Design and Optimization of Energy Systems

Prof. C. Balaji

Dept of Mechanical Engineering

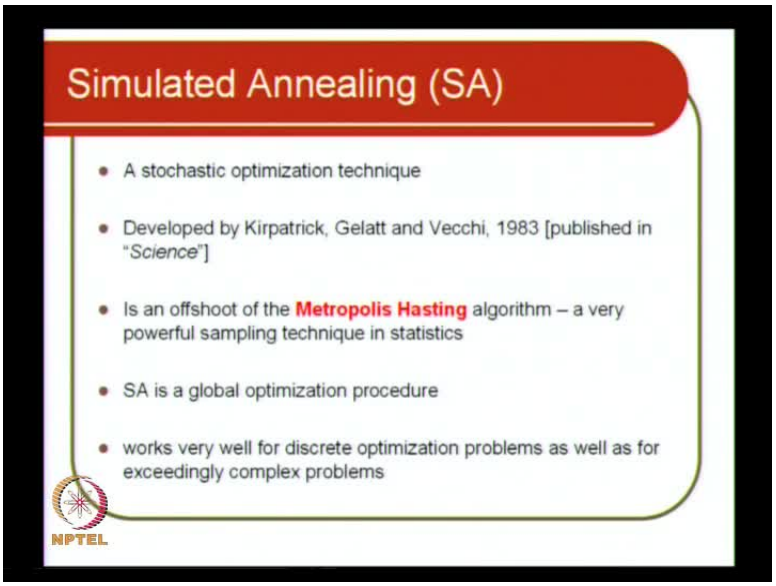
Indian Institute of Technology Madras

Lecture No # 40

Simulated Annealing and Summary

So, we have come almost to the end of the course. So, I will teach you yet another nontraditional optimization technique namely Simulated Annealing, S A as it is known in its abbreviated form. It is a nontraditional optimization technique, in the sense that we do not use regular calculus techniques or we do not use search technique which is based on dividing the interval and eliminating a portion of the interval and so on. However, it is also a search technique, we use some probabilistic laws. So, it is a stochastic optimization technique and apart from that we draw some certain principles used in metallurgy; for example annealing, annealing is basically a slow cooling. So, you mimic the process of annealing in metallurgy. So, it is a simulated anneal; that is why the name Simulated Annealing.

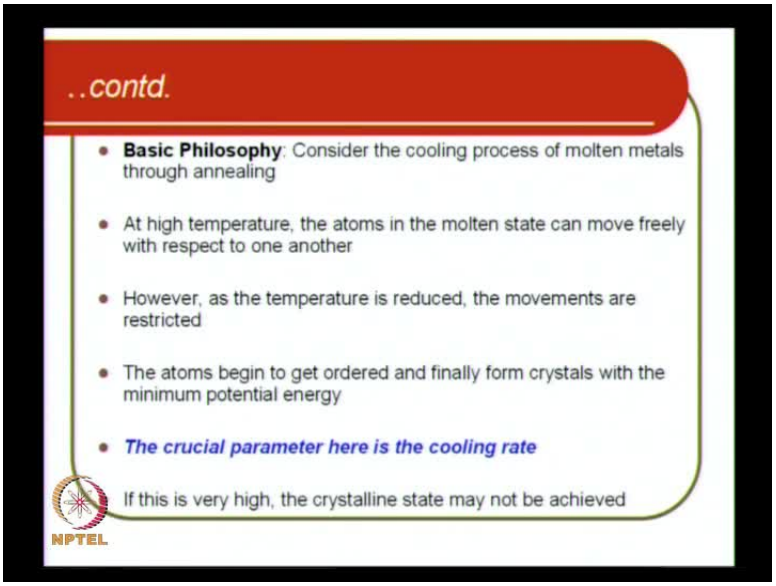
(Refer Slide Time: 01:09)



It is a stochastic optimization technique; that means it is based on probabilistic rules. So, that means it is similar to genetic algorithms in one sense. It is developed by Kirpatrick,

Gelatt and Vecchi in 1983, the article was published in Science. I hope you are aware of this journal Science; it has got one of the highest impact factors considered very prestigious to publish in Science. Then is an offshoot of the Metropolis Hasting algorithm. It is basically an offshoot of this Metropolis Hasting algorithm which is a powerful sampling technique in statistics; you will see that in a little while. So, it is a global optimization procedure just like genetic algorithms. So, there is no premature convergence to local minima or maxima. So, it is very robust. It may not be very efficient in the sense that it will not quickly converge, but it is very robust. Then it works very well for discrete optimization problems as well as for exceedingly complex problems, right; for very very complex problems it works very well.

(Refer Slide Time: 01:57)



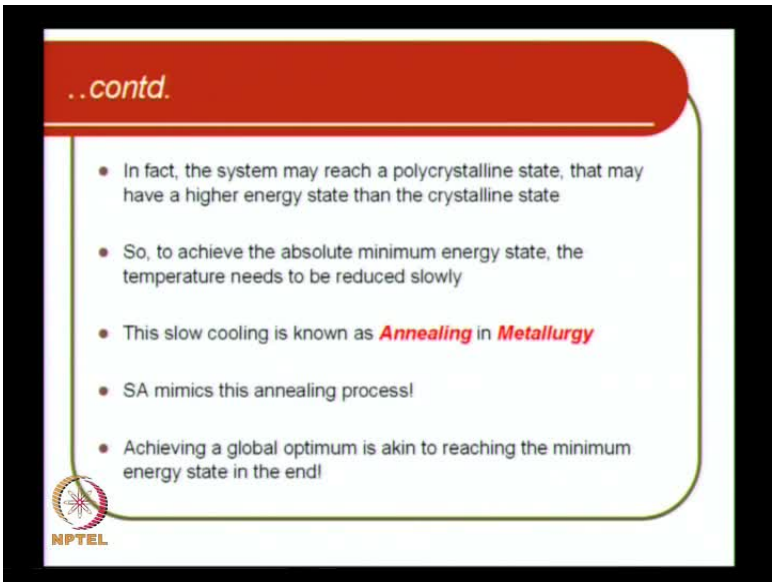
I will put this up on Moodle; the genetic algorithm is already there on Moodle. So, the basic philosophy is, consider the cooling process of molten metals through annealing. All of you have studied annealing, right, in one of the earlier semesters. At high temperatures the atoms in the molten state can move freely with respect to one another, right. At high temperature they have more energy; so, they can move freely with respect to one another. However, as the temperature is reduced the movements are restricted, correct. So, analogously during the initial iterations the samples are free to move anywhere in the domain; what is a sample? If it is a single variable the x; the x can move anywhere in the domain. Just like during the starting process of annealing, the atoms have a probability of being in any state but once the energy level is low the probability of attaining a particular

state becomes low and so on. That is the problem of attaining all states becomes lower, are you getting the point? When the energy is high it has equal probability of attaining any of the states, right. So, that means in short the freedom gets reduced as the energy level goes down.

So, similarly in the initial iteration the freedom is very high. The x, if it is x 1 and x 2, x 1 and x 2 can move here and there but as the iterations proceed the conditions for accepting a particular sample, what is the conditions for accepting a particular sample? That is you are going from x 1, x 2 of i to x 1, x 2 of i plus 1 the condition for accepting it becomes stricter and stricter; that is, quiz 1 you make a paper very easy, then quiz 2 you make little more difficult, the end-semester, you make it very difficult. I mean that is one way of looking at it. So, it becomes more restrictive; it becomes more restrictive as the iterations proceed. So, analogously if you look at the equivalent in metallurgy, the annealing, the atoms begin to get ordered and finally form crystals with a minimum potential energy. You have learned about this minimum potential energy.

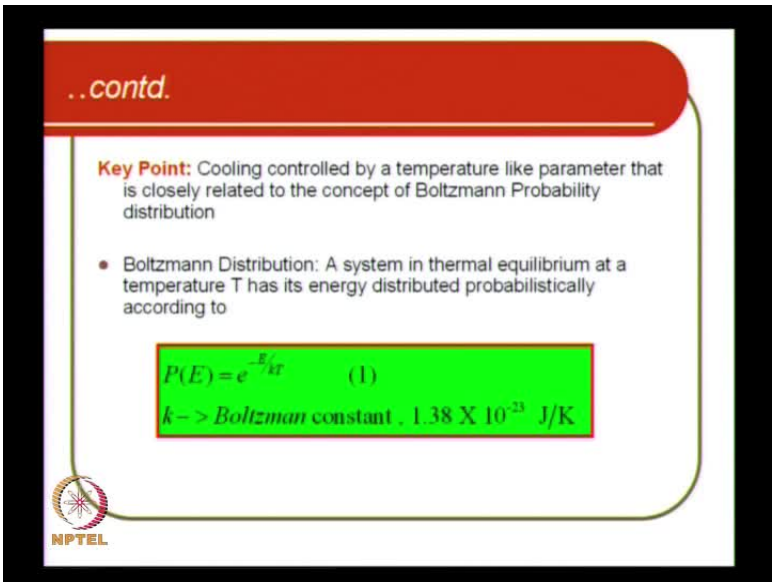
If the cooling takes place very fast, it may not reach that, it may not reach the final state of minimum potential energy, are you getting the point? So, the crucial parameter here is the cooling rate; the cooling rate decides whether eventually you will reach a state which has a minimum potential energy. Therefore, the cooling rate has to be tweaked or fine-tuned or controlled in such a way that you get the optimum end product in metallurgy. Similarly, the convergence rate, the acceptance rate of sample, that is the convergence rate of the algorithm is tweaked or fine-tuned or controlled in such a way that you reach global convergence, right. So, whatever we are saying in English, eventually you have to translate into mathematics and we should able to solve, right. So, if this crucial parameter is very high, the crystalline state may not be achieved. So, this is not what we desire.

(Refer Slide Time: 04:55)



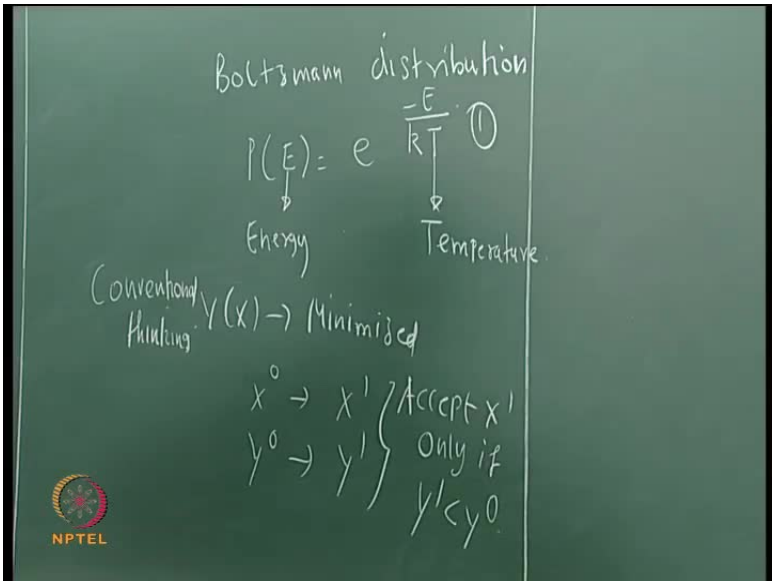
In fact the system may reach a polycrystalline state that may have a higher energy state than the crystalline state, are you getting the point? This is if the cooling is very fast; that means, analogously for the optimization problem you may get a solution which has converged prematurely. It is an optimum; unfortunately it is a local optimum. There is no guarantee that there is a global optimum. So, to achieve absolute minimum energy state the temperature needs to be reduced slowly; the temperature needs to be reduced slowly for annealing. So, this slow cooling is known as Annealing in Metallurgy. The Simulated Annealing mimics this process, mimics this annealing process. Achieving global minimum is akin or equivalent to reaching the minimum energy state in the end.

(Refer Slide Time: 05:45)



What is the key point? The key point is cooling is controlled by a temperature like parameter that is closely related to the concept of Boltzmann Probability distribution. What is this Boltzmann Distribution? A system in thermal equilibrium at a temperature T has energy distributed according to.

(Refer Slide Time: 06:05)



So, the k is Boltzmann constant, T is the temperature. So, this you have studied in Quantum mechanics, right. Now how do we apply this for S A? So, this is equation 1. What does the equation 1 suggests? Equation 1 suggests when the T increases the system

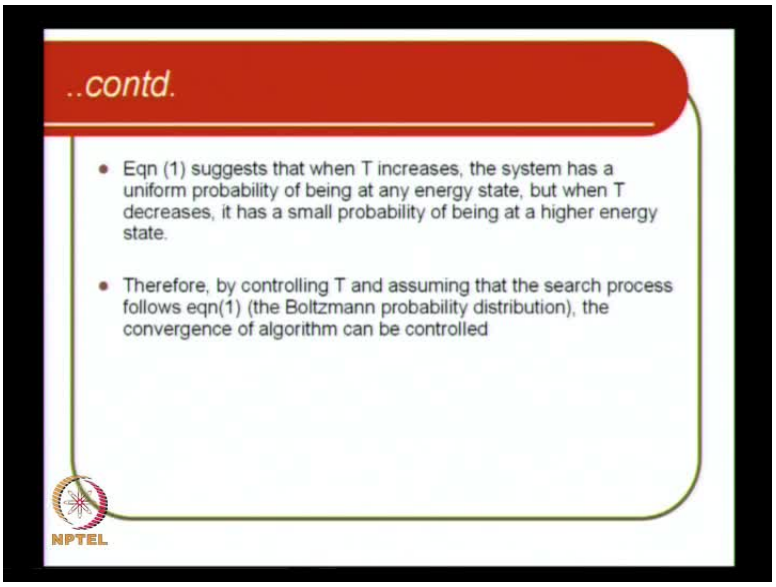
has a uniform probability of being at any energy state. What does it mean? If T is very high what does it mean? It is e to the power of minus of a very low quantity; E to the power of minus of a very low quantity is?

Student: 1.

Student: Close to 1.

It is close to 1.

(Refer Slide Time: 07:22)



So, equation 1 suggests that when T increases the system has a uniform probability of being at any state, but when T decreases what happens is e to the power of minus it becomes a large quantity. Therefore, this P of E becomes very small. When the P of E becomes very small, it has a small probability of being at a higher energy state, are you getting the point? How does it work? It is exponential, whether it is e to the power of minus 4, minus 5, minus 6, minus 7; all are 0 for me, are you getting the point?

Once it has reached a certain threshold, e to the power of minus will approach 0, but if it is e to the power of minus 0.1, e to the power of minus 0.2, e to the power of minus 0.05, there is a chance of getting different numbers. But once you have reached e to the power of minus 4, 5 or then everything will become 0, are you getting the point? Close to 0, I mean your calculator or computer will not be able to recognize that. Therefore, by

controlling T and assuming that the search process follows equation 1 that is the Boltzmann Probability distribution the convergence of algorithm can be controlled.

So, you use a Boltzmann distribution like condition to decide whether the next sample will be accepted or not, are you getting the point? Where is the question, now the question arises what is this, sir; for the first time you are saying something which is different. What is it? What is it that is different in what I am saying? I am saying that, if you are looking for a search algorithm, the conventional thinking is if Y of x has to be minimized, so I go to x 0 to x 1; that is I am going from Y 0 to Y 1. When will you accept x 1?

Student: When Y 1 is less than Y 0.

You put it pretty simple; otherwise, it looks stupid, correct. So, the simulated annealing also there is no problem with this. If you are taking 2 samples, x need not be just one variable x; x is the design vector, x can be x 1 to x n, right. It is a simple notation I am saying. So, when you are proceeding from x naught to x 1 and it translates from Y naught to Y 1, accept x 1 only if Y 1 is less than Y naught. So, what we are doing in simulated annealing is if you seek a minimum, if Y 1 straightaway decreases compared to Y naught, there is no doubt; there can be no doubt in your mind that x 1 has to be accepted.

But the beauty is if x 1 is such that Y 1 is higher than Y naught, do not reject it right away; reject it with a probability. How do you decide that probability? Please use the Boltzmann distribution. How do you use the Boltzmann distribution? This E replace it by e to the power of minus, this should be I think it is e to the power of delta E; that is the change in the energy, right. This change in the energy in the Boltzmann distribution is equivalent to change in the objective function; Boltzmann constant, you can make k equal to 1 for our optimization. So, you make k equal to 1, you make delta E equal to delta Y and T is the temperature.

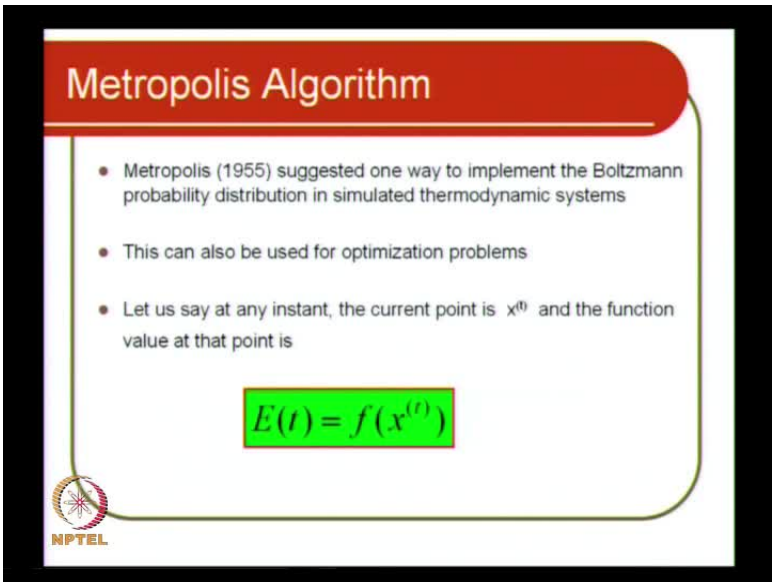
So, this temperature could be there are several ways to depict or represent this temperature. This temperature could be the average value of Y, when you start with a particular iteration. For example, how the system proceeds is like this. Initially, you want to start, what you do is if you have got only one x, let us say x, you will take four values of x, arbitrarily four values of x. Calculate the four values of Y; take the average. We

started the G A was like this. You have a Y bar, you just assume that Y bar is equal to T. Assume k is equal to 1, fine. Now you draw a new sample x 1 from x naught. How do you generate x 1 from x naught? There are several ways of doing it; you can use random number table and Gaussian distribution whatever.

Now you decide whether x 1 is such that Y 1 increases or Y 1 decreases. If Y 1 decreases, very good because you are seeking a minimum, straight away accept it, but if it increases, you apply this probability criterion. So, you will get a number between 0 and 1; generate another random number between 0 and 1, some other number called R. Compare R with P, and then if R is greater than or less than P, you decide a criterion and accept. What are you doing in this? What are you doing this way? Even if the objective function become worse, that is for a maximization problem Y decreases or for a minimization problem Y increases, initially you just allow; let him be like that, allow him to misbehave.

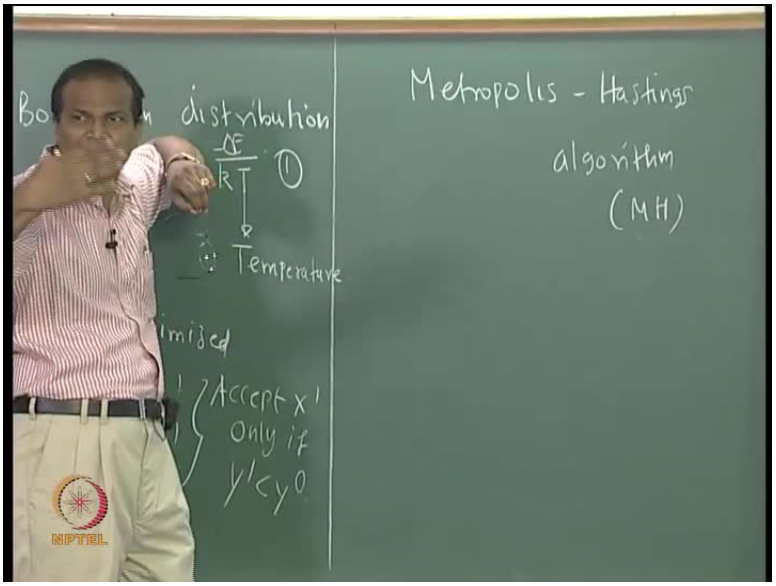
But as you proceed, what will happen is this T will come down because over generations the T will come down because T represents the Y. When T comes down, it has a small probability of being at a higher state. Therefore, compared to that random number the other random number R which you are generating from the table is always varying between 0 to 1. But this random number is also with this P is also varying from 0 to 1. But when the T decreases, the probability will be such that if the function decreases for maximization, or if the function increases for a minimization, as the iterations proceeds when temperature T comes down, it becomes more and more difficult for you to accept the sample. It will proceed like a general algorithm conventional thinking only after initially the solution space has been thoroughly searched, is that clear?

(Refer Slide Time: 13:57)



Now the slides are meaningless. Anyway I have explained the whole algorithm to you without the slide. So, Metropolis 1955 suggested one way to implement the Boltzmann probability distribution in simulated thermodynamic systems. Whatever Metropolis suggested I have already explained to you. So, this can be also used for optimization problem.

(Refer Slide Time: 14:25)

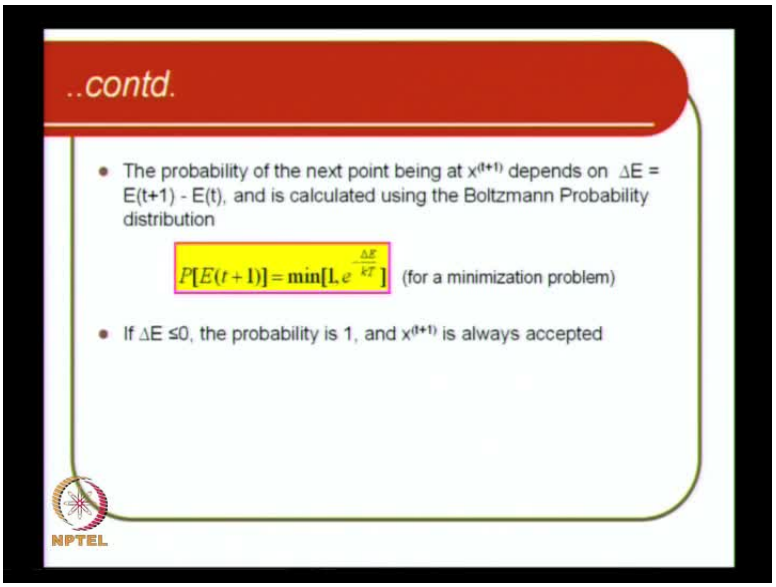


So, let us say that so the Metropolis Hastings algorithm, what I have explained to you now quick in 2 minutes or 5 minutes, the M H or the Metropolis Hastings, it is basically

a sampling algorithm. What is a sampling algorithm? A sampling algorithm is an algorithm which helps you get samples. What is a sample? x naught to x 1, x 1 to x 2. How will you generate new samples? New samples will be based on some laws; that is whether you accept a new sample or not depends on some condition. If this condition is based on the Boltzmann distribution, that is the Metropolis Hastings algorithm.

The Metropolis Hastings algorithm has been listed as one of the ten most powerful algorithms ever developed by man in any field. It can be used to solve a variety of engineering problems because it is actually called an MCMC method; I mean for an MCMC method you use this; that is Markov chain Monte Carlo method. Under MCMC method, the Metropolis Hasting algorithm is one of the most powerful sampling techniques. Okay, we have achieved a lot of success in our research in satellite meteorology, in inverse problems; in our group we extensively use the Metropolis Hastings algorithm. Now let us say that the current point is x t and the function values E t is f of x t, right, that is, y of x.

(Refer Slide Time: 15:54)



The probability of the next point at x t plus 1 depends on delta E. Delta E is your change in the objective function which is E t plus 1 minus E t; that is y of x 1 or y 1 minus y naught and is calculated using the Boltzmann probability distribution.

Now we apply the Boltzmann probability distribution P of E t plus 1 is minimum of 1 e to the power of minus delta E by k T. If delta E is less than equal to 0, that means the y is?

Student: Less than 1.

If delta E is less than 0, y 1 is less than or greater than?

Student: Greater than.

Delta e is less than 0, y 1 is less than y naught; for a minimization problem if y 1 is less than y naught, you want to accept or reject?

Student: Accept.

Accept, okay. So, if delta E is less than 0 the probability is 1 and x t of 1 is always accepted. We are not questioning conventional thinking, but while acceptance is straightforward, rejection is not straightforward. That is the essence. You do not mess up both the things. You accept when it is going in the right direction, but you do not reject if it is not going in the right direction. You reject it with some probability. You do not give you grade straight away. You reject it with a probability and this probability will be such that rejection will become higher, the rejection rate or the rejection will become stricter and stricter as the iterations proceed.

Student: So, the random number we generate will become higher and higher.

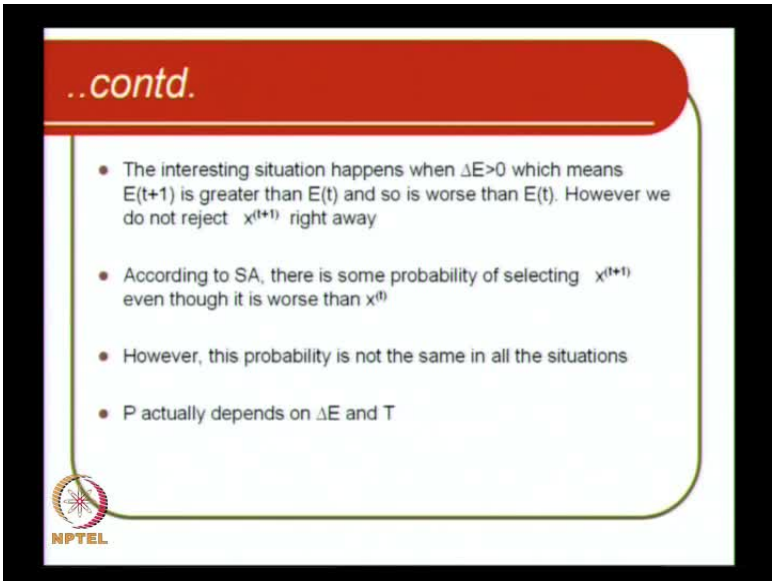
No, random number will become 0 and 1 only; but the P will be such that P will become very close to 0, are you getting the point? Because as iterations proceed, the T what happens to T?

Student: No sir, I am saying the criteria will be.

Criteria same, generate a random number r, r is less than equal to P or r is greater than equal to P. You decide something and stick to that criteria and random number you will continuously generate; random number will always vary between 0 and 1. P will also vary between 0 and 1, but as the iterations proceed P will become closer to 0, but the random number will be between 0 and 1.

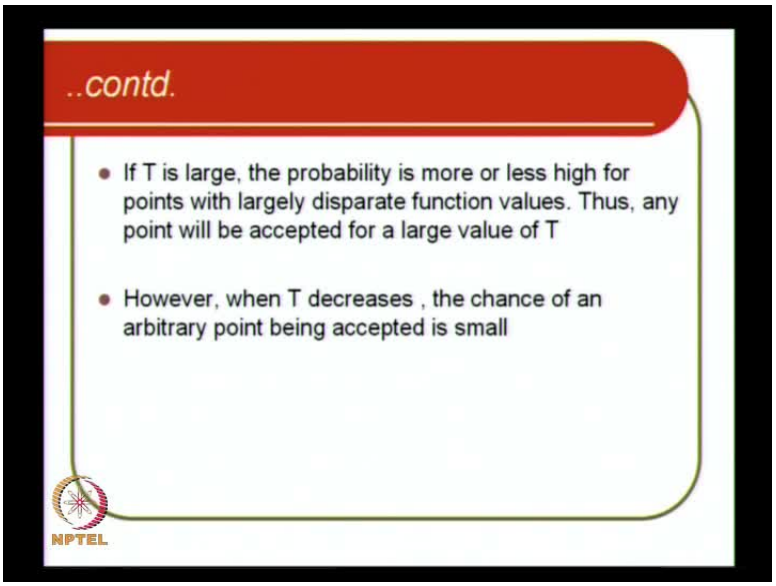
So, if you put a condition r less than equal to P, if r is less than equal to P I accept the sample, are you getting the point? So, r, you can generate 0 to 1. P also it will be 0 to 1, but initially P will be close to 1. So, out of 3, 4 times when this rule is violated, that is y 1 becomes more than y naught you will still accept it. But when the cooling proceeds that is the T is decreasing, then the chance of P becoming high will go down. r has always a chance of going up and down between 0 and 1, are you getting the point, fine.

(Refer Slide Time: 18:51)



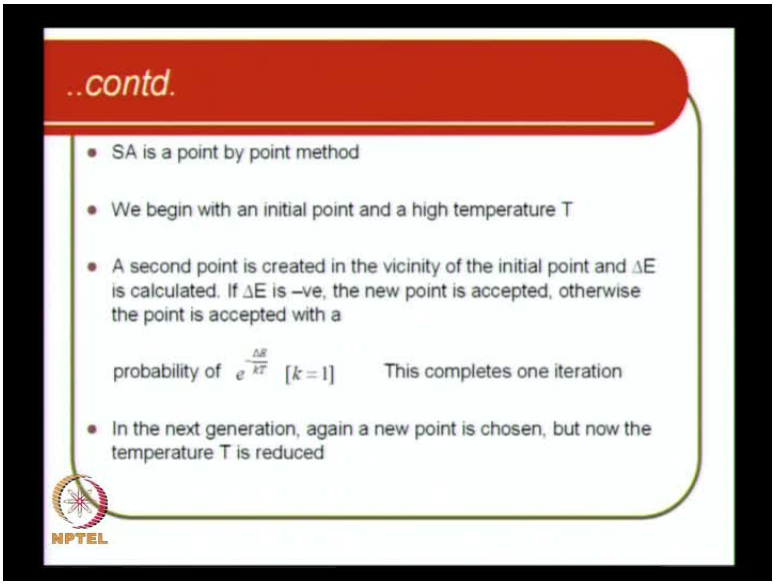
So, the interesting situation happens when delta is greater than 0 which means Y of t plus 1 is greater than Y of t. So, it is worse compared to E of t; however, we do not reject x of t plus 1 right away. According to simulated annealing, there is some probability of selecting it even though it is worse than x 1. However, this probability is not the same in all the situation, P actually depends on delta E and T. That is what I told you.

(Refer Slide Time: 19:14)



If T is large, the probability is more or less high for points with largely disparate functional values. Thus, any point will be accepted for a large of T. Initially, any point will be accepted; however, when T decreases, the chance of an arbitrary point being accepted is small.

(Refer Slide Time: 19:30)



So, S A is a point by point method generally. So, unlike G A we start from one point and follow that point but however, S A for multiple points is also there. That is also possible, but the original S A was a point by point method. So, as usual it is a search technique; we

begin with an initial search point and a high temperature T. The high temperature T is equal to y bar for 3 or 4 values of Y you take the average.

A second point is created in the vicinity of the initial point and delta E is calculated. If delta E is negative the new point is accepted; otherwise, the point is accepted with a probability of e to the power of minus delta E by k T where k is equal to 1, this completes one iteration. In the next generation again a new point is chosen but now the temperature T is reduced; that is you control the cooling rate. For the purpose of this class and in exam and all that you can reduce T by 0.5; each time you can reduce it by 0.5.

Student: Sir, every iteration we reduce or only for?

Every iteration you reduce it by 0.5. You can reduce it 0.25, 0.3 also but for uniformity we will reduce it by 0.5.

Student: Here T represents Y bar, is it? T represents Y bar. So, for maximization problem, Y bar increases.

No, I am explaining the algorithm for minimization.

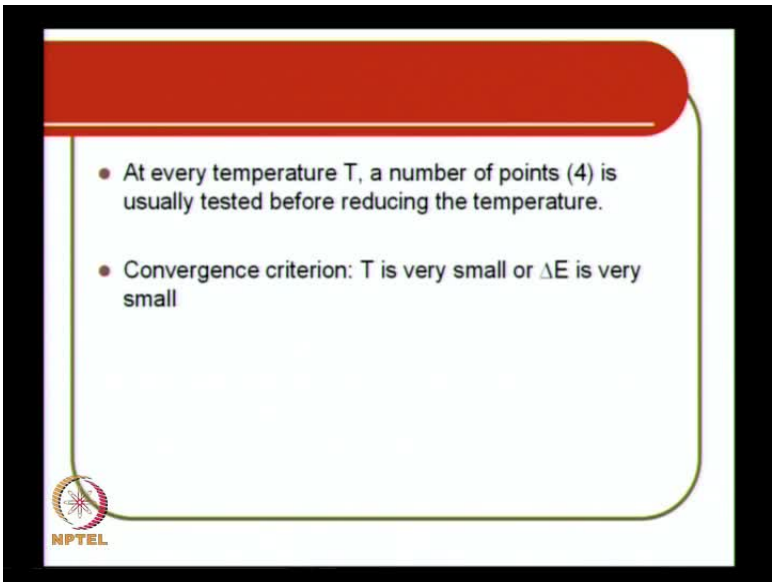
Student: Only for minimization?

You convert it into an equal minimization problem.

Student: Sir, how do we choose the next point?

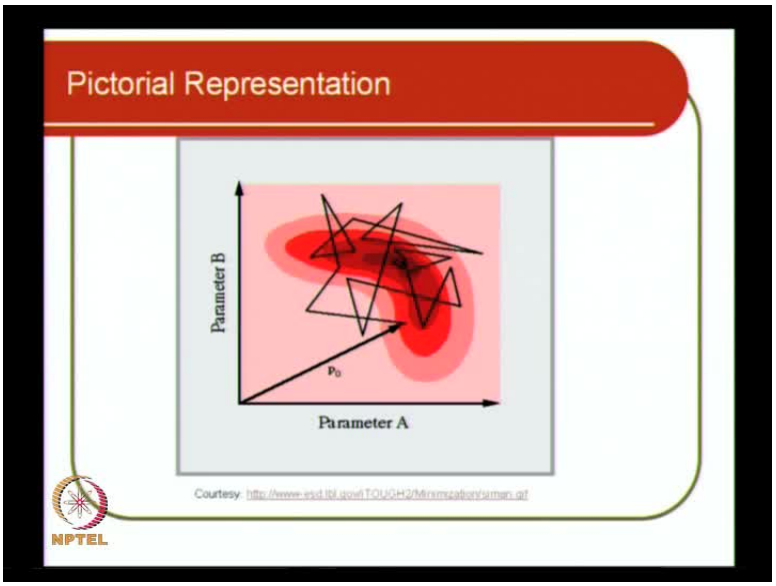
You have to use random number table and I will explain it to you. We will work out a problem, right.

(Refer Slide Time: 20:58)



At every temperature, a number of points are usually tested before reducing the temperature, right. Convergence criterion T is very small or delta E is very small, right.

(Refer Slide Time: 21:06)



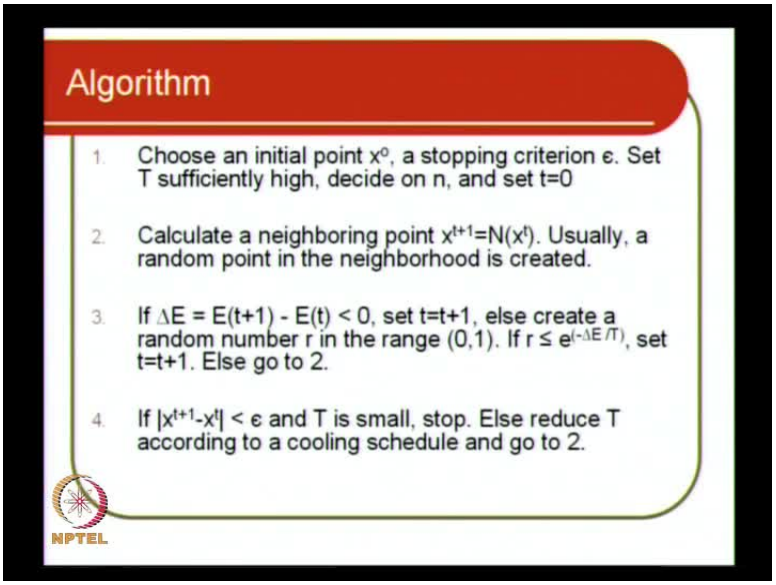
This is how it is; I have taken it from some these things. So, you can see that Parameter A and B, if this is the solution initially it will go zigzag, zigzag, zigzag finally it will reach there.

Student: Sir, in this we are using the Boltzmann distribution here, but there are lot of other distributions which have similar properties what we have seen.

Boltzmann because the Boltzmann distribution works very well for annealing and they have got success with annealing, it has been used in Metropolis, okay.

What is the algorithm? If you want you can write down. It will be better, okay.

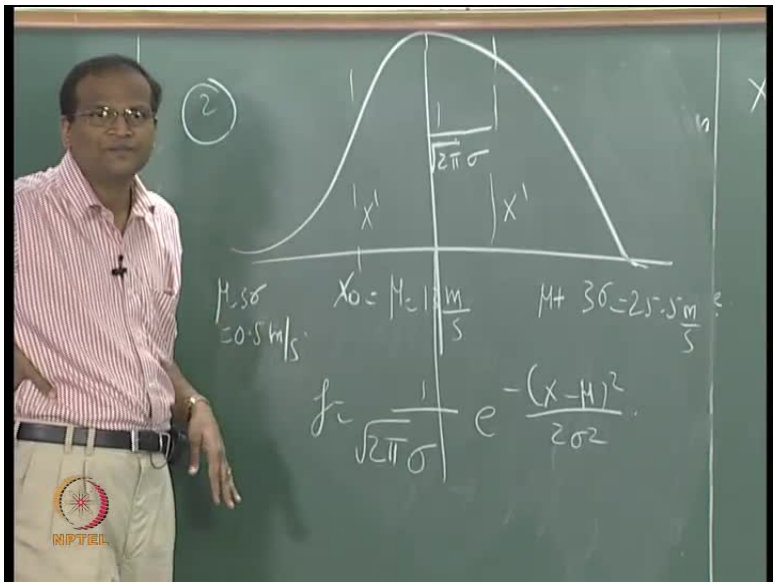
(Refer Slide Time: 21:52)



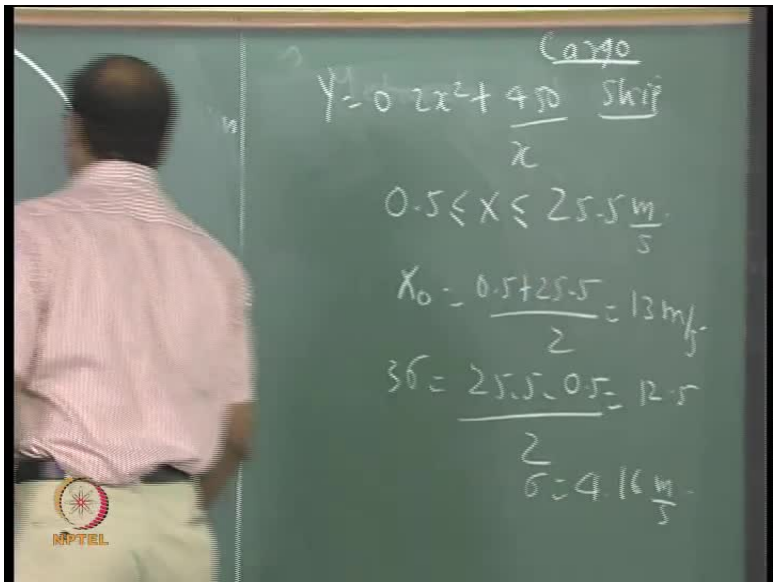
Choose an initial point x naught, a stopping criterion epsilon. Set T sufficiently high, decide on n. Can you just copy this down? It is going to be helpful to you.

So, choose an initial point x naught, a stopping criterion epsilon. Set T sufficiently high, decide on N and set t equal to 0. t is the counter, that is small t is the iteration counter. Calculate a neighboring point x t plus 1 to N of x t. How do you do that?

(Refer Slide Time: 22:46)



(Refer Slide Time: 22:58)



Okay, please look at the board; stop for a while. For example, suppose I want to optimize; so, this is the cargo ship problem, right. Let us say that 0.5, correct. Now what I do is let me take first sample x naught equal to. So, I will take this as the mean. I do not want the samples to exceed; I do not want my samples to exceed 25.5; I do not want my samples to fall below 0.5. So, 99 percent of the time I can do this, if I follow a Gaussian or a normal distribution whose mean is equal to 13 meters per second and whose 3 sigma is given by, are you getting the point? So, 25, 12.5; sigma is 4.16 meter per second. So, I will have mu equal to 13 that is initial iteration. So, I start with mu, that is x naught.

How will you generate x 1 now? Ashutosh you asked this question, right. How will you generate x 1? No, he is not able to, how will you generate x 1? So, if you use the normal distribution f; so, what is the ordinate of this? What is the ordinate of this distribution? When x equal to mu what happens? f equal to?

Student: 1 by root 2 pi sigma.

So, this will be the maximum probability you are getting. This will be the maximum probability you are getting. That will happen when x equal to mu. Therefore, x minus mu equal to 0. So, what I will do is I will use the first column; generate a random number, I will assign that to f. Sigma is known to me, mu is known to me; I will generate the new x but what will be the problem with this procedure if you straight away apply it? The distribution is correct.

Student: You get points which are outside?

No, point is outside; sometimes it may lead to some meaningless this thing because sigma is very high here, are you getting the point? And f is between?

Student: 0 and 1.

0 and 1 but in the maximum f should be only 1 by root of 2 pi sigma. Therefore, you have to use a normalized standard distribution, are you getting the point? Or how else can we take care of this, any suggestions?

Student: We have to generate something like a random number and we have to convert it to whatever.

Yes, how? We have to implement now.

Student: We use the value of sigma and we just add that random number to this to the existing value.

What, what, what?

Student: We generate a random number.

Okay no, no, you have the table you generate the first random number. Okay, then?

Student: From that random number we generate a random normal. We can generate random normal.

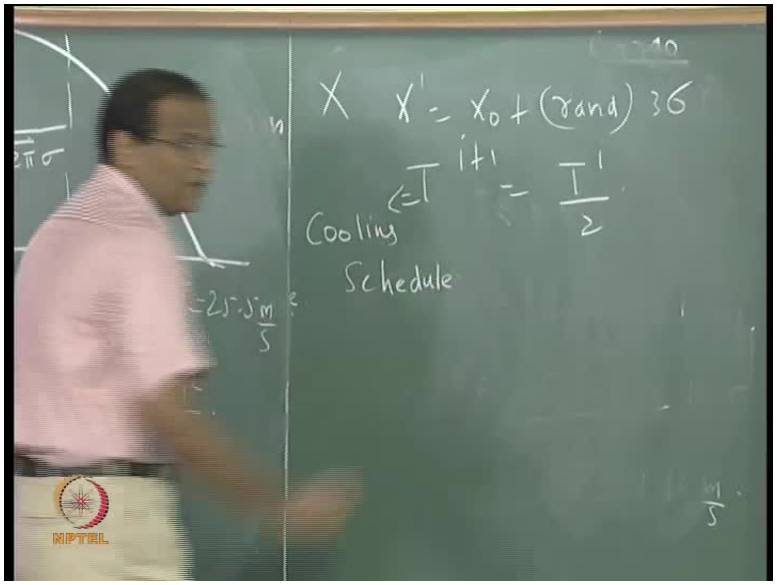
Random normal means what, what does it mean?

Student: Standard normal n 0 1.

Okay. So, I generate n 0 1, I convert that to n 0 sigma. Choose some value of sigma which I want which is some small part of the domain and add that to x and x t to get x t plus 1.

No, no, okay. So, is it like this? If I understand you right, are you saying that?

(Refer Slide Time: 27:35)



If suppose I say x 1, this is not correct, is it not? Then you are not following the random number, is it not? Is it correct?

Student: No sir. But random normal is on both sides.

No, but what is that rand now? Is it the random number you are taking from the table?

Student: No.

Okay. So, I am not sure whether this is, why I am discussing it at length is it is this Simulated Annealing is a research algorithm; it is not discussed in text books. There is

no standard procedure available, it is not written in 10 text books. We solve it in a particular way; I want to see whether first time when I propose this to you, you can come up with something. Now I want suggestions. You still have 20 minutes.

Student: Can you just repeat the question?

Now we want to generate randomly a sample. We have this sample; we want to go either here or here. So, I want x 1.

Student: We can generate a random number between 0 and 1 by root of 2 pi into sigma.

We can generate a random number?

Student: between 0 and 1 by root of 2 pi.

Now we are cooking with fire. So, you can generate a random number between 0 and 1 by root of 2 pi. How do we do that?

Student: You just divide the random number.

You just divide the random number by root of two pi sigma that is it. The problem is solved. Ashutosh, does it answer your question? It leaves you more confused. Then, see if the right hand side is varying between 0 and 1 there is no problem, but right hand side the ordinate is when this become 0 it becomes only 1 by root of 2 pi sigma; that means I am not using a standard normal, I mean normalized normal distribution. I do not want to complicate things; after all I want one sample. Suppose, I want to normalize and every time I normalize my sigma may change; it will lead to a lot of mess. I want to quickly get over to this thing. I do not care about Gaussian distribution, but I want a random sample. I have to follow some rule; that is why I am using this fellow, right.

Now because f can vary between 0 to 1, if suppose by chance I am getting f 0.99, one random number could be 0.9. If I put 99, this may lead to completely arbitrary results, are you getting the point? Because this will be e to the power of minus will be something. So, it will go completely outside the range or it will lead to some silly results. So, therefore it is important for us to keep f between 0 and 1 by root of 2 pi sigma. So, generate a random number, divide it by root of 2 pi sigma and then equate it. Now there is a thing that it is x minus mu whole square. When you generate a new x 1 it will always

go to one side, is it not? Is everybody able to follow what I am saying? I am teaching funda concepts. So, when you take x minus mu whole square, it will always go to one side.

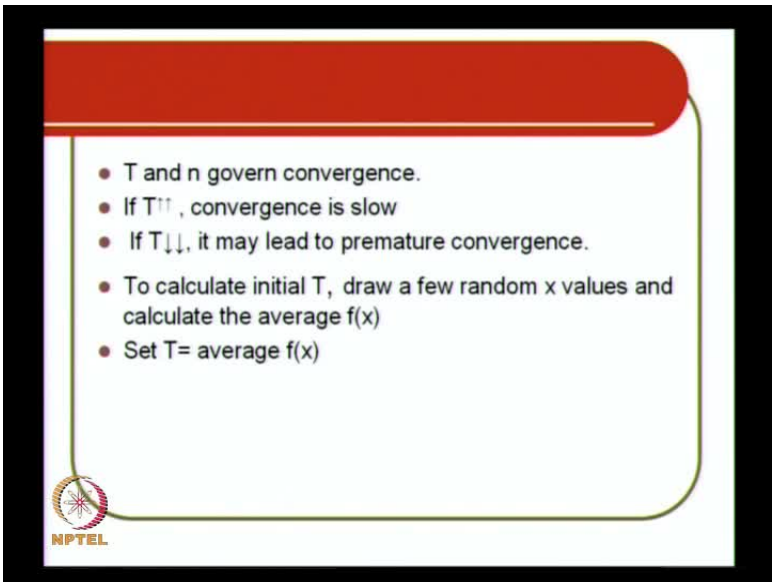
Student: That again you could use another random variable.

You use another random number. Take the random number in the row 5; take, generate another random number gamma or k. If k is between 0 to 0.5 put plus delta x; if it is less than 0.5 minus delta x. So, so many things are required. But this is the variety you are introducing; by putting so many stochastic things, you are doing that zigzag. So therefore, if there is a treacherous function which goes up and down, this fellow will not leave; he will catch him. But your Golden section search and all that will work only for the unimodal; this fellow and GA can catch any fellow; are you getting that point? But they will be slow, but they will catch. And, it is infinitely superior to exhaustive search, because there is some funda based. Exhaustive search, there is no funda. It is funda based, are you getting the point?

Now have you written all the four; everybody through with this? So, calculate a neighboring point x t plus 1 is N of x t; N of x t is the normal distribution which we have seen. So, using a random point in the neighborhood it is created. If delta E is E t plus 1 minus delta T is less than 0, set t is equal to t plus 1; that is move to the new point else create a random number in the range 0 to 1. So, this random number should be in from the row 3 or row 4. In your random number table, if r is less than equal to e to the power of minus delta by T, set t is equal to t plus 1. So, this is P probability r less than equal to P, you can set the criterion. So, please note that you have to use three sets of random numbers for simulated annealing. The first set of random number is for the sampling.

So, you can stick to row 1 and 2 for this. Row 3 of the table you can use for generating r and row 4 or row 5 you can use it for generating that k or whatever which will decide whether the delta x will go to the right side of the mean or to left side of the mean. So, if x t is less than your criterion stop, else reduce and cooling schedule. What is our cooling schedule? Cooling schedule is, so this is our cooling schedule. T and N govern the convergence. If T is very very high convergence is low; if T is very small it may lead to premature convergence.

(Refer Slide Time: 33:32)



Please note this; to calculate initial T, draw a few random x values and calculate the average f of x. This you have to note down, how to generate the initial temperature and then set T equal to average f of x. So, you know the initial sample, initial x will be the mean of the range. Initial standard deviation you know, right. Initial cooling rate I have given you, use random number, generate and proceed it, right. Yes, that is it. Now problem number, for calculating

Student: T.

Yeah, you do not have to do. For second iteration onwards T is equal to T by 2; you do not have to do that. I will give; I will put the question paper on moodle. So, one question on DP, one question on LP, one question on GA, one question on SA; 4 questions, one question on Lagrange multiplier; 5, one question on non-linear regression, right. Then, Golden section and other things depending upon same problem I will say solve it by Fibonacci or Golden section; golden section everybody all of you know it just takes 10 minutes. And then initial simulation I will give a tough problem which will involve some dou y by dou x and all, either Newton-Raphson or I can give Gauss-Seidel, system of 3, simulation is also involved. So, you have a fairly good idea of what to expect, right.

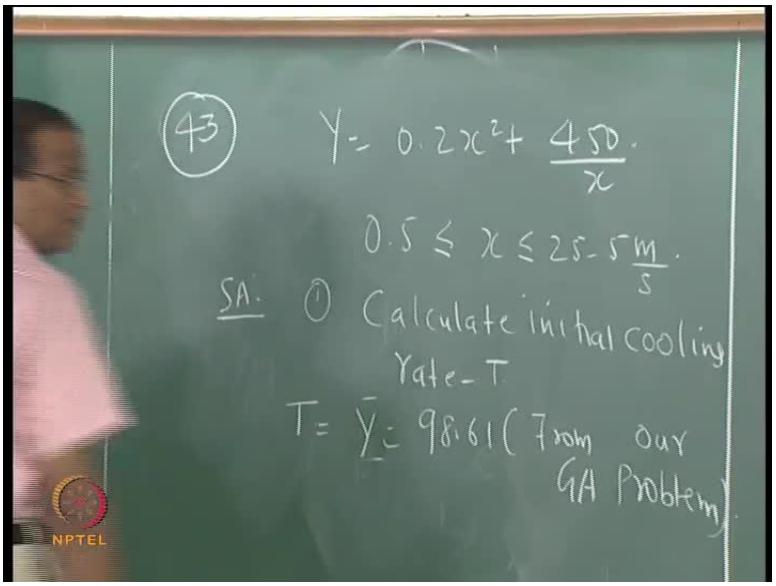
Now we will start solving, because simulated annealing we have not really solved. You should know how to use a random number table. So, problem number 43; consider the cargo ship problem. We would like to solve it using SA. Consider the cargo ship

problem. Anyway, problem statement can be conversational know instead of being very formal and. We would like to solve it using SA. Perform four iterations of the SA. Perform four iterations of the SA, with an initial interval of uncertainty of 0.5 less than equal to x less than equal to 25.5 meters per second. Perform four iterations of the SA for this problem with an initial interval of uncertainty of 0.5 less than x less than equal to 25.5. Use the random number table provided to you.

I want to use the board. Shall we minimize? Shall we put it on standby? I will use it again. So, I told you the difference between B Tech, M Tech, PhD know? You know that; you do not know? Anyway before we start solving before everybody gets. If you think you know everything, you will get B Tech; if you begin to doubt that you know anything at all, you will get M Tech if you are convinced that you do not know anything and you are also convinced that others also do not know anything, you will get PhD.

Now if you are convinced that you do not know anything at all, if you are convinced that others also do not know anything at all and more important you are also convinced that in your lifetime, nobody can ever figure out that you do not know anything at all, you become a Prof. So, 12 years back one fellow was very serious, very serious when I told this. He put his hand up, ‘sir, I have no doubt that I do not know anything, sir’, then he is trying to trap me. Then, why am I registered for M Tech sir, I should get PhD’, he tried to trap me. I said that, ‘you still think that I know’. It took some time for him to understand.

(Refer Slide Time: 38:34)

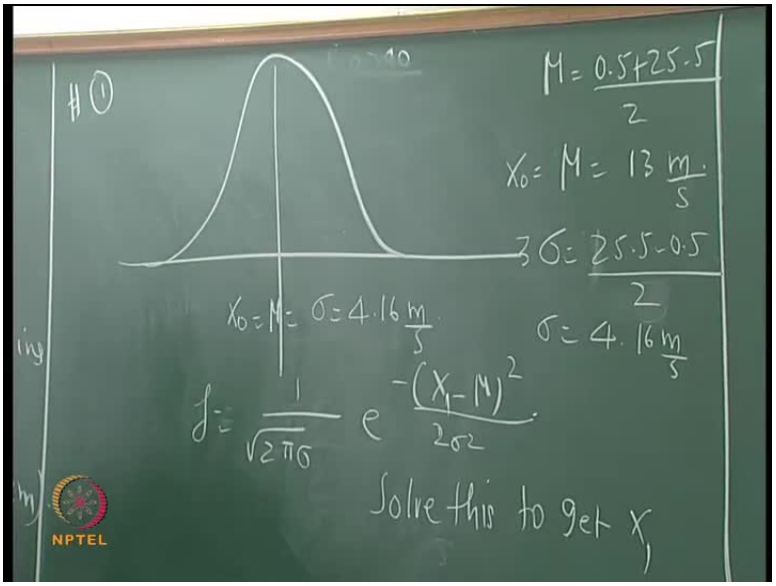


Now first step; that means use 4 values. See all of us calculated 4 values using GA, right; we will use that itself. Go to the problem 42, using GA we started with 4. So, we got Y bar. What was the Y bar?

Student: 96.85.

So, Y bar.

(Refer Slide Time: 39:58)

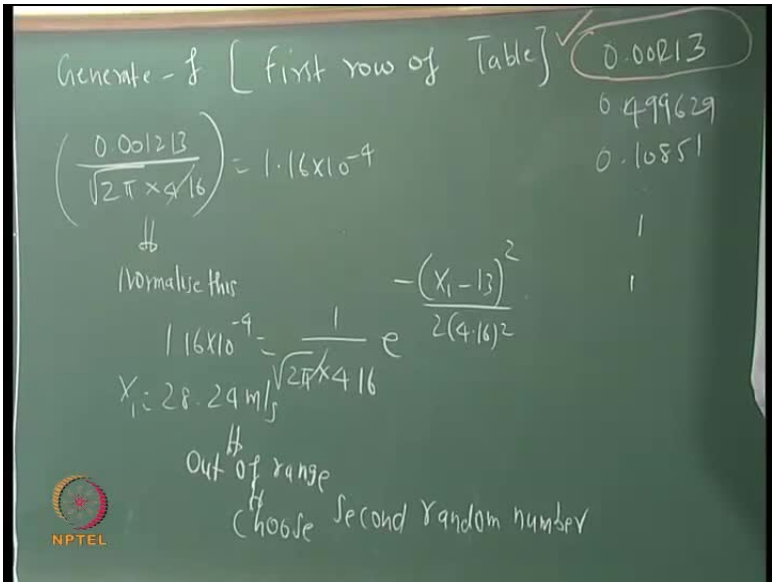


So, now iteration 1; so, every iteration you have to draw this. In the exam also you have to draw this and indicate the mean, only then I will be convinced that you have understood, right. So, X naught equal to mu equal to. So, the X naught, sigma, no, no, no, 3 sigma, right; each time you have to write this, X 1. What is the objective of writing the Gaussian distribution? You have to solve this equation to get? Solve this equation to obtain?

Student: x 1.

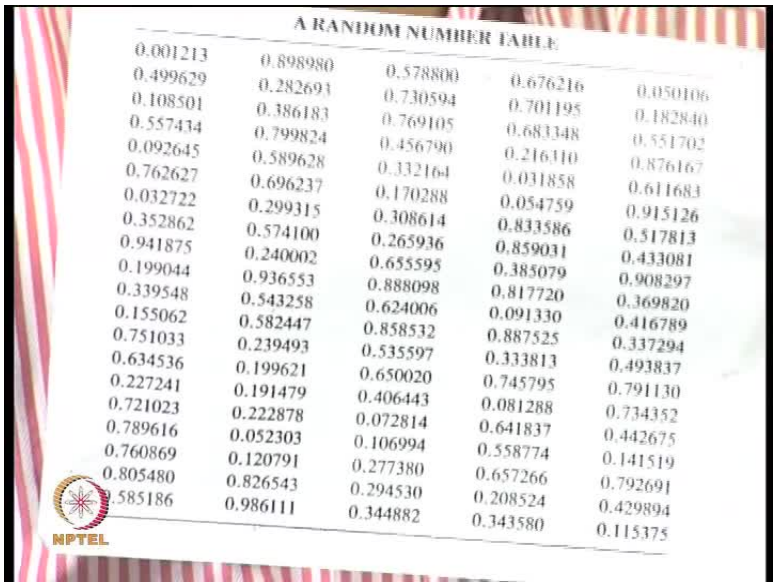
What is it? x 1, good. So, solve this to get X 1 and then apply your funda. What is that funda? e to the power of minus delta by k t. If Y straightaway decreases, no need to apply that funda. If it goes in the wrong direction, you have to apply that funda.

(Refer Slide Time: 41:44)



Now generate f. First row, 499629, can you show this? Please show that because the other students do not have this or you can show this, okay, not for you guys. So, we want to use this.

(Refer Slide Time: 42:49)



This is a random number table. So, lot of such random number tables are available freely available in the internet and also available in the appendices of standard mathematics text books or probability and statistics text books. Here is one such table. So, we will start with this. Generate the first random number. First random number is 0.001213; now that is the f value. So, we use this; divide it by 2 pi normalize this. How much are you getting? Very bad, very small know. So, 1.16 into 10 to the power of minus 4 equal to 1 by, yeah, solve for X 1. Is it okay?

Student: Sir why do we divide by each time and then, here it gets cancelled.

Should we divide it or multiply it?

Student: We should divide it. No sir, actually we can just equate that e power minus 1, because e power the maximum value of that is between 0 and 1.

Yeah, but anyway it is also going on the, you do not have to divide it. When you are actually doing it, get rid of the 1 by 2 pi sigma, to make it very formal and since this lecture is also going out, I will. What they are saying here, what people are saying is, anyway these two get cancelled. You have to just look at, okay. First step you write like this; next step onwards you do not write, so that you do not get confused. It may so happen that you may use a normalized normal distribution then you should not get confused. Now what is this now? X 1. I can only say delta X, right because X 1 minus

mu, does it give arbitrary answers? No, it would not give arbitrary answers. What is it giving? Which is out of range?

How much are you getting? Has anybody finished?

Student: 28.24 because this 0.001 is going out of.

It is X 1 is 28 point?

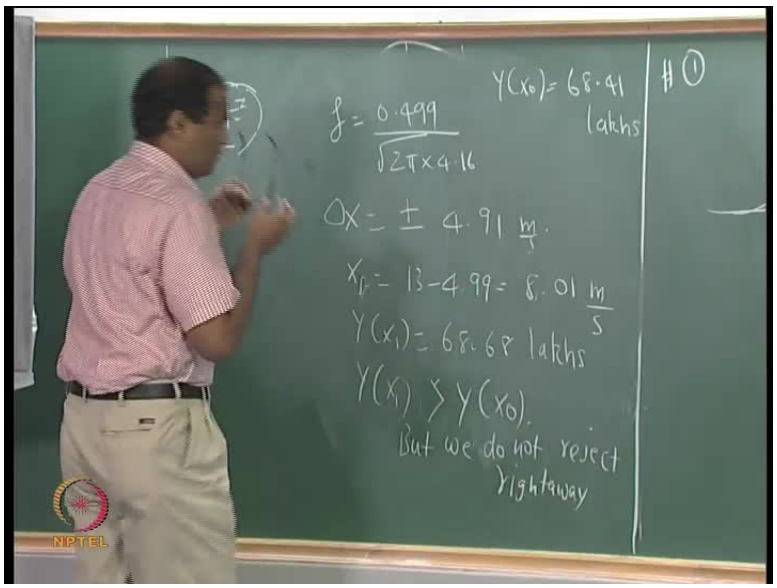
Student: 24.

I am sorry, out of range. Choose second random number.

Student: 38.91.

X, anyway it is going in the wrong direction. Now next random number f, what is f?

(Refer Slide Time: 46:41)



Student: 0.499.

Initially, yeah yeah okay, what Varun is saying is correct; that gets cancelled, it is okay. Now I will do some little bit of cheating, I mean, first we got 28.24 by adding to the positive side of 13. Now I am allowed to do the negative side. But this if you do not believe me, go to row 5 and take another random number. What is the first random number in row 5 column 5?

Student: 0.05, sir.

0.05. Then, if it less than 0.5 I will add. What is the second number?

Student: 0.182.

Okay, we will start from second because I know the answer man 10.4 meter per second, are you getting the point? But you should not do all this in the exam. So, go strictly by the random number. See I told you whether delta X will become positive or negative, this is a square. So, whether delta X is positive or negative, you have to decide by another random number; you set up an algorithm based on column 5. If random number generate a random number, less than 0.5 you go the left side; greater than 0.5 go to the right side whichever way. Now I will say delta X is how much? How much is it Sampath?

Student: 4.9.

4.91. So, let me say X 2, X 1, 8 point? So, the first step, what is Y of X 1? What was Y of X naught? I think so far you did not calculate Y of X naught. What was Y of X naught?

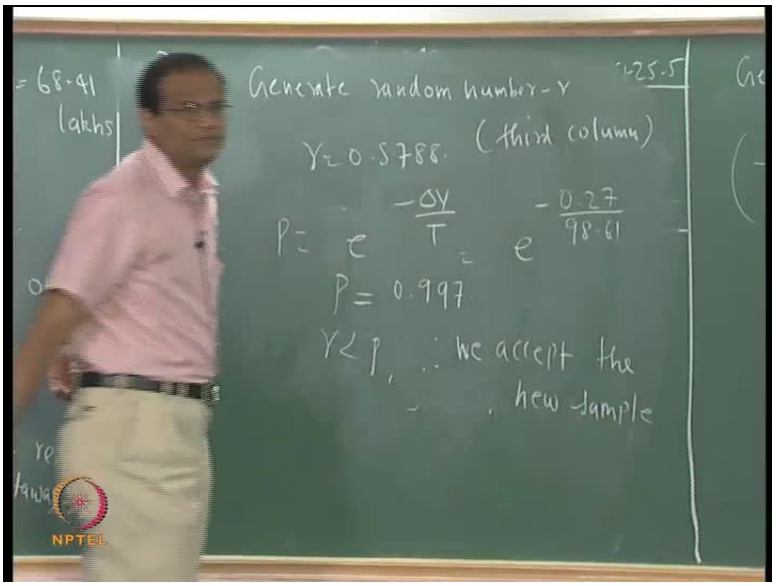
Student: This is 68.68, sir.

This is lakhs. What about Y of X naught? No, no, no, no Y of X naught?

Student: 68.41.

At 13, now it is good. I wanted to tell you the algorithm in one iteration itself we got. Now Y of X 1 is worse compared to Y of X naught, we are seeking a minimization, correct, but I do not want to reject it right away. I want to use the Boltzmann distribution, Y of X 1 greater than Y of X naught, but we do not reject right away.

(Refer Slide Time: 50:09)



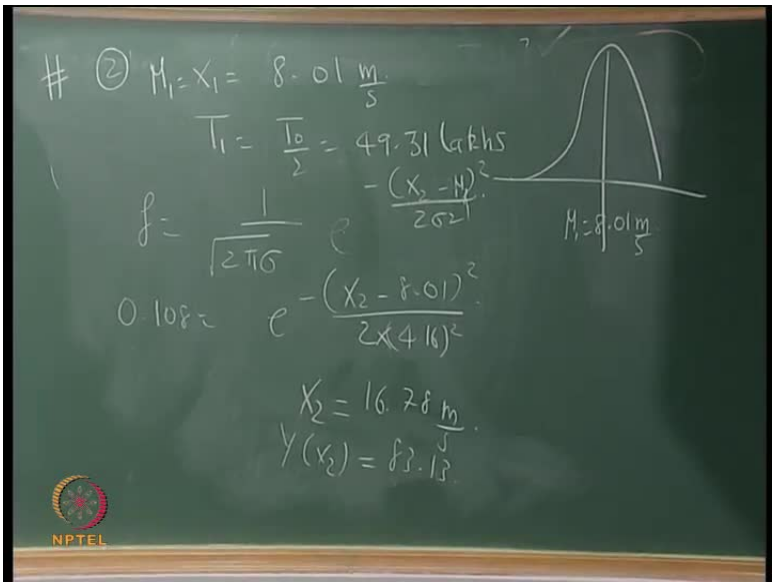
So, generate random number r third column. So third column, what is the first one, 57, r equal to 0.5788. Now P, delta is how much; 0.27, 98.61?

Student: 98.61.

Okay, I am very happy, it is very close to 1. So, if r is less than equal to P we accept, right. What did you say? r is less than P, therefore we accept this new sample even though it went down. That is the simulated annealing. It may look counter-intuitive, but in the long run it works. Because now next iteration I am going to say T 1 equal to T naught by 2, okay. We will do one more iteration and then we can close. You can see that it can be eminently programmed, easy to write an SA program.

People who want to do their B. Tech project, M. Tech project, dual degree project, whatever or you are interested in optimization, you can use Mat lab, you can code GA, you can even use code SA, take any of the problems which we discussed in this class; quiz too we had a good problem Lagrange multiplier d to the power of 1.5, you can code it and then use various strategies; that is a good learning experience. Ideally I think this course we should have a lab, right. Make it a four credit course and you should have a lab and all of you will work on a system, we will do the iterations, we will plot contours of; it will be real fun optimization lab. Maybe in the future I should think about it.

(Refer Slide Time: 52:53)



So, iteration number 2, okay. X 1 is, T 1; that is lakhs of rupees, right. T is lakhs of rupees where delta Y T is dimensionless. There is no e to the power of minus rupees is not there. Do not worry how e can be raised to the minus of rupees. That is dimensionless. Now what should you do? f equal to, we will retain the same sigma, right, or you want to change? Now what will be the mu?

Student: 5.01.

Do not get stuck with the old mu, this is a mu 1, are you getting the point? But then the 3 sigma that (( )) will be there know. There is a possibility that 3 sigma may go out; sigma will also be changed but this simulated annealing one on one. I mean so we are just trying to learn simulated annealing. In fact the actual Metropolis-Hastings algorithm, the sigma will change with respect to the current mean but let us not complicate the algorithm, we will keep the sigma same. If you keep the sigma same and it so happens that you generate a random number you get a new estimate which exceeds 25.5, throw it out and take the next random number. It is not very complicated in the exam. Only thing it will be lengthy; that is all. So, you should allot sufficient time for this simulated annealing. I mean, no traps, I mean, it will not let you down, I mean, the considerable labor is involved.

In the actual Metropolis-Hastings algorithm, what happens is sigma is 5 percent of the current mean. Watch carefully. When we do research, when we actually apply the

Metropolis-Hastings algorithm for research problems, what we do is sigma is dynamically updated to 5 percent of the current mean; that means the sigma will be 5 percent of this, will be 0.4. So, it will be 7.6 to 8.4. But if you do that we will never reach the solution for this problem. So, we keep the sigma that way. Now that is required for the research problem because such a high value of sigma the solution will oscillate when you are looking at high dimension problems. The high dimension problems are problems with lot of variables, are you getting the point? Let us keep the sigma the same. Let us keep mu 1 as this. So, what is the next random number? You are already going to the third random number, correct, 0.108. Please note we are using this sequence for f, this sequence for r, this sequence for deciding whether delta x is positive or negative. Officially, that is our stand; Ashutosh is it clear?

Student: What difference does it make which one you use?

No, no, because that f has to be a particular random number.

Student: Random numbers are random numbers; it does not make any difference with it.

No, no, no, when you start with this you have to proceed. In this case you have to start and sometimes you will get 0.5, 0.5, 0.5 always, 0.6, 0.6. They are all in a particular sequence; when you are doing iterations you have to follow it. This is the way the computer will generate. In a do loop, if you put rand of x it will generate in this order. Otherwise, you will pick and choose all those things which are more than 0.5, are you getting that point? That sequence of random number is very important. So, what is the new random number? 0.108. We will leave this. And tell me whether it is going to be positive or negative? What are you getting Varun?

Student: Sir it needs to be positive.

Why?

Maybe you are getting high value now?

Student: Yeah, that is correct sir, 8.75.

Which one?

Student: Delta x.

Delta x is 8.7. X 2 is? You add plus 8.7.

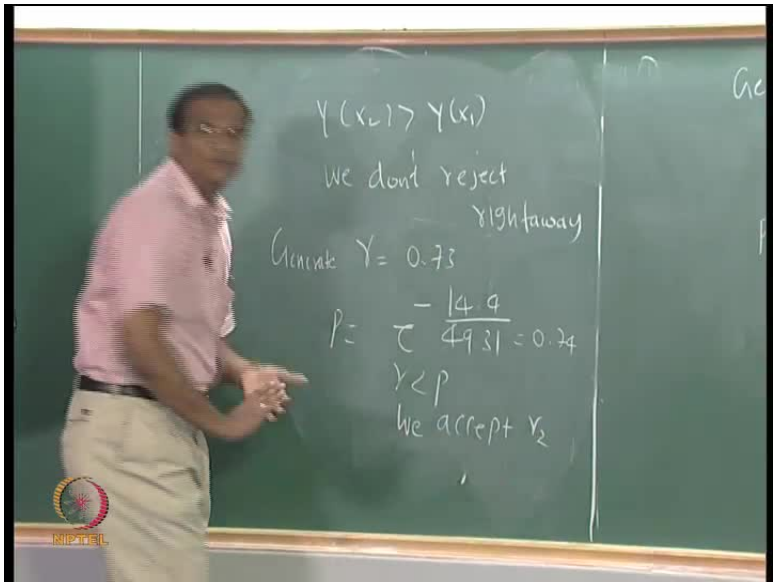
Student: 16.78.

Now we will apply the MH once more; So, Y of X 2?

Student: 83.13.

83.13. Now, is it getting worse?

(Refer Slide Time: 58:50)



Correct, we do not reject right away. r, what is the new r?

Student: 0.45.

0.45.

Student: 0.73.

0.73, very good; you are giving me some hopes. e to the power of minus?

Student: Rejected values.

Rejected, I am so happy. I can stop it, e to the power of minus?

Student: 14.4.

14.4.

Student: 49.31.

49.31. What is this?

Student: 0.74.

0.74, then how can you reject? Accept?

Student: It is close, sir. You can reject sir.

You have to accept. But it is going in the wrong direction but you have to accept. But you see it is getting stricter because the denominator is going down. Abhishek is that clear? Senthil are you able to see? The denominator is going down; therefore, it will become tough to accept it. Okay, this time r is less than P, we accept r 2. So, you can complete the other two iterations at home. Anyway, there is no funda involved; you know how to accept or reject. This is how the simulated annealing works. After some time it will cover the zigzag path.

So, it will cover the whole of the solution space fairly well, so that global optima are not missed. It is a pretty powerful technique. So, if the objective function is computationally expensive, for example, you want to solve a tumor. You want to solve for a tumor using the peens by e transfer equation or you are doing FEM, ANSYS or CFX or FLUENT to generate all this and for getting each value of Y, it is going to take a lot of time. I keep telling you, you can develop a neural network. You can run for some so many combinations of X, validate it, train it, such that neural network is just like regression; just like Y equal to a x squared plus b x plus c. It is gives you a regression of the independent variable. And then, it gives y. And then keep playing with that y using simulated annealing and then finally see and then finally you will get the optimum.

Now what you will have to do is after you get the optimum, you can substitute those values of x back into your original forward model. Which is the forward model? That is you want to find out the temperature distribution in a tumor or something. After you get all these parameters, you can substitute it into your original governing equation and generate all the temperatures and check whether temperatures which are predicted by the neural network are the same as the temperatures you are getting by the full model. This

completes the loop. This is a standard operating procedure for doing research. Or one step further, you do this experiment, have a heat source, change the volumetric heat generation rate, put thermocouples, measure it and then use the experimental data. Then you can go to the highest, you can go to a very high journey.

And, if you can come up with some algorithm which is more powerful, you can prove that for the Himmelblau function or Banana function, it is superior to what Kirkpatrick has done, you can aim at Nature or Science. So, it is possible and they did not come from heaven the people who are publishing there. You have to put effort that is all. Normally engineers do not try to work in those, try to publish in those kinds of journals but it is possible. Once you have Nature and Science, the advantage is you have lot of citations, lot of people will look at you work and all that. Suddenly, overnight you will have greatness thrust upon you. Then how to handle this will be another problem, okay.

Now, I will summarize the whole course.

That is by cooling rate algorithm. I am suggesting that T can be reduced by half. You can have different rates. So, ultimately what is the best reduction in cooling rate that you have to decide based on your problem. You do not want it to reduce it by 4 times or 8 times or 10 times, because the rejection will become very this thing. Rejection will become very strict. It will accelerate your convergence but premature convergence. So, between the devil and deep sea, you have to.

That is for starting. Only for starting, you have got the average. Then because initially what is that T, you did not know. So, you took for average of 4 values and, okay

Student: Sir mu value, what is to compare?

That is a mu value.

Student: That will become the new.

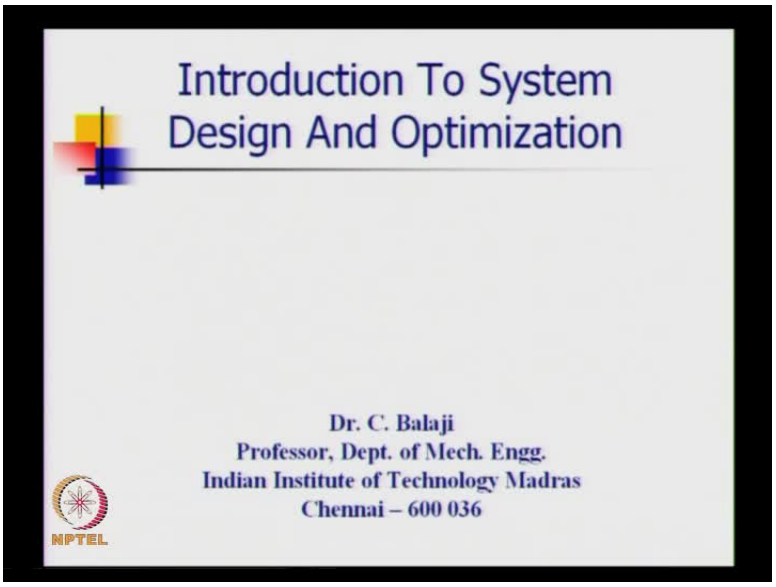
Which one?

Student: The new value which is there.

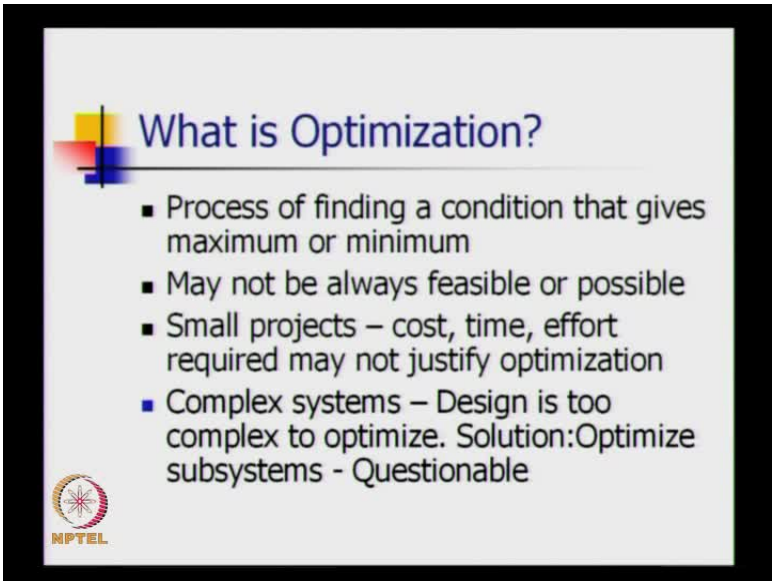
I am putting my distribution around the new mean. The new value of X becomes the mean of the distribution. That is the way all sampling algorithms will work. When you

are proceeding from 13, the mean is around 13. From 13 if you come to 8, the mean is around 8. From 8 if you come to 8.6, it is, the mean is around 8.6, are you getting the point? The new value of X becomes the value of mu automatically, are you getting the point?

(Refer Slide Time: 1:05:02)



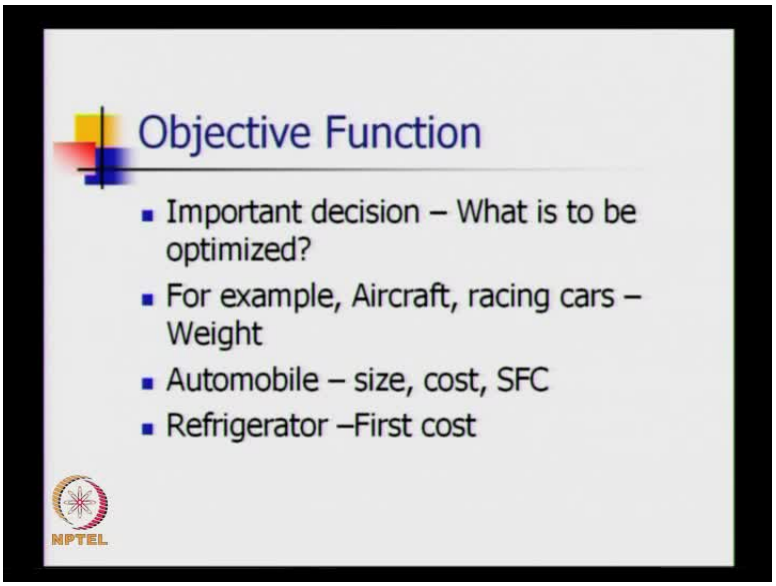
(Refer Slide Time: 01:05:10)



So, it is just a quick summary of the whole course. Yes, next slide. Oh, what is this? Now it is a dumb question what is optimization after going through 43 lectures. Keep on pressing. Process of finding a condition that gives the maximum or minimum; I told you

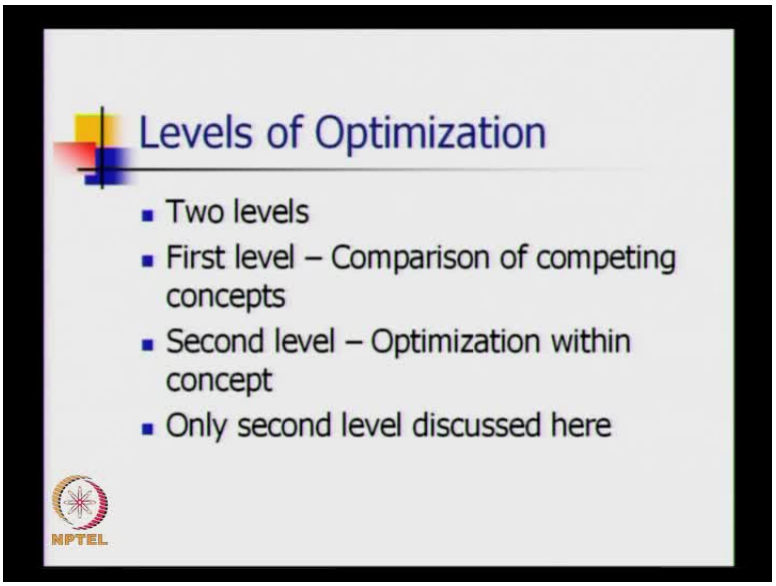
it may not always be feasible or possible because of the complexities, time and money involved. Small projects that cost time and effort may not justify. Complex system design is too complex. One possible strategy is to subdivide the problem into optimization of subsystems and proceed.

(Refer Slide Time: 1:05:38)



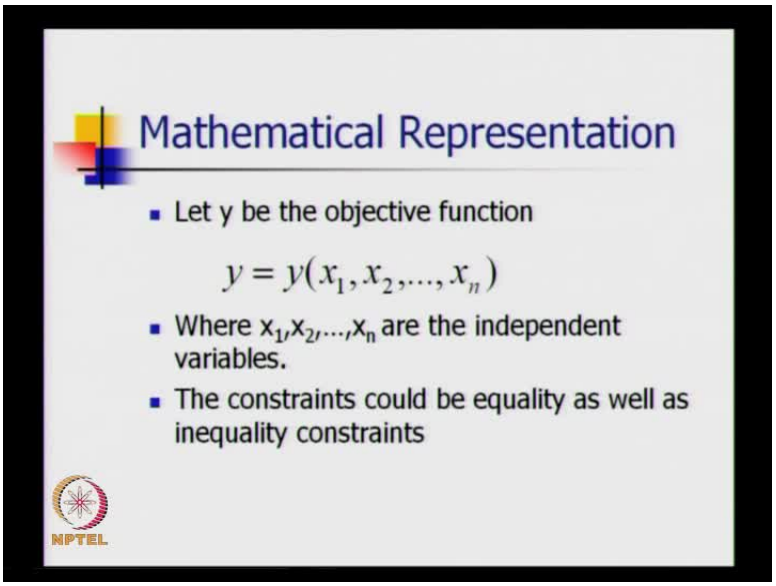
The important decision is what is to be optimized; that is called the Objective Function. For example, aircraft, racing cars, it will be the weight; for automobile, it could be size, cost and specific fuel consumption or it could be for a racing car, it will be BHP per ton or weight; for the refrigerator, it is the first cost. What is the first cost when you buy it in the market? For the air conditioner, more important is it will be the running cost.

(Refer Slide Time: 1:06:01)



Several levels of optimization are there. For example, if you want to look at optimization from two levels, in one level it is a comparison of competing concepts which we have not considered in this course; it is not possible to consider. That is there is a, are you getting the point? You want to solve the power problem in Tamil Nadu. There are so many ways; you can import power, you can have a coal-based power plant, you can have a gasification plant, you have a nuclear power plant, then you have to optimize each of this and find out; that is a big task. But once you decide that I want to have a nuclear power plant, then we can give for a nuclear power plant near Chennai what will be the optimum. So, comparison of competing concepts is more difficult. Second level is optimization within a concept. Only second level was discussed in this course.

(Refer Slide Time: 1:06:41)



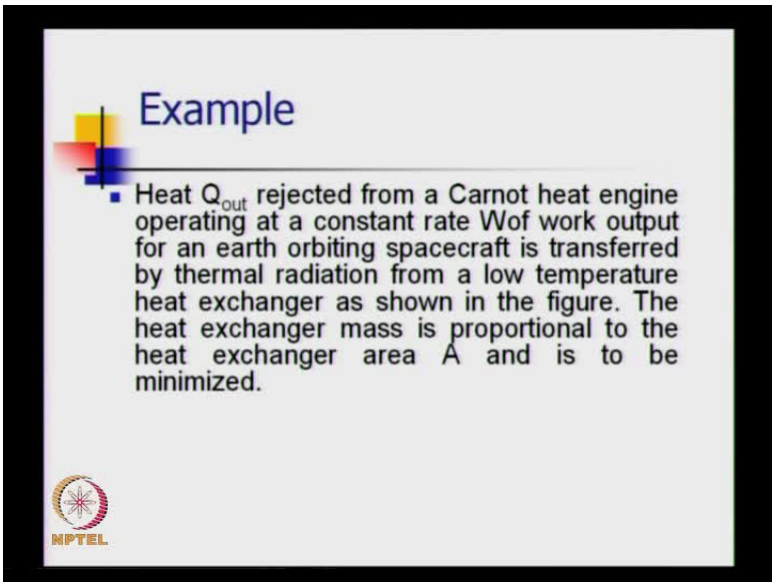
Mathematical representation, y is written as y of x 1 to x n; x 1 to x n are independent variables and usually it will be constraints could be equality as well as inequality constraints. Economics is all about constraints; the constraints actually bind the solution. The economics is all about, economics in one line?

Student: Unlimited wants and limited resources.

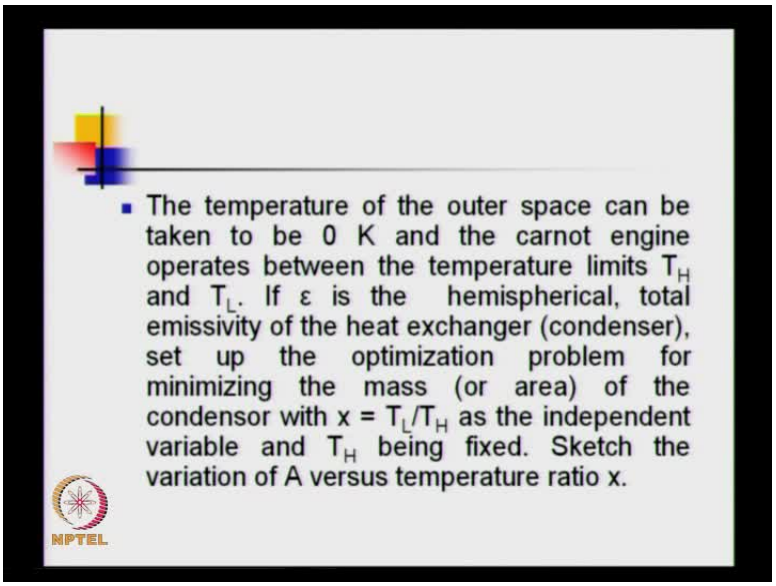
Unlimited wants and limited resources. That is, the whole point is unlimited wants we have, but we have only limited resources. The resource could be anything; it could be time, it could be money, whatever, right. So, the equality constraints are given as phi 1 or phi 1 to phi m and the inequality constraints are psi i to psi j less than equal to l a; that is only representation. We can also have greater than equal to. Inequality constraints are more difficult to handle, okay.

The following relations hold. Minimum of A plus y is A plus minimum of y; max of y is min of minus y, okay.

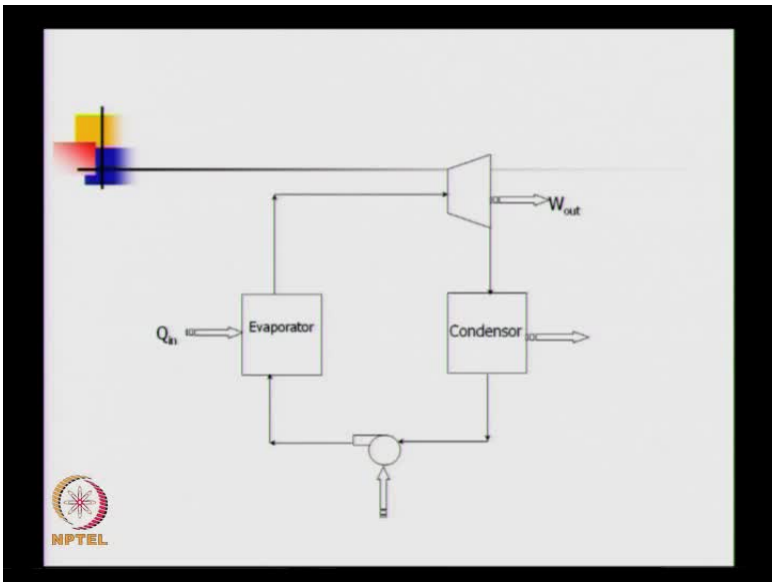
(Refer Slide Time: 1:07:42)



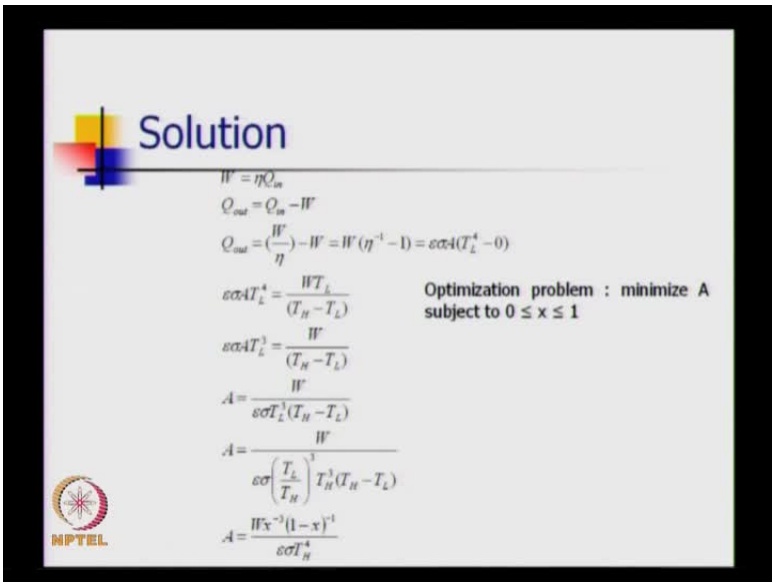
(Refer Slide Time: 1:07:45)



(Refer Slide Time: 1:07:46)



(Refer Slide Time: 1:07:53)



So, we have seen this problem heat rejected from a Carnot cycle. Go to the next one, next one. So, evaporator condenser is operating in outer space we solved this problem. So, only radiation is possible. We set up the optimization problem to minimize A; A was the area and we wrote it in terms of the temperature ratio. Temperature ratio is T L by T H. We optimized it straight using calculus. We did this?

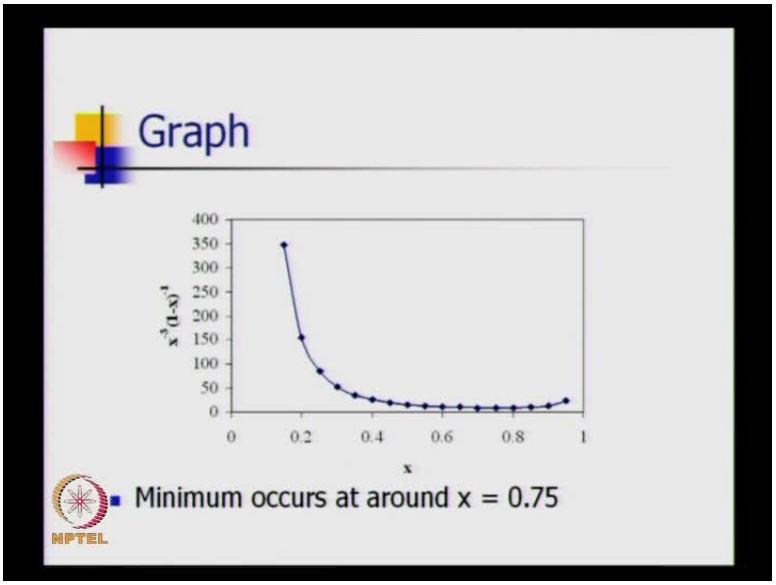
Student: No sir.

No, we did not do this? I will put it on Moodle. So, this is simple. Do not worry, you have solved complicated problems. Okay then you want, you have time? I will show. It is okay, you are getting very conscious. Now you know how difficult it is.

So, the work output is efficiency into the Q in. From thermodynamics you know that first law of thermodynamics Q out is Q in minus Q W. So, I am writing Q out in terms of W and eta where eta is efficiency. I am keeping W fixed. So, whatever heat is rejected is also the heat which is rejected from the condenser, right, in a power plant. The heat rejected from a condenser of a power plant operating in outer space is basically heat rejected by radiation alone. That is that can be given by the Stefan-Boltzmann constant. We assume the outer temperature to be 0. You have epsilon sigma A T L, temperature of the condenser to the power of 4 minus 0, okay.

Now I have written epsilon sigma A T to the power of 4 is W into eta for a Carnot engine can be written as 1 minus T 2 by T 1. T 2 is T low, T L, T 1 is T H. I write it in terms of T L and T H. Now I am just doing some mathematical manipulation and I am writing out an expression for the area A. I want to minimize the area A. A is W, W is fixed, emissivity is fixed, Boltzmann constant is fixed, T L is fixed. I want to find out what is the ratio of T L by T H; that I call as A. So, I am posing this optimization problem in terms of X; X is the temperature ratio. What is the temperature ratio?

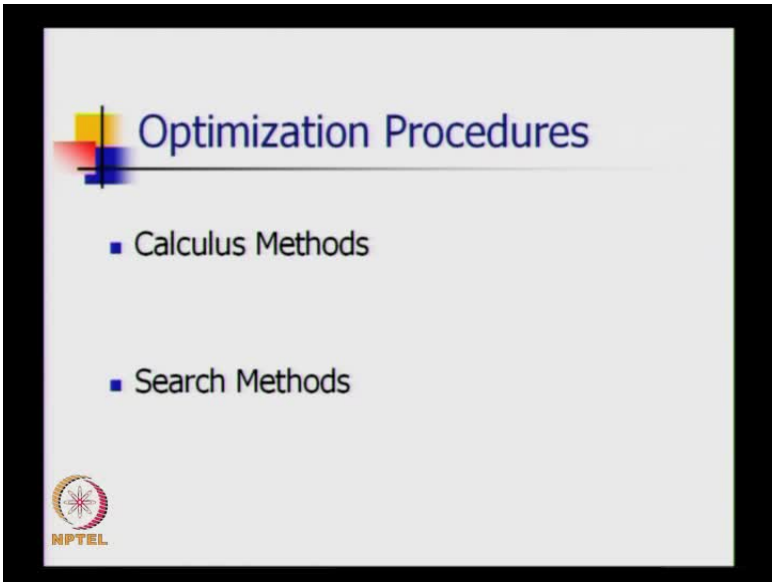
(Refer Slide Time: 1:09:50)



Now you can solve it by calculus. If you solve it you are getting the minimum occurs around 0.75; that is T L by T H is 0.75. This is a simple calculus based approach for a one variable problem.

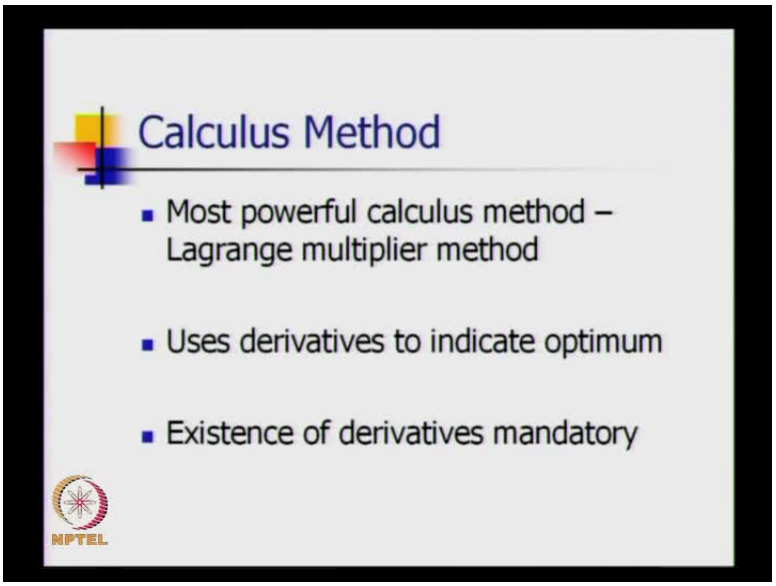
This is just to give you an idea of how to formulate an optimization problem; from English, you apply the laws of physics and convert it into mathematical form. After you convert it into mathematical form, you have to decide an appropriate strategy for solving the optimization problem. After you get the results you have to do post processing. You should be in a position to analyze; you should be in a position to analyze the results and then, in Lagrange multiplier and all that it is possible to do post optimality analysis sensitivity coefficient.

(Refer Slide Time: 1:10:30)



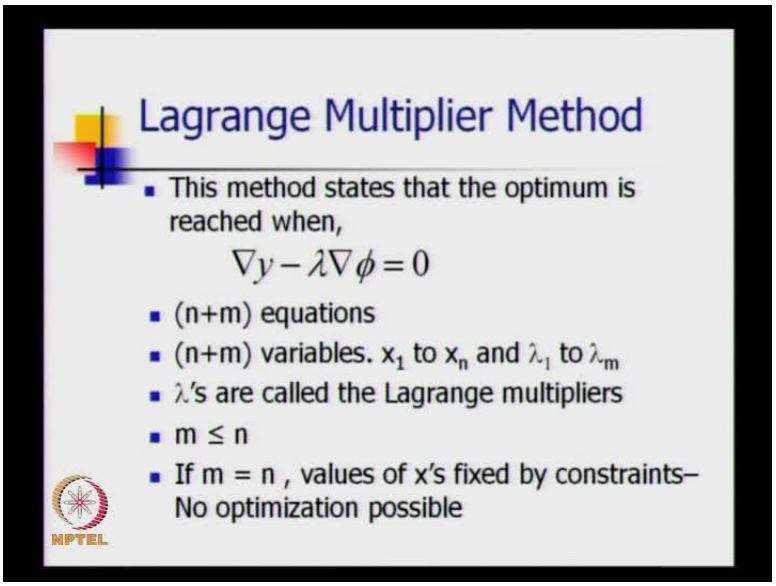
So optimization procedures, calculus methods and search methods, two broad categories.

(Refer Slide Time: 1:10:35)



Most powerful is the Lagrange multiplier method; uses derivatives to indicate optimum. So, the existence of derivatives is mandatory.

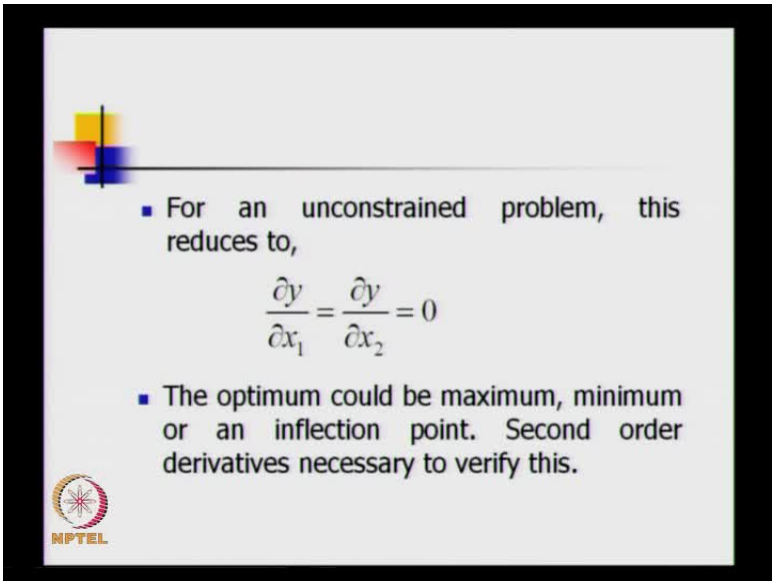
(Refer Slide Time: 1:10:41)



So, this method states that the optimum is reached when del y is minus lambda del phi equal to 0. Graphically we saw how this works. Here itself we can see that lambda equal to del y divided by del phi; lambda is the change in the objective function with the change in the constraint. So, it is a sensitivity coefficient. It is also the shadow prize, right. So, there are m constraints. So, m constraint equations, n variables, so n plus m

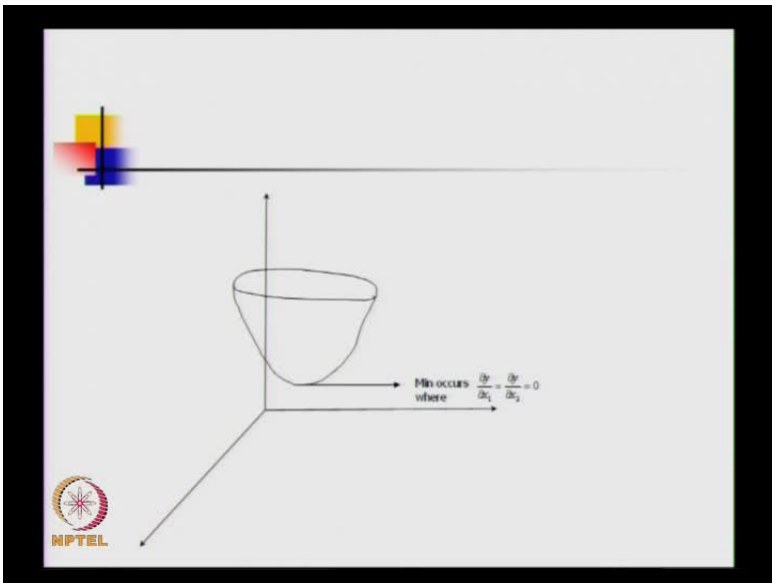
equations, n plus m variables, lambdas are called the Lagrange multipliers. So, m must be less than equal to n. So, that is the problem with equality constraints. If m equal to the n directly the constraints can be solved. That is the solution whether you like it or not.

(Refer Slide Time: 1:11:22)



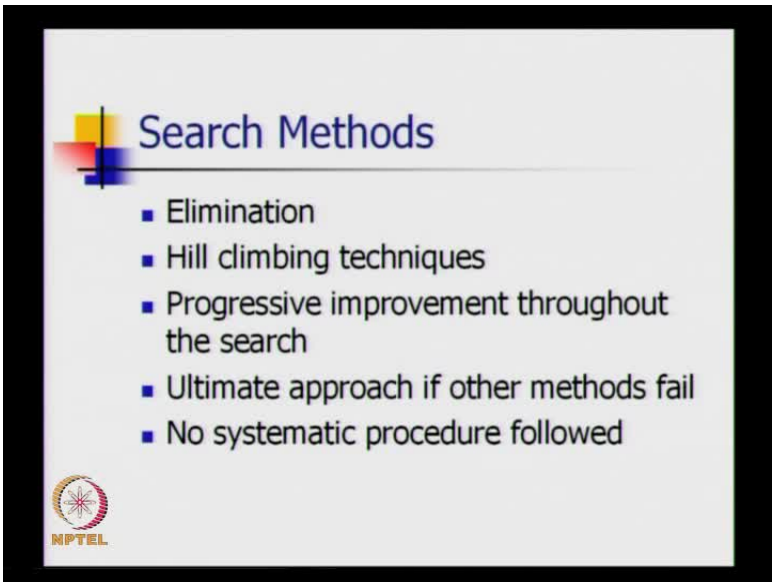
For an unconstrained problem this reduces to del y by del x 1 is equal to del y by del x 2 and so on. Second order derivatives are necessary to verify this. We use the Hessian and evaluate it, okay.

(Refer Slide Time: 1:11:32)



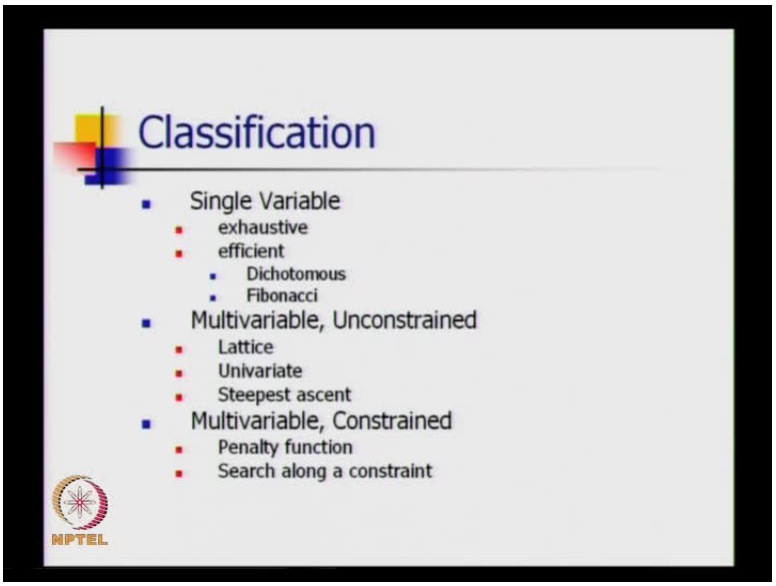
So, it is a depiction of a two variable problem where the minimum occurs, dou y by dou x 1 equal to dou y by dou x 2 equal to 0.

(Refer Slide Time: 1:11:39)



Search methods are based on eliminating a portion of the interval or on systematically climbing to the top. So, you have elimination or hill climbing techniques. In both these techniques there is a progressive improvement of y. y keeps on increasing or decreasing, it is depending upon whether you are solving a maximization or minimization problem. It is the ultimate approach if other methods fail, right. But sometimes there is no systematic procedure which is followed, so you may feel that it is a Helter-Skelter method of searching. But generally in many of the methods even though it is mad as they say there is a method in the madness, okay.

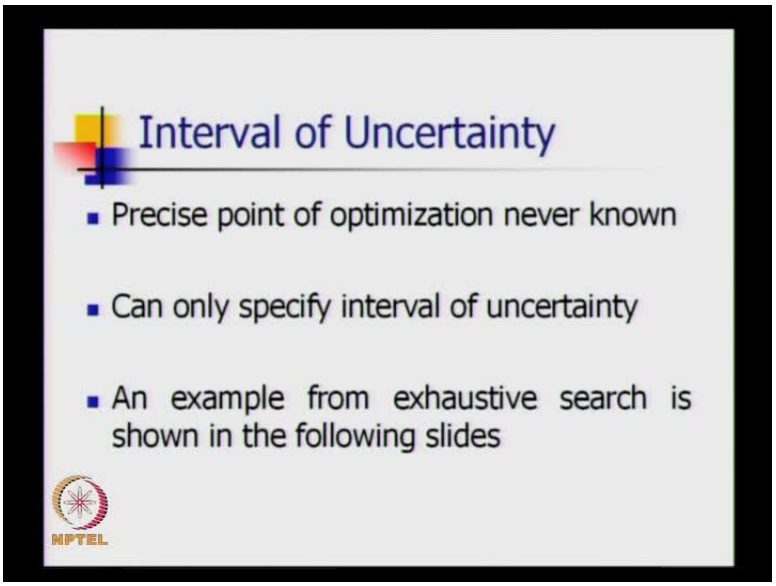
(Refer Slide Time: 1:12:14)



So, single variable, multivariable; broadly optimization problems can be classified as single and multivariable, constrained and unconstrained. So, we saw the exhaustive search, the efficient search, dichotomous and Fibonacci search. We also saw the Golden section search, multivariable unconstrained I told you the lattice method; east, west, north, south, northeast, northwest, southeast, southwest and univariate search method is converting it to one variable problem and solving one variable at a time, then steepest ascent or steepest descent. Please remember lattice univariate and steepest ascent can be applied only to unconstrained optimization problems, are you getting the point?

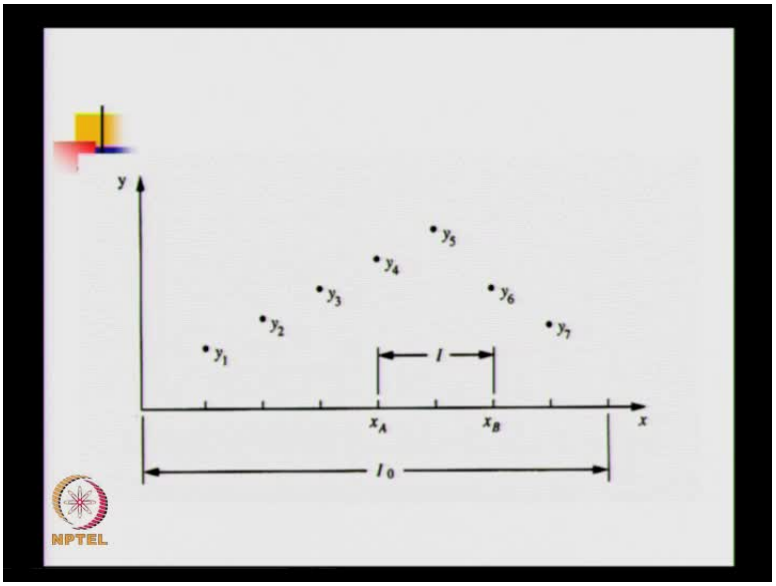
So, we have delta x 1 divided by dou y by dou x 1 is equal to delta x 2; that is the steepest ascent. You fix delta x 1 and then get all the other delta x's or choose in terms of alpha and decide how much you will go. That is there are two strategies for that. Multivariable constraint we did not look at many techniques, but I told you how to convert it into an equivalent unconstrained problem by putting a penalty on violation of the constraint and the penalties there will be a square of the constrained term, so that it is always positive. For a minimization problem the penalty will be plus. For a maximization problem the penalty will be minus. For a minimization the penalty will be addition of cost. For maximization it is a reduction in profit. Though you would love to put a minus sign for the minimization problem, it is counter-intuitive, right. I have explained this to you several times.

(Refer Slide Time: 1:13:37)



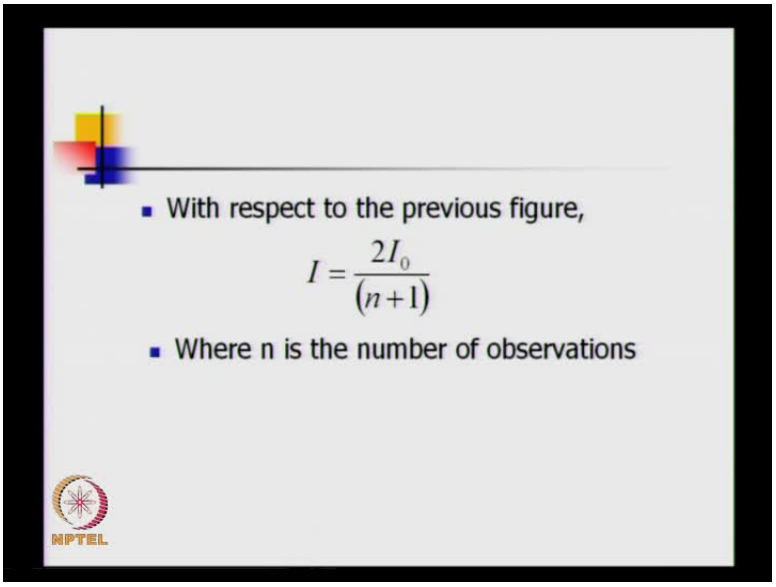
So, the most important funda in any search method is the interval of uncertainty. Your final answer lies between which two limits; that is the interval of uncertainty. So, the precise point of optimization is never known because you do not solve using calculus. So, you can only specify the interval of uncertainty. The interval of uncertainty should keep on reducing and the original interval of uncertainty divided by your new interval of uncertainty gives the RR or the reduction ratio of the algorithm.

(Refer Slide Time: 1:14:04)



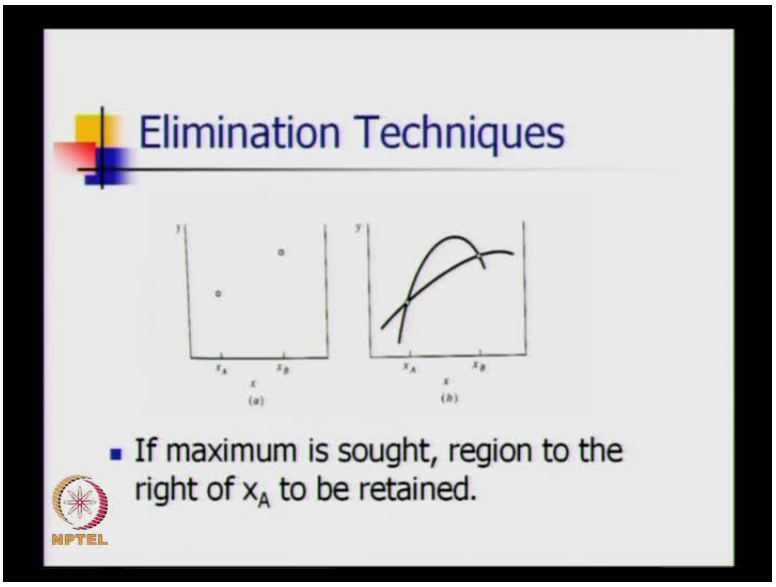
So, this is basically a very simple depiction of how to use the two point method. Basically, you can see that the function is increasing; it is monotonic. So, I can only say that the optimum lies somewhere between y 4 and y 6. I cannot say that it is left of y 5 or y 5, x 5 or the right of x 5 but I am sure that it is lying between A and B or x 4 and x 6, okay.

(Refer Slide Time: 1:14:30)



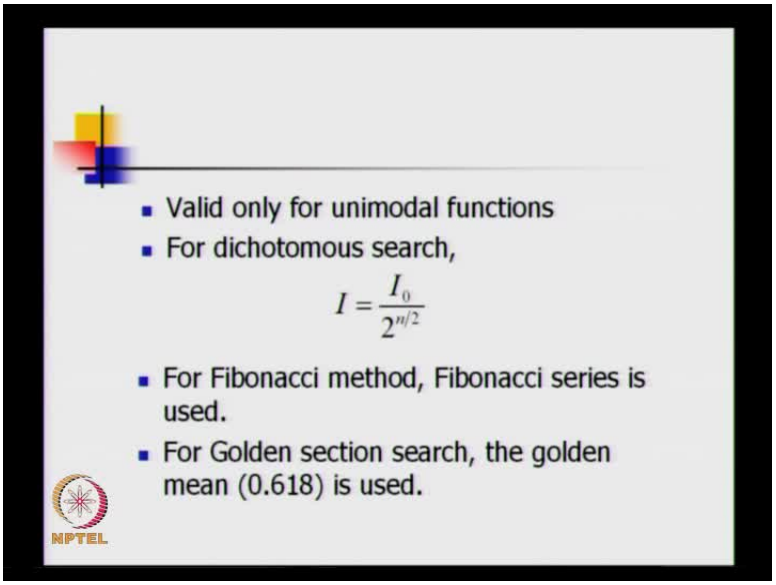
So, with respect to the figure that was an exhaustive search equal interval 2 I naught by n plus 1 or I naught by n plus 1 by 2. So, n plus 1 by 2 is the RR of this algorithm, okay.

(Refer Slide Time: 1:14:43)



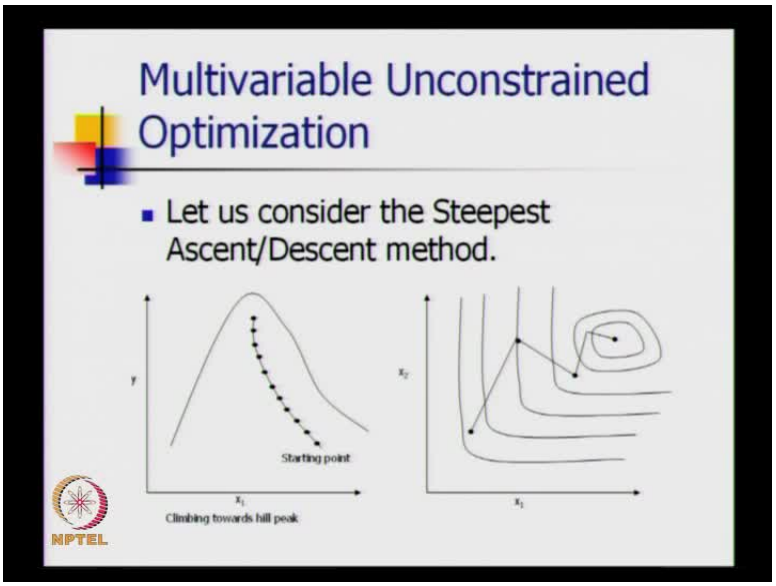
So, this is basically if it goes like this, if the function goes like this, if maximum is sought. So, since y of x b is greater than y of x a, region to the right of x A has to be retained.

(Refer Slide Time: 1:14:57)



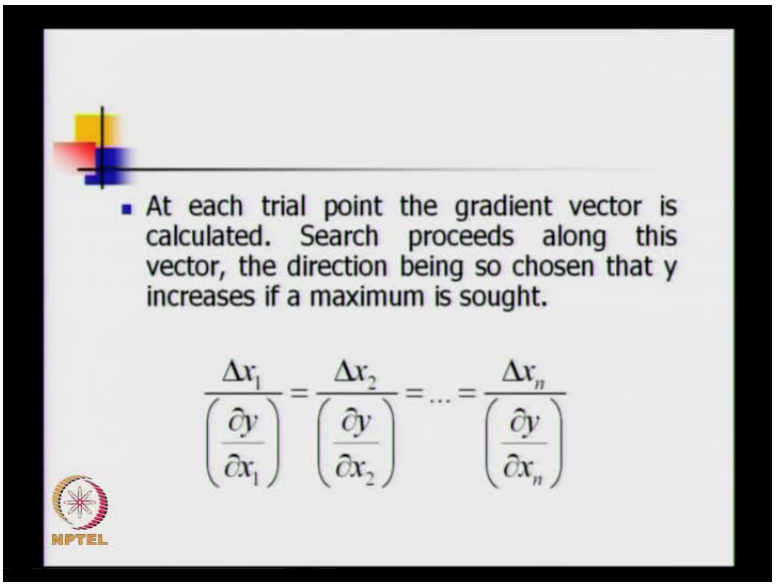
Valid only for unimodal functions; for dichotomous search it goes as RR goes as 2 power n by 2; for Fibonacci method Fibonacci series is used; for Golden section search the golden mean is used, 0.618 or 1.618. So, the ratio of the consecutive numbers in the Fibonacci search also approaches 0.618.

(Refer Slide Time: 1:15:16)



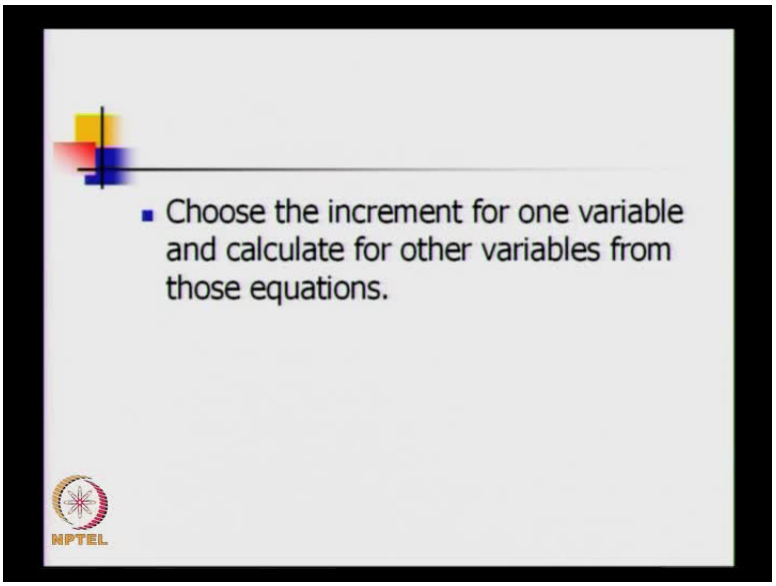
Let us consider the steepest multivariable unconstrained. Let us consider the steepest ascent or steepest descent. You start for a single variable problem; you start with a particular point and systematically reach the top. For a two variable problem you go in the steepest descent direction or ascent direction and reach the top.

(Refer Slide Time: 1:15:33)



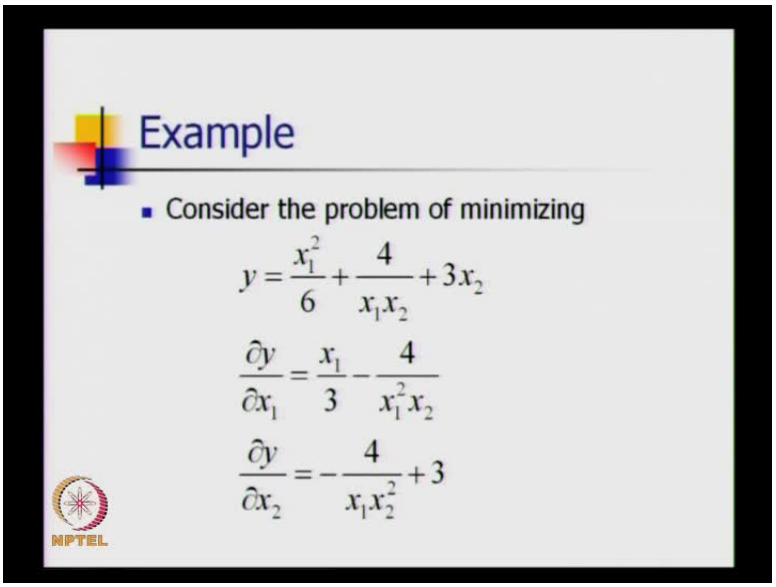
At each trial point the gradient vector is calculated. The search proceeds along this vector. The direction is chosen, so that y increases if the maximum is sought, are you getting the point? Whether it is going to be positive delta x or negative delta x depends upon whether d y by d x is positive or negative and you want a minimum or maximum, that condition was there. So, this is the condition. So, if you choose delta x 1, all the other delta x's can be obtained or simultaneously you can choose by defining an alpha and solving for alpha each time.

(Refer Slide Time: 1:16:03)



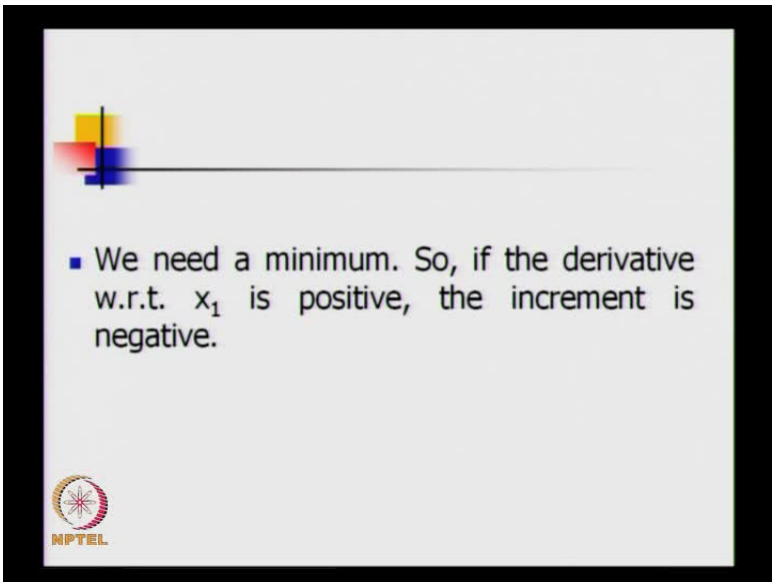
So, choose the increment for one variable and calculate for other variables.

(Refer Slide Time: 1:16:07)



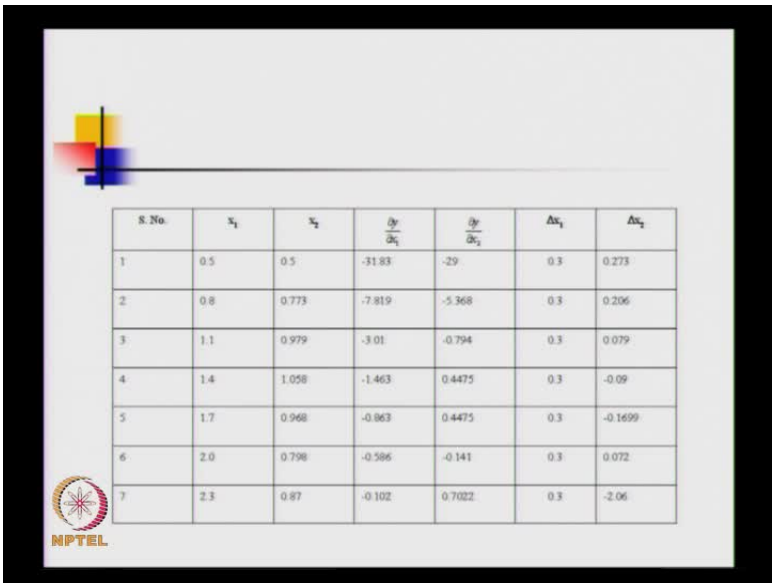
So, consider the problem of minimizing. So, x 1 squared x is the numerator first term, x 1 x 2 is the denominator in the second term, third term. So, you can do d y by d x, d y by d x 2; yeah, next.

(Refer Slide Time: 1:16:17)



So, we need a minimum. So, please go to the next one.

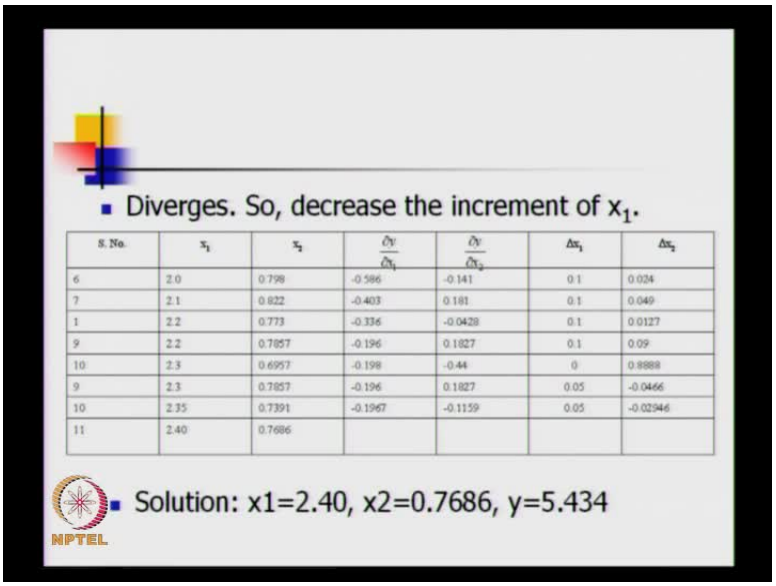
(Refer Slide Time: 1:16:21)



S. No.	x_1	x_2	$\frac{\partial y}{\partial x_1}$	$\frac{\partial y}{\partial x_2}$	Δx_1	Δx_2
1	0.5	0.5	-31.83	-29	0.3	0.273
2	0.8	0.773	-7.819	-5.368	0.3	0.206
3	1.1	0.979	-3.01	-0.794	0.3	0.079
4	1.4	1.058	-1.463	0.4475	0.3	-0.09
5	1.7	0.968	-0.863	0.4475	0.3	-0.1699
6	2.0	0.798	-0.586	-0.141	0.3	0.072
7	2.3	0.87	-0.102	0.7022	0.3	-2.06

So, you can start with 0.5. I have a delta x of 0.3. So, I think we solved this, right. So, 0.5, 0.82, it becomes like this. When suddenly the function becomes funny, then you reduce the delta x 1.

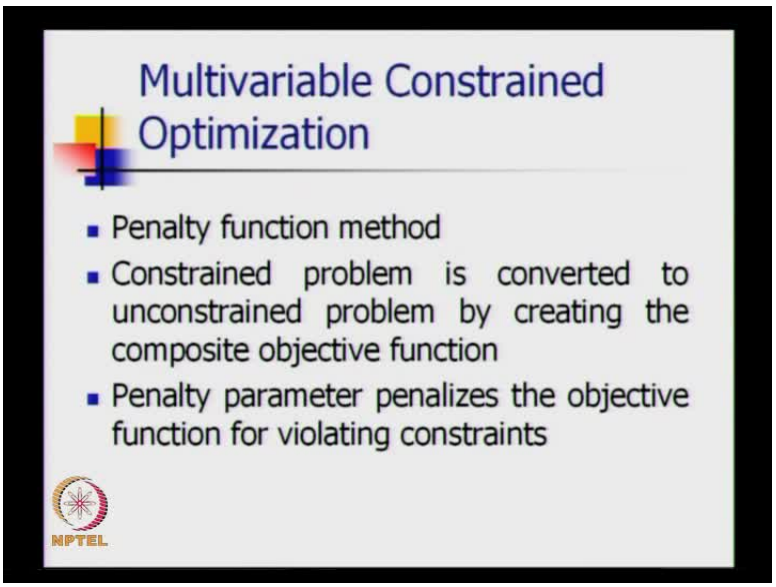
(Refer Slide Time: 1:16:34)



S. No.	x_1	x_2	$\frac{\partial y}{\partial x_1}$	$\frac{\partial y}{\partial x_2}$	Δx_1	Δx_2
6	2.0	0.798	-0.586	-0.141	0.1	0.024
7	2.1	0.822	-0.403	0.181	0.1	0.049
1	2.2	0.773	-0.336	-0.0428	0.1	0.0127
9	2.2	0.7857	-0.196	0.1827	0.1	0.09
10	2.3	0.6957	-0.198	-0.44	0	0.8888
9	2.3	0.7857	-0.196	0.1827	0.05	-0.0466
10	2.35	0.7391	-0.1967	-0.1159	0.05	-0.02946
11	2.40	0.7686				

So, this is an easy method or you have to calculate alpha. So, this is basically, finally, you can see that d y by d x 1, d y by d x 2 are not changing much. So, the solution is 2.4 and 0.7686.

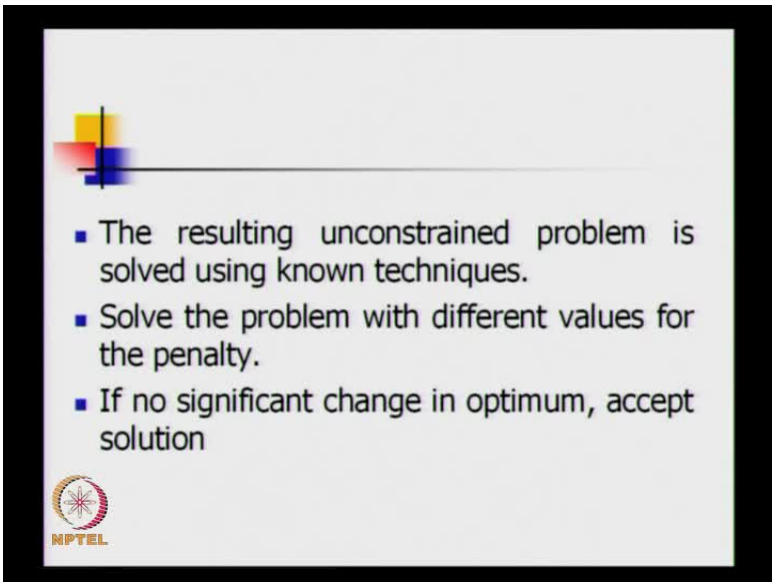
(Refer Slide Time: 1:16:48)



For multivariable constrained optimization, penalty function method is very important; it is very powerful. The constrained optimization problem is converted to unconstrained problem by creating what is called the composite objective function which takes care of

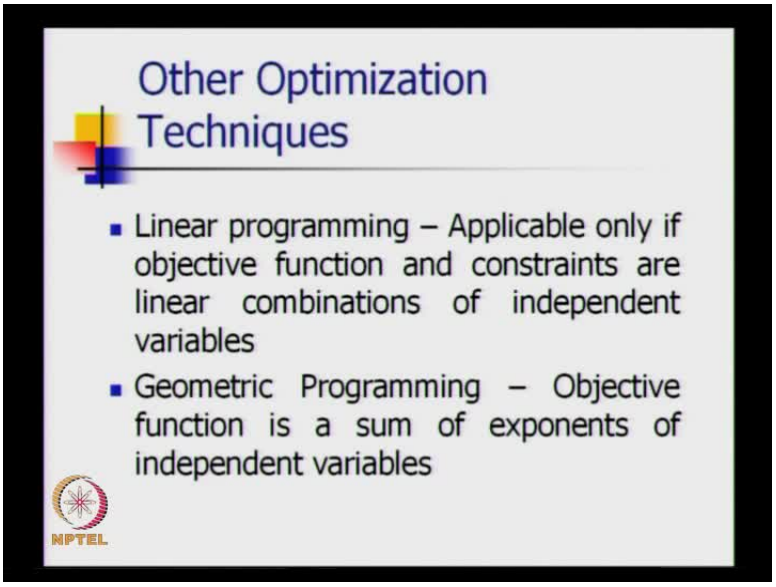
both the objective function and their constraints. The penalty parameter penalizes the objective function for violating the constraints.

(Refer Slide Time: 1:17:08)



The resulting unconstrained problem is solved using known techniques. Solve the problem with different values of penalty. If there is no significant change in optimum, then stop, okay.

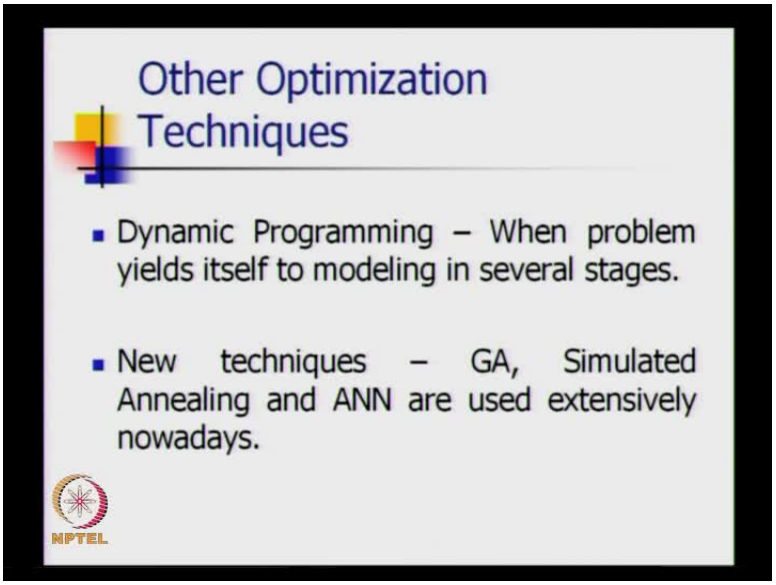
(Refer Slide Time: 1:17:18)



Other optimization Techniques, linear programming applicable only if the objective functions and constraints are linear combinations of the independent variables. So, we

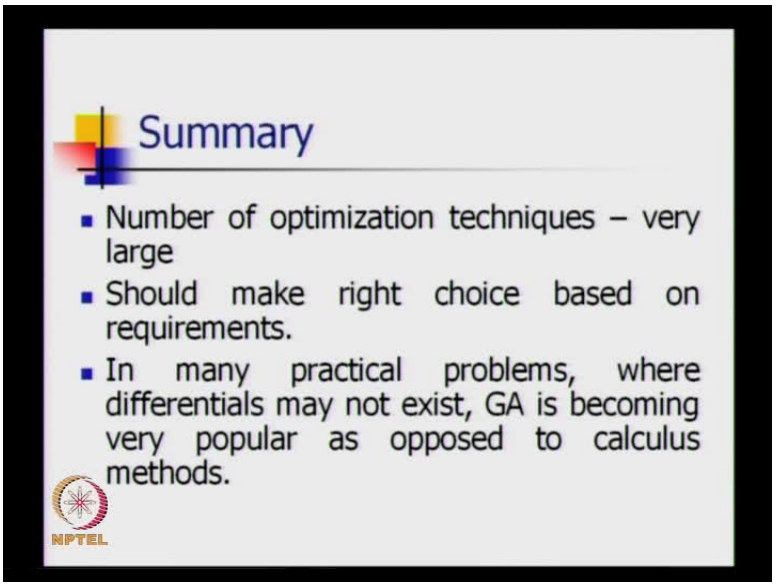
try. So, we solved the LP problems using two techniques, the graphical method for two variables and also the method of slack variables. A systematic way of performing the method of slack variables is the simplex method which some of you must have learnt in operations research. Geometric programming is basically objective function is a sum of exponents. I did not cover this, because there are in many problems in thermal sciences which are amenable to geometric programming.

(Refer Slide Time: 1:17:54)



Dynamic programming: When a whole problem can be subdivided into stages, then you try to optimize with respect to each and every stage and proceed from your starting point to the destination. New techniques: Several new, nontraditional techniques or non-classical optimization techniques like Genetic Algorithm, Simulated Annealing and also Neural Networks can be used for optimization; that means you train a neural network and exhaustively search; you trivialize the problem to a certain extent, but sometimes very complex problems cannot be handled this way. So, at least two of these techniques we have seen in this course, right.

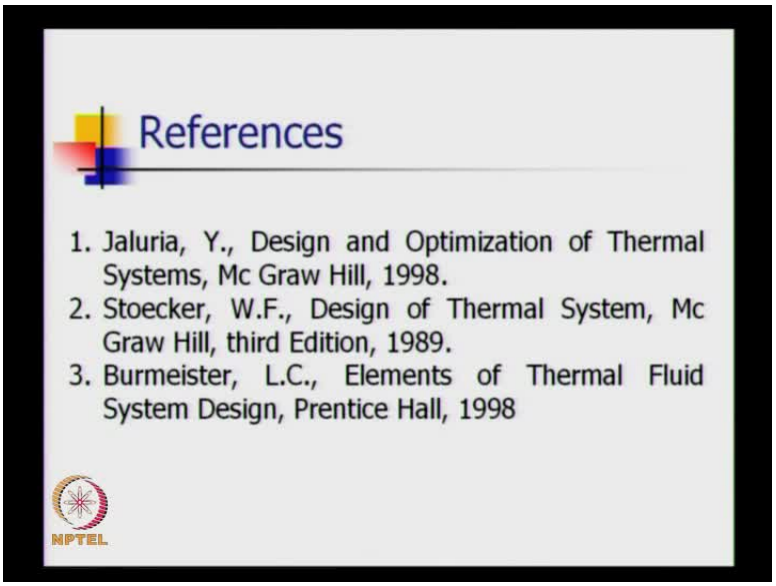
(Refer Slide Time: 1:18:26)



So, what is the summary of the summary? So, the number of optimization techniques is very large. Optimization is very powerful tool in the hands of the engineer. Regardless of the field you work, regardless whether it is electrical engineering, mechanical, chemical, whatever, always there is scope for optimization. And once you know the basic methodologies and tools, then it is a lot of fun to optimize and try to seek because we all try to seek improvement, right, in whatever we do. So, the number of optimization techniques is indeed very large.

The analyst should be able to make the right choice based on requirements. So, the idea behind this course is you have got a flavor of all these techniques. So, when you actually encounter an optimization problem, you know which is the methodology you have to choose. And then you will probably code or use a standard code and solve it. In many practical problems where differentials may not exist, so these non-classical techniques like Genetic Algorithms, Simulated Annealing, Particle Swarm and all these are gaining popularity.

(Refer Slide Time: 1:19:20)



So, these are some of the books I have used.

(Refer Slide Time: 1:19:24)

Thank you.