider this example, that we have considered y is equal to mx plus c.

(Refer Slide Time: 28:09)



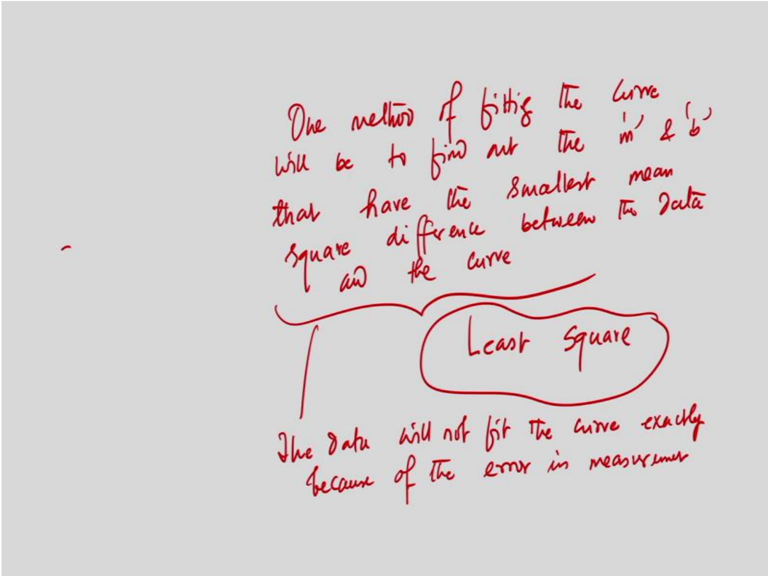
So, as I said today, we will consider these curve fitting analysis for a you know single variable function, if we have data set, yi and that is almost equal to xi. So, this is the data set we have. Now, this yi and xi these data points will have an associated measurement error. So, that means, if we are you know varying xi then we will calculate yi but the measurement error those are you know intrinsic, we cannot eliminate them, but we can reduce that is what we have discussed. So, the xi and yi will have an associated measurement error.

So, as I said that we know a priori that this can be already represented by a straight line. So, we are considering y is equal to mx plus b and that is equation number 1 and this m and b these two are constants, so we need to calculate from the given data set and we will, we can and our objective should be to select m and b in such that the difference between curve and the data point will be have minimum.

So, this is error, but what I would like to say that you know, here in this case, as I said that we have measurement error in this, in this case, we will be varying xi but the measured value yi will have an measurement error, we will have a measurement error. So, but in this case, the only error will be in the measurement of y.

So, this is the case and as I said that the straight line y is equal to mx plus b. Now we have to calculate m and b. Now, we would like to find, we would like to find the function that fits the data best. Our objective will be like this, that the function that fits the data best, now functional relationship that we have already considered that is straight line. Now, in this straight line, we have 2 different constants that is m and b. So, so I am writing select the best parameters, m and b for the data set. So, we have to select the best parameters m and b for the data set.

(Refer Slide Time: 31:48)



And if we go, one method of fitting the curve will be to find out, to find out the m and b that have the smallest mean square difference between the data and curve.

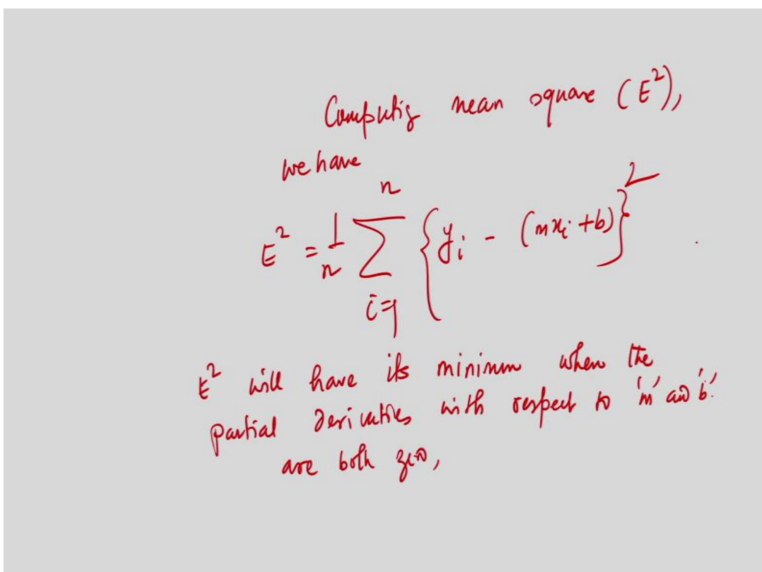
So, one method of fitting the curve. So, what we have understood from the last slide that we should have best value of m and b and one method of fitting the curve will be to find out the best parameters that is m and b that have the smallest square, mean square difference between the data and the curve and that is why this method is known as least square.

Smallest square difference. So, that is why the curve is known as least square. So, this is one of the methods which is used to fit a curve from the data set and in this method what that, what is ensure is that the you know constant m and b what we are trying to find out, these two parameters should have the best value.

I mean these two parameters should have the you know best parameters so, that the difference that is the, difference between curve and data should be, should be minimum. And using this method it is ensured that the smallest mean, mean square difference between the curve and data and that is why the method is known as least square. What I would like to say that is very important that the data although we are telling that we are not telling the difference will be 0, but we are telling that the smallest mean square difference, the difference between curve and the data point should have minimum difference.

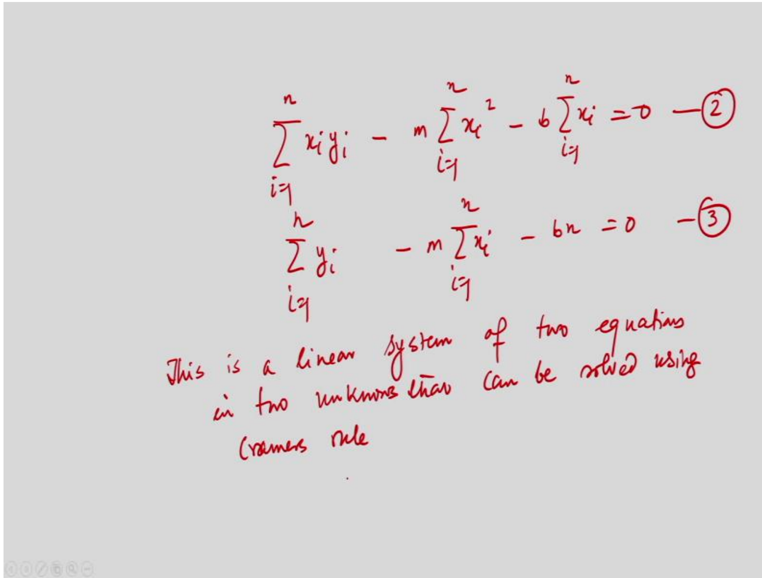
So, we are telling the difference will be 0, why? That is the data you know the data and that is why the data, so I am writing and that is why the data will not fit the curve exactly, because of the error in measurement. So, we have the measurement error and that is why the data set will not fit the curve exactly and we should have the smallest difference. But our objective should be to make the difference more smaller. So, now if I go to the next slide. So, as I said that fitting the curve will be to find out m and b that have the smallest mean square difference between the curve and the data.

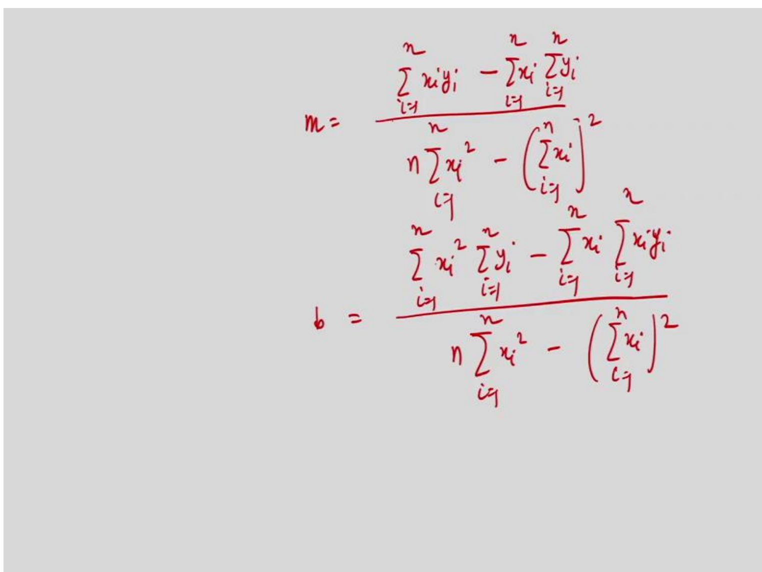
(Refer Slide Time: 35:28)



So, and the computing mean square, E square, I am writing this mean square is E square. We have E square equal to summation of i equal to 1 to n,1 upon n, yi minus mx plus b whole square. So, this is the mean square. Now, because so we need to find out m and b. So, that we know from calculus that E square will have its minimum, when the partial derivatives with respect to m and b are both 0, that we know from the calculus. Thus, if we do, rather if we perform the differentiation, we have the following two equation. So, E square will have its minimum when the partial derivatives with respect to m and b are both 0.

(Refer Slide Time: 37:08)





And if we do, if we will perform this, then I can write these two equation that xi, yi minus m summation of xi square b into xi, i equal to 1 to n, i equal to 1 to n, i equal to 1 to n that equal to 0, and summation of yi minus summation of xi, i equal to 1 to n, i equal to 1 to n minus bn equal to 0.

So, this is equation 2, this is equation 3. Now, so, this is, this is a linear system of two equations with twonknowns. So, this is a linear system of two equations in two unknowns that can be, you know, confidently solve using the Kramer's rule. Now, if we solve these two equations using Kramer's rule, we will get the value of m and n.

So, value of unknowns that can be solved using Kramer's rule, and if we solved and then we will get m equal to you know summation of xi, yi minus summation of xi into summation of yi, i equal to 1 to n, i equal to 1 to n, i equal to 1 to n divided by you know n summation of xi square i equal to 1 to n minus you know summation of x i, i equal to 1 to n square.

So, this is m and similarly we will get b you know that is very important b. So, if we write the value of b, b will be. So, this, the equations can be solved conveniently using Kramer's rule and this is not very difficult. So, this will be xi square yi, i equal to 1 to n, i equal to 1 to n minus summation of xi, i equal to 1 to n summation of xi, yi, i equal to 1 to n yi divide by n xi square i equal to 1 to n minus xi, i equal to 1 to n square.

So, these m and b these two value and values of these two parameter, value of these two parameters is important that means, the parameters m and b are the important parameters for the curve such that and our objective was rather our objective is to find out the value of m and b in such a way that these two values will have you know, that is what I have written, the you know the, the minimum smallest mean square difference between the curve and the, between the data and the curve.

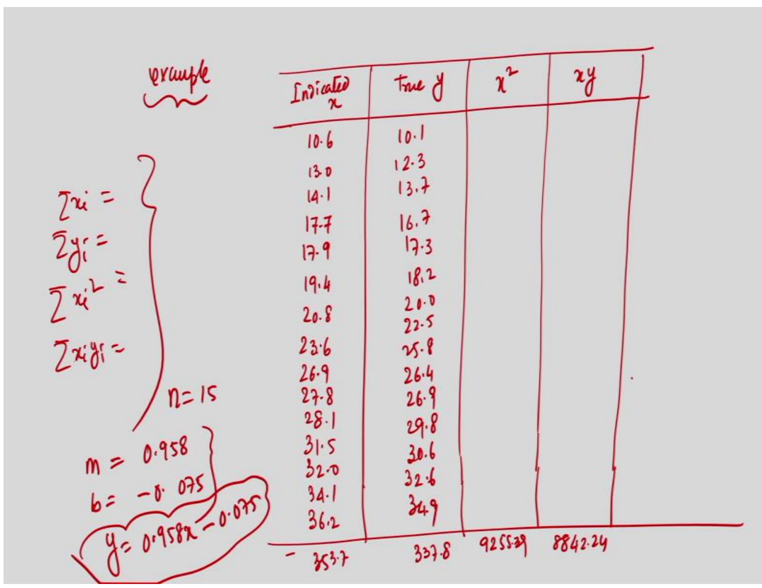
So, this is very, you know, important. Now, so, if we have a data from there we can calculate what will be the value of m and n and finally, we can fit the curve and after fitting the curve, we can check whether the distance, for a given value of xi, what is the value of y, yi, we are getting from the experiment and from the curve.

And we can check whether the difference between the experimentally measured value and the value is predicted by the curve is minimum or not. So, I can now you know, workout on example, to check the efficacy of this method. That means, what I said that, from the given data set, we can calculate the value of yi.

So, that is yi is equal to xi, if we, if you consider the example that I have given that is the displacement of a spring. Now, if we keep on changing the load, you know force xi, then we can calculate the change in length. Now, for a, and from the data set we can fit a curve that is a straight line that we know from the physical law.

Now, what we will do from a given value of xi we know from the experiment what is the value of yi. So, that value and the value for that xi from the curve, if we try to find and we can check whether the difference between these two values is minimum or maximum. So, now we can have the example.

(Refer Slide Time: 42: 20)



So, this is very important. These four different columns we need to consider. So, indicated x, true y, if we go to the previous slide, we can see, we need to calculate x square and xi, yi. So, these two parameters you have to calculate, either you have to calculate xi square and xi, yi and also summation of xi.

So, find true y. So, we can calculate x square and xy. Now, for the values of indicated values of 10.6. So, we may have a few data points, 13.0, 14.1, 17.7, 17.9, 19.4 you know 20.8 if we have a few in order, you know the number of data points, if the number of data points is large then the curve which will be fitting will be very close to the data point that means the difference between the curve and the data points will be minimum.

So, 20.8, 23.6, then 26.9, 27.8, 28.1, 31.5, 32.0 and 34.1, 36.2. So, if we calculate and the measured y that is 10.1, 12.3, 13.7, 16.7, 17.3, 18.2, 20.0, 22.5, 25.8, 26.4, 26.9, 29.8, 30.6, 32.6, 34.9. So, the value of y, measured value of y for the given value of x, these two you know data sets are given. We can calculate x square and xy and ultimately, here, you know, summation of xi is 350, but I am writing over here that is, this is 9255.39, 8842.24, 337.8 and 353.7.

So, this you know, if we just you know, summation of xi, summation of yi, summation of xi square, summation of xi, yi. So, these four you know values we have calculated from this table and if we find out the value of m and n, total number of n is equal to 15, 15 data sets are, data points are there. Then if we try to calculate using the formula that we using the expression we have derived in the last slide, m and b will get m equal to 0.958 and b equal to minus 0.075.

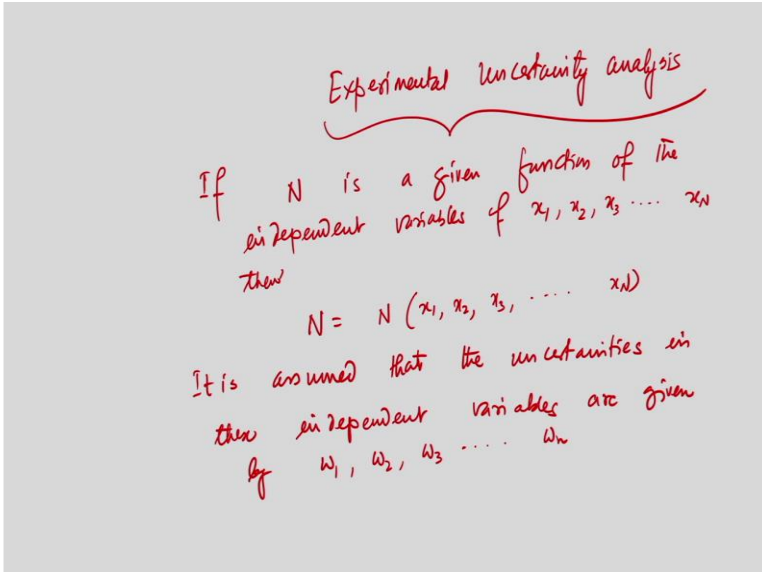
So, that means, our straight line will be, equation will be y is equal to 0.958x minus 0.075. So, this is the fittest curve. Now, what you can check for a given value of xi, you know from the given values of xi in the table, you know what is the value of yi that you have obtained from the experiments.

Now, if we just check whether the fittest curve is correct or not, what we can do for the guess, for the same value of xi can calculate y from this equation, and we can compare the calculated value from this equation and the measured value using experiments, whether these two values are you know closer or not and from there we can check the, you know I can say efficiency of this line in predicting data points.

Now, the, what I said in the beginning of this class is that, suppose we would like to measure the value of y for a given x which is in between 10.6 and 13.0, seems we did not do the experiments in between, for this value, what we can do, we can now calculate, we can predict the value of yi from the curve which we have predicted using this formula.

So, with this I complete this part that and that is why and I have already discussed why it is called least square. Now, finally, I would like to discuss another important part of this you know course is that, uncertainty analysis. So, we have discussed about the statistical analysis from the, you know for the last 2, 3 classes.

(Refer Slide Time: 48:16)



And finally, we will discuss about what is the uncertainty analysis, other experimental uncertainty analysis. Experimental uncertainty analysis is very important, we have seen that the measured I mean, what he can say that uncertainties will be always there. So, if we need to calculate what will be the uncertainty in the measured, you know measured quantity, just I can give you an example that if we are try to measure velocity by measuring pressure, so why we are measuring pressure? Maybe we are measure using a U Tube manometer.

Now, the equipment itself will have some you know uncertainty also we can have kind of uncertainty in measuring the distance or the change in height of the liquid column in the U Tube manometer. So, these two uncertainties will leads to an uncertainty in the measured quantity. So, we need to ultimately calculate what will be the uncertainty in the final measurement.

So, just I am giving an example that if capital N is a given function of the independent variable of x1, x2, x3 upto xN, then N is equal to N x1, x2, x3 upto xN. So, this is the example that I have given you, say y is equal to root x that is what a single variable function, by measuring pressure you can calculate velocity.

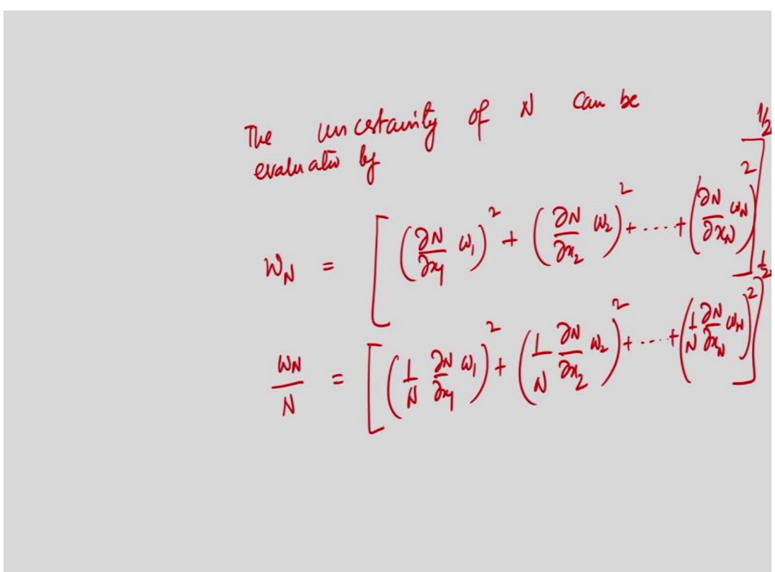
So, while you are measuring pressure, as I said, the uncertainty is you know inherent to the measurement, we are you know measuring the distance in the U Tube manometer. So, that when you are measuring distance, we may have kind of error, but ultimately we will have uncertainty in the measurement of the pressure and that uncertainty, uncertainty will leads to an uncertainty in the measurement of the velocity.

So, just in mathematically if we try to calculate it, then how we can. So, if the example that I have given you that is single variable, but it is not a case always that we will always have a single variable function. So, in a genetic way, we can calculate that if N is the given function of the independent variables of x1, x2, upto xN, then and also it is assumed that, that the uncertainties in these independent variables are given by w1, w2, w3, upto wn.

So, instead of say if you consider example y. Now y is equal to root x plus root x2, root x1 plus root x2 that means, if just I am giving an example, so the uncertainty will be there in measuring x1, uncertainty will be there in measuring x2. Now, that this total uncertainty will be, the uncertainty, uncertainty in measuring the variable x1, x2 will be, the total uncertainties in measuring x1 and x2 will be the uncertainty of the y.

Similarly, if we consider here that the independent variable are given by uncertainties in these independent variable, you know variables are given by w1, w2, w3, and wn, then the uncertainty of N can be evaluated. So, we need to ultimately find out uncertainty in N, similar to what we are, you know seen in the context of y is equal to root x.

(Refer Slide Time: 52:47)



So, I can write in the next slide that the uncertainty of N can be evaluated by WN is equal to del N del x1 w1 square plus del N del x2 w2 square plus del N del xN wn square to the power half, to the power half. So, this, this we can write this equation further that wN by N. This is the uncertainty in the uncertainty of N that is nothing but 1 upon N del N del x1 w1 square plus 1 upon N del N del x2 w2 square plus 1 upon N del N del xN wN square to the power half.

So, this is the you know uncertainty of, so if N is ultimately our measured quantity and to measure N, if we need to you know if N is the function of several independent variable, then the several independent variable will have several uncertainties and ultimately, if you would like to predict the uncertainty of the measured quantity N, then we can use this formula.

So, to summarize today's class, we have discussed about the curve fitting analysis and we have discussed that the curve, we have discussed the objective of you know curve fitting, why you need to go for the fitting curve, fitting a curve from the available data. And if you would like to fit a curve then what are the important points we need to know that if we do not have a prior guess about the functional relationship, then we can assume, we can grab the data and from there we can predict.

But if we know a priori about the functional relationship, then the task is very easy. And finally, we have walked out and we have discussed about the least square method. And using this method if you would like to fit the curve using the available data, then what are the you know procedures. And finally, we have taken up one example and we have calculated like the, we have calculated the, we have that we have analyze curve fitting method.

And we have seen that if the, you know function which you have considered is a single variable function, and from there we have tried to calculate rather than we have tried to fit the curve using a straight line and finally we have calculated the constants. And we have discussed that for, from a given value of xi, we know the experimentally measured yi. And for a same value of xi if you would like to calculate the yi from the curve, we will get one value and we can compare these two values to check whether the fitted curve is good enough or you know, predicting all the output.

And finally, we have discussed about the calculation of uncertainty which is an important inherent part with the experimental methods. So, with this I stop my discussion today and this is the last and today, what I have discussed that was the last module of this course. I hope you have enjoyed all, you know the course material and wish you good luck. Thank you.