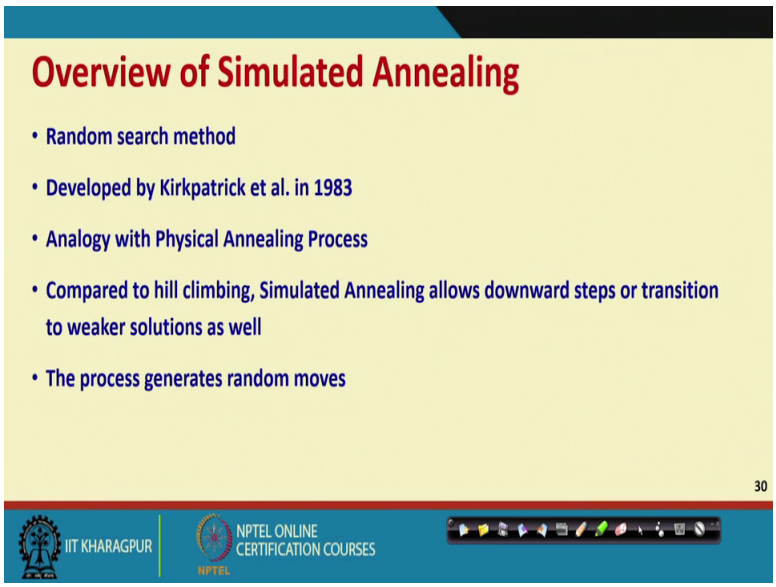
Selected Topics in Decision Modeling
Prof. Biswajit Mahanty
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

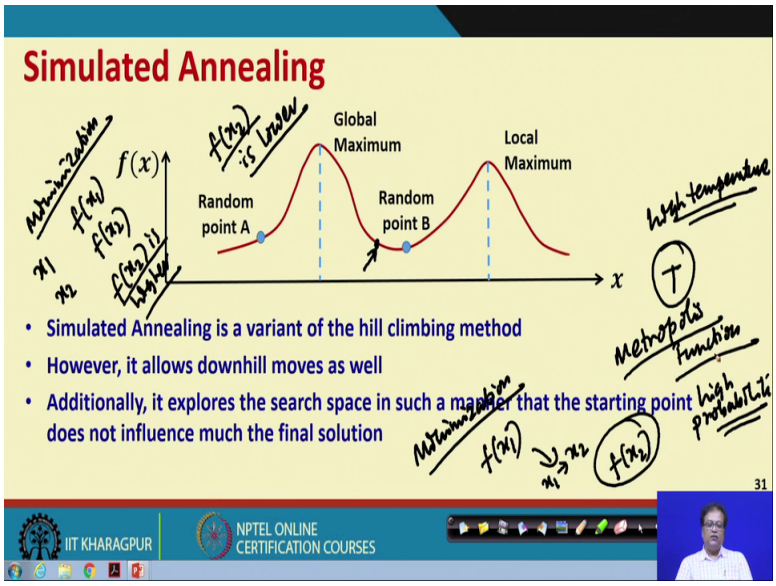
Lecture – 36
Simulated Annealing

(Refer Slide Time: 00:27)



So, in our subjects Selected Topics in Decision Modeling, we are now in 36th lecture that is simulated annealing. Simulated annealing, as you can see you know it is a random search method developed by Kirkpatrick and others in 1983 based on the physical annealing process. And highly compared to hill climbing, only thing it also allows downwards steps or transitions to weaker solutions and through a series of random moves.

(Refer Slide Time: 00:54)



Now, how it happens, look here in this case you see a function has got multiple peaks and traps. So, we have a global maximum, there could be a local maximum also. If you start the search from a random point A, obviously you move up the global maximum that is the usual hill climbing method where with, every iteration we try to go to a higher value and discard any value that goes lower. And but then in hill climbing method, typical hill climbing method, if you start from random point B, then through hill climbing you do not lead to the global maximum you lead to a local maximum right that is a peak point definitely, but not the best possible peak.

But assume supposing at random point B, if you are allowed a downward movement and if you go let us say to a point say here right. So, what is possible then you can do hill climbing and you can reach to the global maximum. So, this is essentially done by simulated annealing. What is done you know first high temperature is assumed. So, as if you know you reach you raise the entire thing to a very high temperature T which is a very high temperature. Now, at a high temperature, function is used which is called the metropolis function, metropolis function. This metropolis function is such that it actually has got a very high probability rather high probability to you know accept a point which is you know downhill.

So, the point is look here suppose in an iteration you know it is a minimization problem, it is a minimization problem. So, in a minimization problem, you know we have a

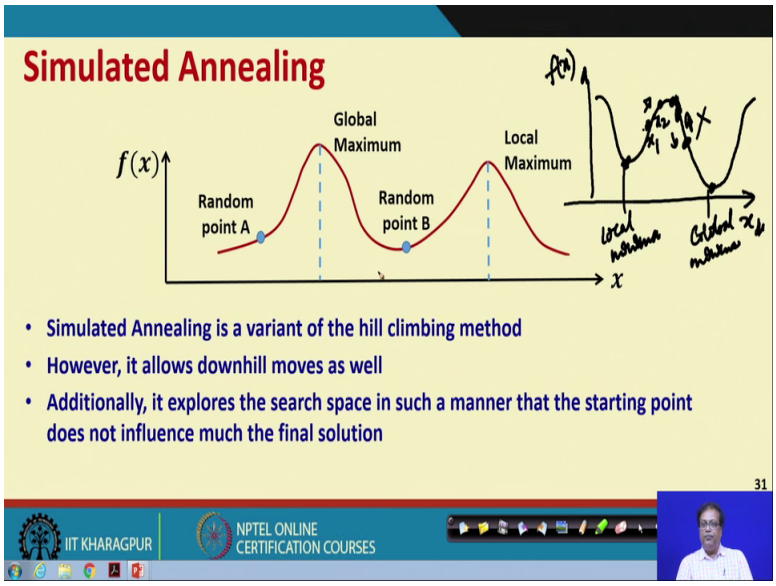
function value f x 1 in the subsequent iteration we got a new point x 2 from x 1; and I got a point x 2 the value is f x 2. So, see what has happened in before iteration, I had x 1 function value f x 1. After iteration, I got x 2 function value f x 2; problem is minimization right.

So, if f x 2 is lower, there is no problem; everything is fine. It is a minimization, we are moving towards the global minimum. Is it not? See, we are moving global or local we do not know, we are really moving lower there is no problem. So, there is no need to really look at the metropolis function or conditions you know you simply accept that point x 2. So, this is how the method goes.

But on the other hand, suppose f x 2 is actually higher, see what has happened. We started with x 1 we got another point x 2 which is higher and the problem is minimizations. So, what has happened we wanted to go towards the minimized points, but here as if you know in a figurative sense, we are going downhill. Here the top of the hill is minimum point and bottom of the hill is you know those points, which are not very rather high values.

So, in that sense, what happens then we have to apply metropolis condition with a high probability of accepting the current f x 2 or the point x 2 even though it has got a higher value right of the functional of the function. So, this is the essential idea that the, you know in a figurative sense for a maximization point. If it is going up, then no metropolis condition; if it is going down, then metropolis condition.

(Refer Slide Time: 06:17)

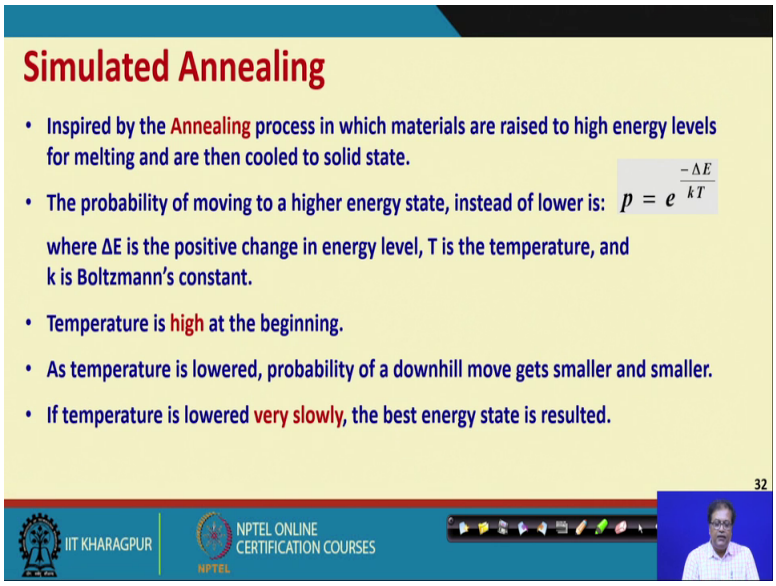


So, however, if it is you know other way round, so let us say the minimization function. So, this is minimization function. So, you see this is a local minima, this is a global minima, so this side is x, this side is f x, so this is the global minima. So, let us say, you know we were here, say we are here, we are going down, so this point is fine. But we are going up, so is it to be taken that up point is taken or not, see the up point could be while in the search for the global minima, or the up point could be somewhere here, and then we are going up.

So, if you are here and we are going up, it is good, because we are avoiding the local. This is local minima, this is global right. So, if you are at starting from here, so this is x 1, and this is x 2, we are going up, this is good, we are going towards global max minima. But if it is here, and we are going up that is not good.

So, therefore, going up has got both good and bad situations. So, what really it does, if you are at a high temperature, so initially when things are at a very high temperature, at the time chances of going downhill is actually favored. The probability of accepting the value, if you are going downwards in case of maximization, and upwards, I mean the function value is higher, you know in minimization, we can still accept. But, slowly very slowly we reduce the temperature, and as the temperature goes up, the probability of accepting the reverse direction movement reduces is not that is the annealing that is where the comparison with the physical annealing process comes in.

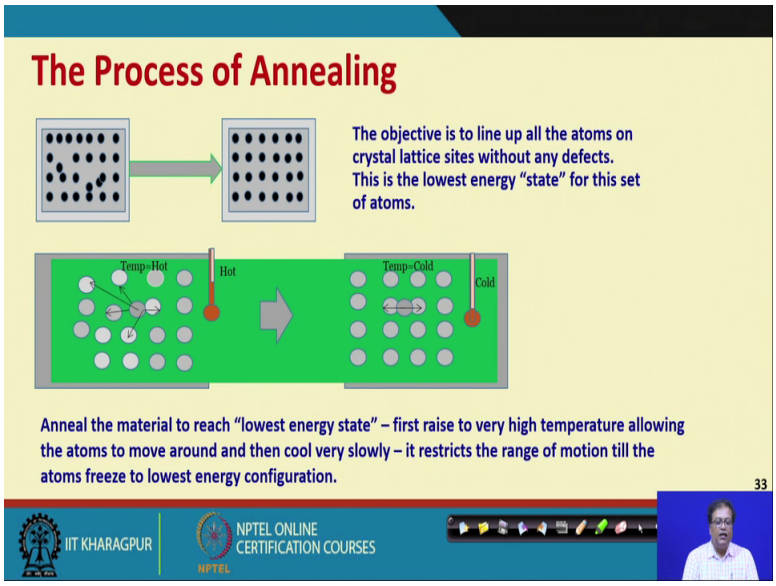
(Refer Slide Time: 08:29)



So, same thing is written here, inspired by annealing processing in which materials are raised to high energy levels for melting, and are then cooled to solid state. The probability of moving to a higher energy state, instead of lower is the metropolis function p equal to e to the power minus del e by k T, where k is the Boltzmann constant, and del e is the positive change in energy level, T is the temperature, and k is the constant. So, for our purpose, we shall take k equal to 1. Although everyone knows in typical you know physical processes, the k is definitely not 1, but ours is not really, truly you know annealing process, we are just drawing an analogy. So, we take k equal to 1.

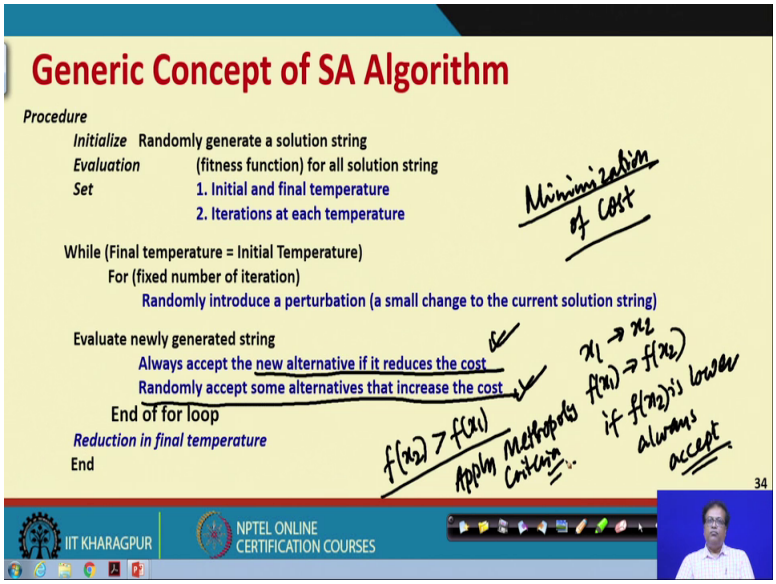
Temperature is high at the beginning. So, when we start at a very high temperature, at the time these probabilities are high right. As temperature goes down, probability for moving to a higher energy state, instead of lower will be less and less. So, as temperature is lowered, probability of a downhill move gets smaller and smaller. So, here, what is uphill; what is downhill, uphill is global maxima or global minima; downhill is opposite direction. So, if temperature is lowered very slowly, the best energy state is resulted. So, as you keep lowering the temperature, the process finally comes to a steady state, and the best energy state is resulted.

(Refer Slide Time: 10:19)



So, this is the physical annealing process. The objective is to line up, all the atoms on crystal lattice sites without any defects. This is the lowest energy state for these set of atoms is it alright. So, you see, they are very nicely arranged. So, when the temperature is hot, you know then all those things, and your temperature is cool, it is a low energy state, and it is much more stable. So, anneal the material to reach lowest energy state, first raise to very high temperature allowing the atoms to move around, and then cool very slowly, it restricts the range of motion till the atoms freeze to lowest energy configuration, so that is the physical annealing process.

(Refer Slide Time: 11:03)

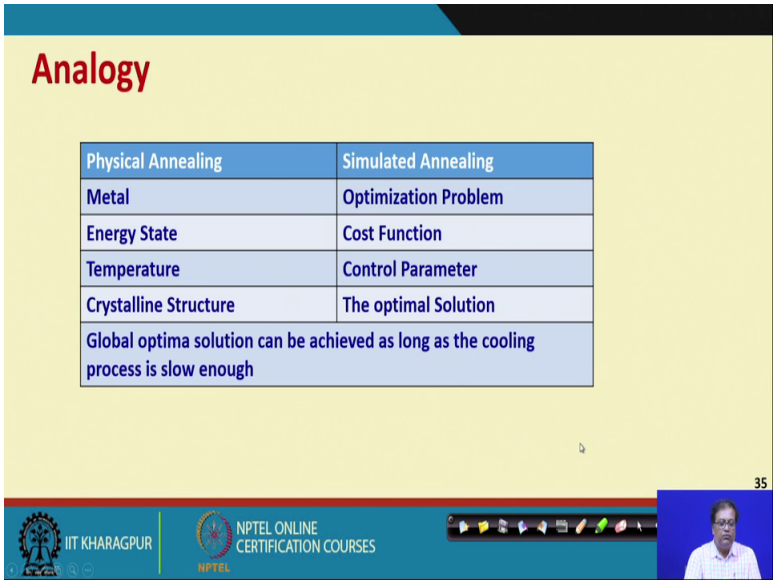


Now, in the essay or simulated annealing algorithm, what exactly we do, let us see that. Randomly generate a solution string, evaluate the fitness function for all the solution string. Then set initial and final temperature; iteration, how many iterations at each temperature. So, while final temperature equal to initial temperature, for fixed number of iterations. Randomly introduce a perturbation right, a small change to the current solution string, so that means, from the point that you start, choose a point nearby.

So, evaluate newly generated string, always accept the new alternative, if it reduces the cost. So, this algorithm is given for minimization, this is given for minimization of cost. This, the example also that we take, we shall take for minimization. So, always accept, if the new alternative reduces the cost, because it is a minimization, so always accept right, so that means, from x 1 to x 2, f x 1 to f x 2, if f x 2 is lower, always accept right.

But randomly accept some alternative that increases the cost. What happens, if the reverse, that means, f x 2 is greater than f x 1, then apply metropolis criteria alright, this is what it is. So, after at certain number of iterations, then you again change the temperature. So, this is the essential process of the simulation simulated annealing algorithm right.

(Refer Slide Time: 13:34)

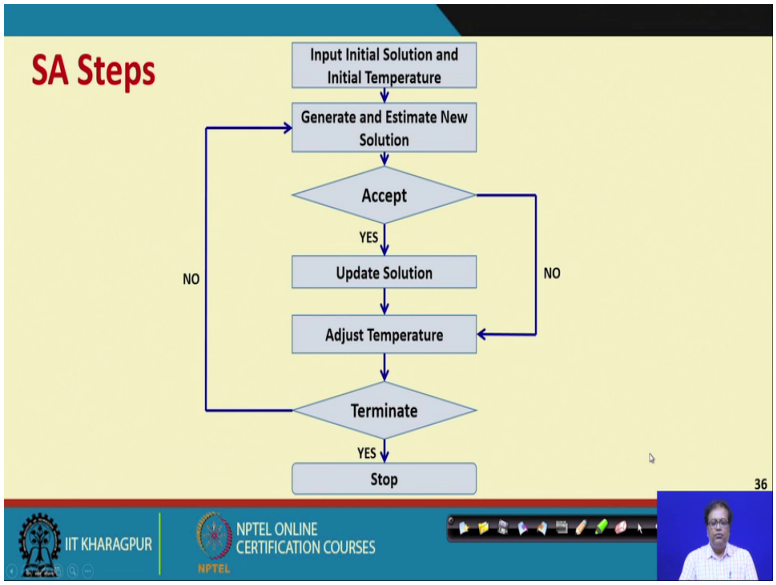


So, now look how the, what is the analogy. The physical analogy that we have the metal versus optimization problem, energy state versus cost function, temperature versus control parameter, crystalline structure the optimal solution. So, global optima should

solution global optima solution can be achieved as long as the cooling process is slow enough that means, very very slow.

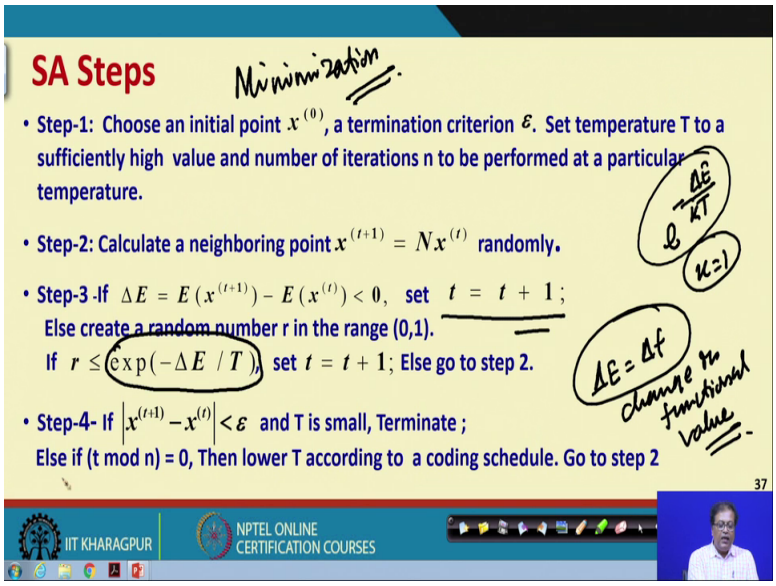
So, why it should be very slow, the point is at in the initially, we should be able to evaluate many points right. And then from those many points, we shall see you know: what is that particular point which is showing a very good; what you call response, so that the system can actually go to the global minima or global maxima. But, please remember, like in GA that global optima is not guaranteed, so it is not an exact method, it is an inexact method. Like all in exact method, the optimum point is not guaranteed right. So, this is shortcoming of all such evolutionary algorithm techniques right, or meta heuristics, which is a general characteristics that I have told you before also right.

(Refer Slide Time: 15:13)



Now, what are the steps? So, here are the step once again, input the initial solution and the initial temperature, generate and estimate the new solution, accept, then update solution, adjust temperature. Now, the update solution is done for a number of iterations obviously, and only then you adjust temperature right; and if you cannot accept, then adjust temperature; and then terminate, no and the go back. So, like this that is the steps of the simulated annealing.

(Refer Slide Time: 15:55)



Let us see the algorithm. So, this is the algorithm that we shall follow for in the example. So, choose an initial point x 0, a termination criteria epsilon, set temperature T to a sufficiently high value, and number of iterations n to be performed at a particular temperature right. So, take a temperature value, decide on how many iterations to be done at that particular temperature get an initial value, and a termination criteria. Now, calculate a neighboring point x t plus 1, which is N x t means, N is a random thing, so randomly. So, then in other words, choose a random point around x t right that call it x t plus 1.

Now, see this is again for the steps are for minimization, so that must be remembered; all these steps are for minimization problem. For maximization problem, the directions should be opposite right. So, the question is that the net change of energy should be negative, because here energy is nothing but the functional value right that is del E is nothing but del f change in functional value. So, if the functional value change is negative, then accept it right without any question. And set t equal to t plus 1 that means, go for the next iteration.

Otherwise, create a random number r in this way that is r from 0 to 1, and this is the metropolis criteria. So, take this is the probability value, the e to the power e to the power minus of del E by T, k equal to 1, already I have said. So, basically original

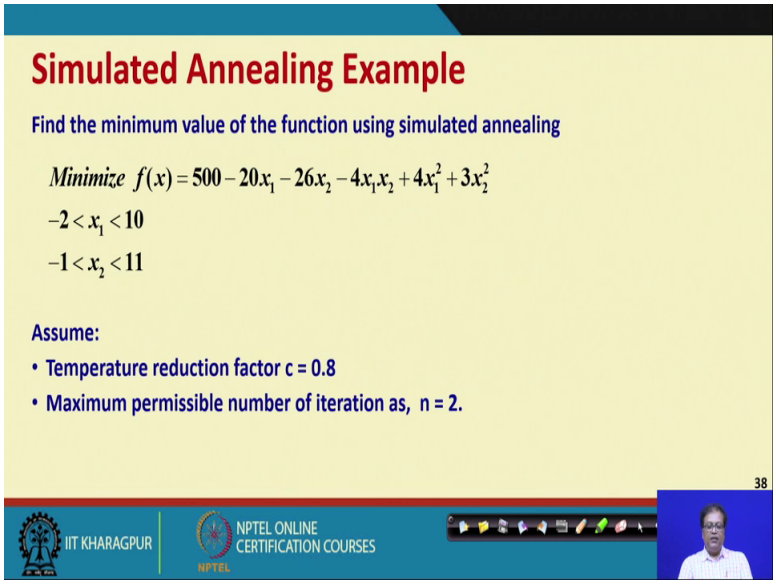
function right, so that was the function e to the power minus del E by k T, where k equal to 1, so that is that function metropolis function.

So, if r that probability that you have taken the random number, if the random number is lower than this, now mind you that as the temperature is reducing, these number will reduce also is it alright. So, what happens that high temperature, this value is high; at low temperature, this value is low. So, chances of acceptance will be high, when T is high is it alright.

So, therefore, there will be high change that assume the r is 0.5, 0.6, around that obviously, it will be accepted. So, if so then accept it, then set t equal to t plus 1. Else go to step two, means discard these point, and generate another point is it alright. So, if this is accepted, point is accepted; if it is not accepted, point is discarded.

Then in the step 4, check whether the new x is very near to the old x right that means, you know no more hill climbing is possible, because if you have reached the peak, any nearing point will not be good enough right. So, T will be small, so terminate. Else if t mod n equal to 0, then lower T according to a coding schedule go to step two, that means, t mod n means that number of iterations have been done right. So, in that case, lower the temperature, and again generate a new point, so that is the essential steps of simulated annealing.

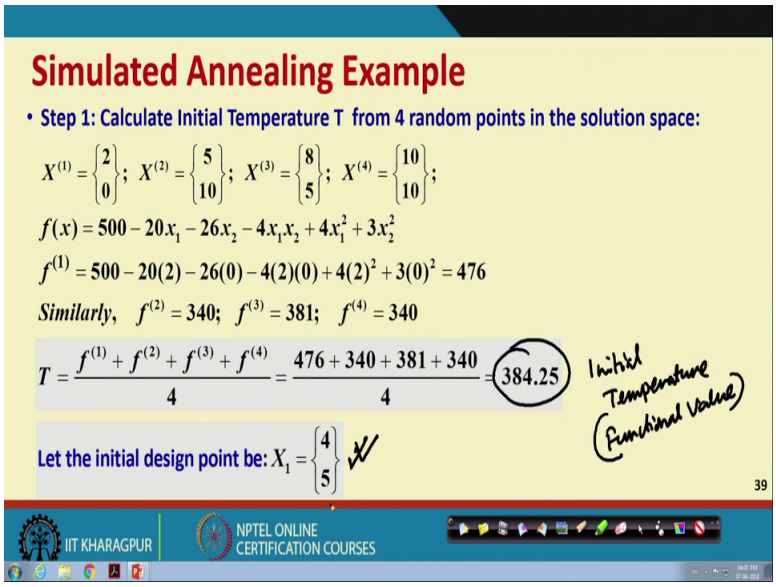
(Refer Slide Time: 20:37)



So, having understood that let us see an example. So, this is the example. Find the minimum value of a function using simulated annealing. Minimize f x equal to 500 minus 20 x 1 minus 26 x 2 minus 4 x 1 x 2 plus 4 x 1 square plus 3 x 2 square, along with some x 1, x 2 conditions. And assume, temperature reduction factor c equal to 0.8, and maximum number of iterations has n equal to 2.

Having said that I have taken these ratios of 0.8, and only 2 iteration, because of we are doing it in the class; When you do computer interactions, you should take more iterations, not just 2, may be 10, 20 iteration, so that you explore the solution space sufficiently for every temperature is it alright. So, those values and probably depending on the problem, you may reduce the temperature more slowly, instead of 20 percent in each step is it alright, so that is how things go.

(Refer Slide Time: 21:49)

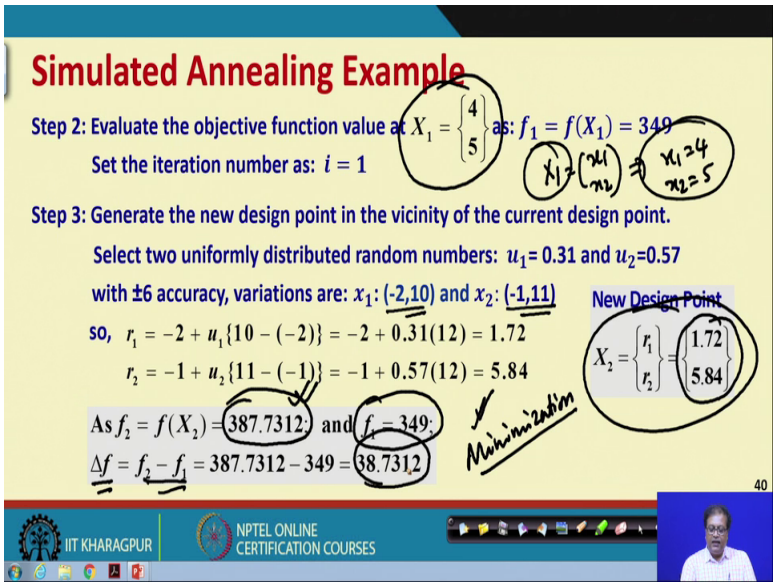


So, how to generate the temperature? So, the first task is to calculate the initial temperature T. So, usually what is done, we take some random points. So, in this case, you have taken four random points. So, in this four random points that is 2 0, 5 10, 8 5, and 10 10. We have calculate the function value, we call that was our original problem minimize f x. So, we find the functional value in each of those random points, for example, in the 1st point 2 10, the value is 500 minus 20 into 2 26 into 0 4 into 2 into 0 4 into 2 square plus 3 into 0 square equal to 476, so that is the functional value at point 1.

Similarly, we can find functional values at point 2, point 3, and point 4. So, when we found out all these functional values, so if I take the average of these functional values 476 plus 340 plus 381 plus 340 by 4 equal to 384.25. So, what is this 384.25, this is our initial temperature. So, the temperature is nothing but a functional value right, it is a functional value. This functional value is a you might think that that is the kind of this is the kind of functional value randomly; you know it is we are around this point right.

Now we have to choose an initial design point randomly. So, again through a randomization process, we got the initial point, let us call it 4 5. So, 4 5 is our initial design point. So, we have computed the initial temperature T by taking four random points, and we have taken an initial design point X 1, which is 4 5 is it alright, so that is the first.

(Refer Slide Time: 24:21)



Then after that you know we go to the next step. So, at the next step, we find the functional value at our initial design point 4 5. So, 4 5 is our initial design point, the function value is 349 right, and this is our first iteration. So, what we do, generate the new design point in the vicinity of the current design points. So, this is our 349 is our current design point. So, select two uniformly distributed random numbers, u 1, and u 2 right. So, two random numbers we have to obtain.

Assuming those two random numbers are 0.31 and 0.57. Why are they 0.31 and 0.57, they could be any other number right. Nothing very great about 0.31 and 0.57, it is a

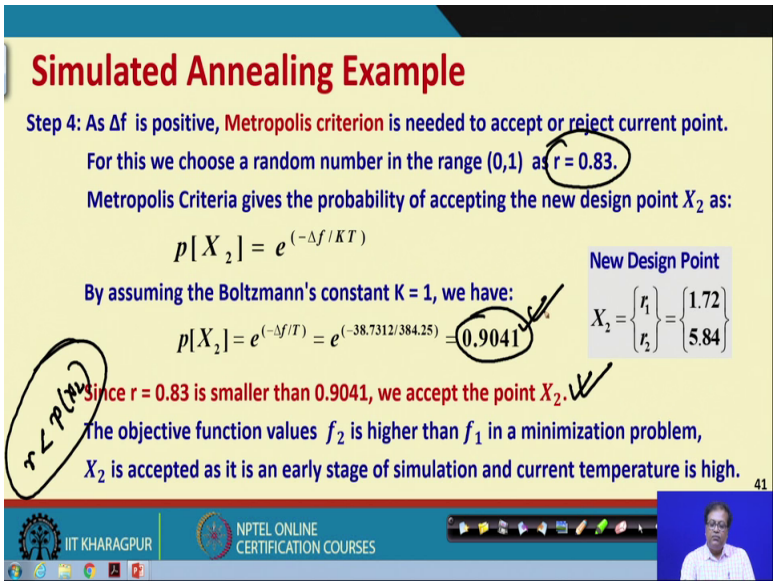
random number; we have obtained from calculator or by some other method. So, these are the two random numbers, assuming plus minus 6 accuracy. So, since X 1 is 4. So, actually X 1 point X 1 is small x 1 and x 2. So, these are the two variable, so which gives x 1 equal to 4, and x 2 equal to 5. So, this is the that is we are calling point x 1 the capital X 1 right.

So, now x 1 will be varying, because of plus minus 6 minus 2 to 10, and x 2 minus 1 to 11, 4 5, so like this. So, using that random number, so the what will be the new value, new value will be minus 2 plus u 1 into 10 minus minus 2, and minus 1 plus u 2 into 11 of minus these two. So, this point comes to 1.72, and 5.84. So, this is going to be our new design point.

Now, look here, in the new design point, what is the function value. So, if we if we put it in f x, this point 1.72, and 5.84, what do we get, we get 387.7312. Now, how good is this point, this is good or bad tell me, the original function was one of minimization right. When X 1 is equal to 4 5, that means, capital X 1 the initial point, then the function was 349.

Now, at after iteration, we got a new point 1.72 and 5.84, and the function value is 387. So, it is a good or bad, in a traditional hill climbing sense, this is a bad point. Why, because it is a minimization problem. And in the new design point, the function value is increasing. But we have seen that to really cut off the global you know minima from our consideration, we might also sometimes except you know lowering of value, or in this case, increase of value going in the opposite direction based on the metropolis criteria. Let us see what is that. So, the del f in this case is f 2 minus f 1 that is 38.7312 that is our del f. So, what we have to do, we have to apply the metropolis criteria.

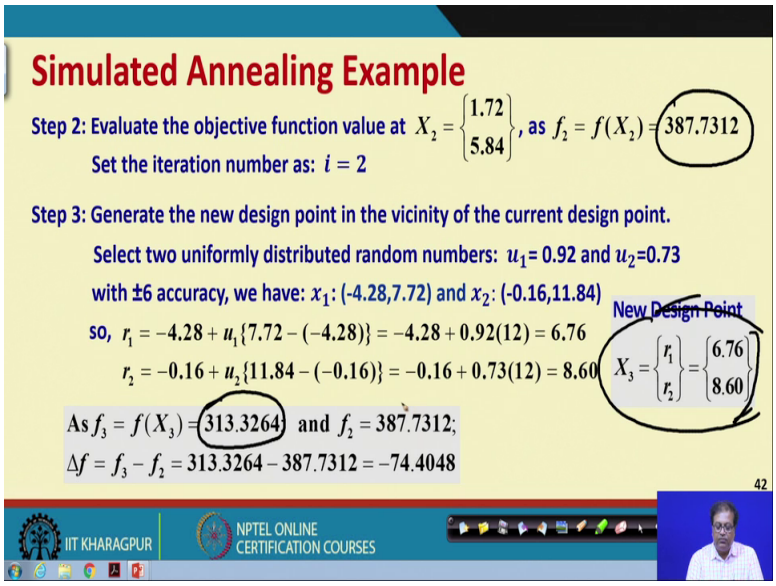
(Refer Slide Time: 28:02)



So, let us take a random number, assume that that random number is 0.83. So, this is our random number 0.83. Now, we have p X 2 is e to the power minus del f by K T, and K equal to 1. So, we see, this value is 0.9041 right. So, what we got, we got r is less than p X 2.

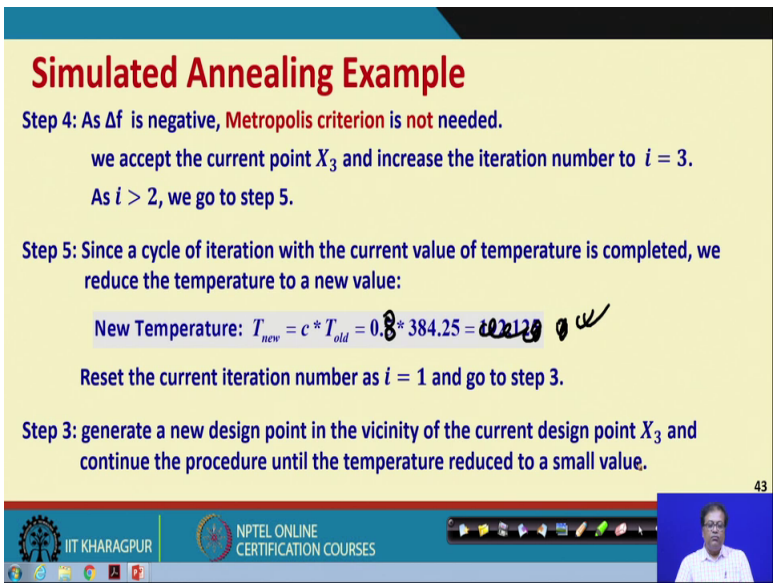
So, our random number is less than; that means, we accept this point. So, although the point is a bad point, but we still accept these point as the objective function value f 2, although it is higher, but still accepted. And usually this is happened, because of high current temperature that is how, this probability has come such a high value. Note the point, so that is how, it is done.

(Refer Slide Time: 29:11)



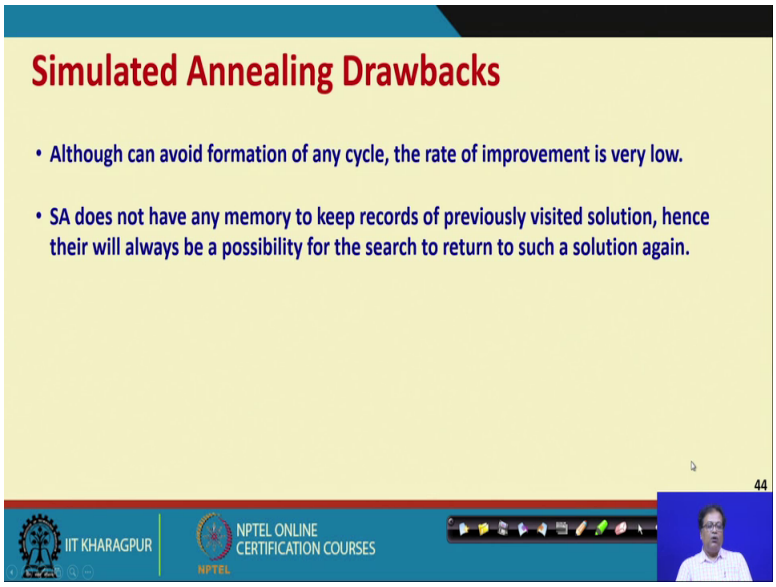
And further if we do iteration at the second iteration, again by the same method. We take two random numbers, and with plus minus 6 accuracy around this point, that means, between minus 4.28, and 7.72. And these two ranges, again we take using these two random numbers a new point. And this new point is X 3, which is 6.76 and 8.6. So, when we put that the new function value comes to 313. Look here, the old function value was 387. So, from 387, we have actually come down by 74. So, this is the good thing, it is a minimization problem, and we are going down, so that means, the point should be accepted, there is no need to apply metropolis criteria is it alright, so that is the thing.

(Refer Slide Time: 30:07)



So, we have done our two iterations, metropolis criteria not required. Since, a cycle of iteration with the current value of temperature is completed, now we reduce the temperature to a new value. Actually it will not be 0.5, this will be 0.8 right. So, whatever is the value, 300 something that will be coming right, so that will be our new temperature. And our step 3, generate a new design point, and like this, we will continue is it alright. So, like this, we do the simulated annealing process right.

(Refer Slide Time: 30:53)



And let us see further some drawbacks of simulated annealing. Although can avoid formation of any cycle, the rate of improvement is very low that is the good bad point. SA does not have any memory to keep records of previously visited solution, so the point is we might come back to the same solution again right. So, this is the simulated annealing method. We shall see more such methods, in our subsequent lectures right.

Thank you very much.