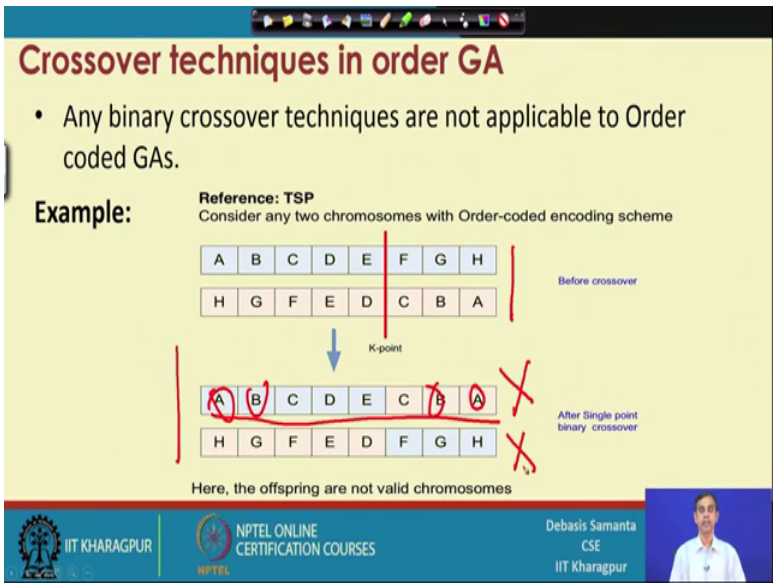
Introduction to Soft Computing
Prof. Debasis Samanta
Department of Computer Science & Engineering
Indian Institute of Science, Kharagpur

Lecture – 22
GA Operator: Crossover (Contd.)

So, different GA encoding scheme follow the different pattern of chromosome, the binary coded GA follow the binary patterns, the real coded GA follow the real values of the gene values and depending on the patterns that it is following binary coded GA, so the binary cross over techniques were used, we have learned over the binary crossover techniques. Now, real value coded GA, again it is a totally different than the binary crossover techniques because it needs the several totally different what is the treatment.

Now, we are going to discuss another GA technique it is called the order GA and the crossover technique that is there in the order coded GA.

(Refer Slide Time: 01:04)



Now, again the order coded GA as we know, it is basically based on the concept of the sequence of the values that is there in the GA, so the sequence is important. As the sequence is important for example, the travelling salesman problem if we follow that all the values that is there in the chromosome should not be repeated and they should follow certain sequence actually. Now, this means that if we follow the binary crossover

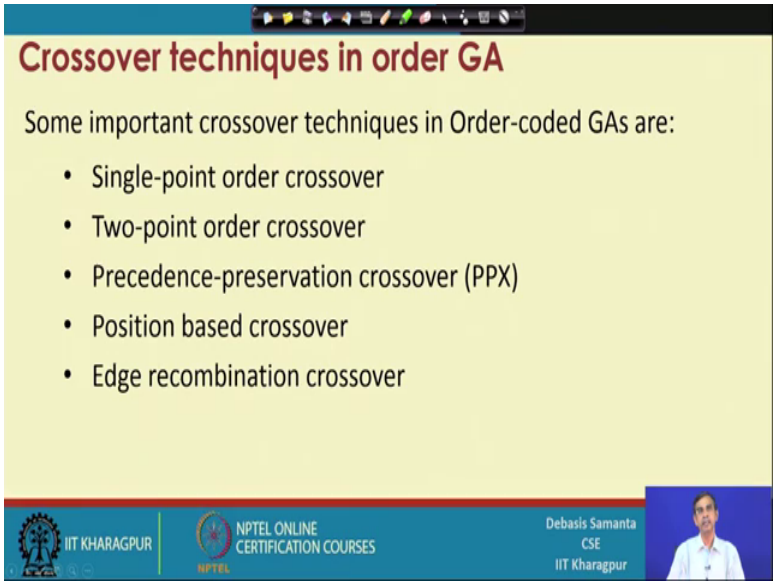
techniques; obviously, these are the basically in terms of symbols, so no real values are involved. So, that is why we cannot apply the real coded GA.

However, the binary coded GA cannot, the crossover technique that is used in binary coded GA also cannot be applied here, this an example because how the binary coded the crossover technique that is followed there in binary coded cannot be applied here. Now, if you considered say binary crossover technique namely the single point crossover technique, we can recall that we have to generate a K point there, so if this is the K point, then basically the swapping these 2, so it is swapping this 2, swapping this 2, we will get this 1 and then swapping this one also will get this one.

So, it is basically from these 2 parent chromosome using the binary single point crossover we will get it, but you can note that this parent chromosome is not a fusible chromosome or in order to acceptable chromosome because A it is common here, B, B is copied here and all the chromosome that is not all so possibly present here. So, this is not a valid chromosome or this is also similarly not a valid chromosome; that means the simple single point crossover technique that is used there in binary crossover, it is not applicable to the order GA in fact.

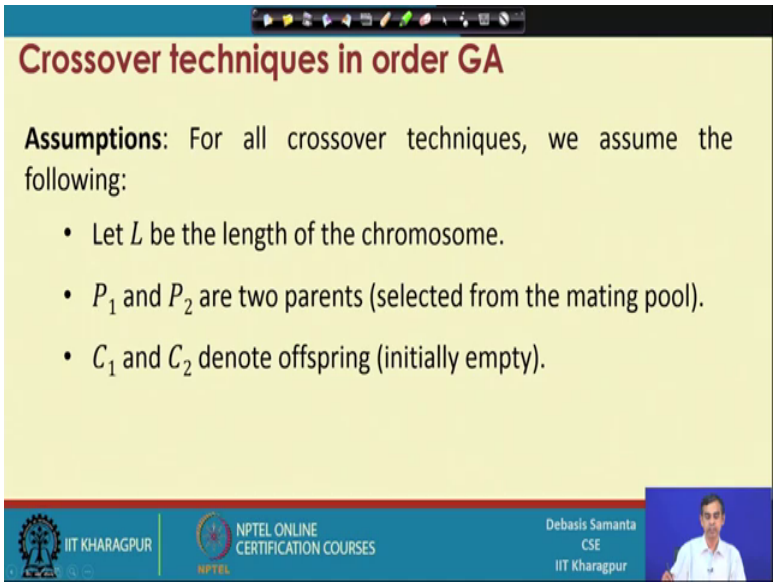
So, this means that order GA needs a different treatment, so far the crossover operation is concerned.

(Refer Slide Time: 03:01)



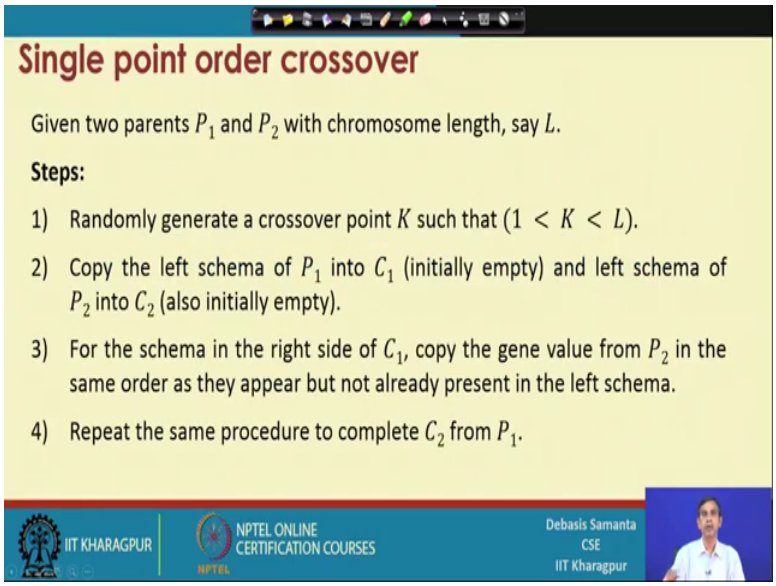
So, we will discuss about the different crossover technique that is followed in case of order coded GA, we have listed few important techniques 5 important techniques are there, it is 1 is called the single point order crossover, then second is 2 point order crossover, then precedence preservation crossover it is called the PPX, then position based crossover and then edge recombination crossover. All these crossover techniques we will discuss one by one in the next subsequent slides.

(Refer Slide Time: 03:29)



Now, in order to discuss the different technique that we have mentioned in the last slides, we will follow certain assumption for all the techniques to discuss, the first is that we will consider that the length of the chromosome be denoted as L, L is an assumption that the length of the chromosome this one. P1 ad P2 are the 2 parents which are selected a random from the meeting pool and C1 and C2 denotes the 2 children's which we want to derived from the P1 and P2 by virtue of the crossover techniques followed there. So, these are the assumptions under this assumption we will be able to discuss each technique one by one. Let us first start with single point crossover technique in order GA.

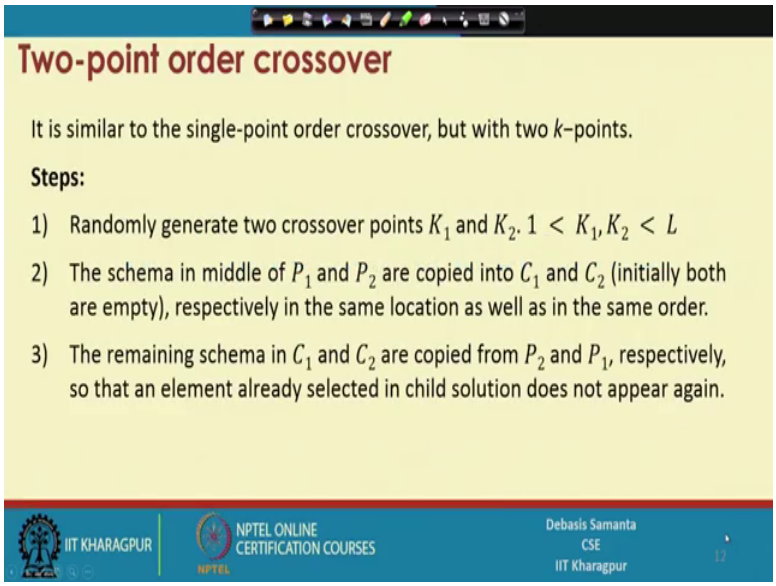
(Refer Slide Time: 04:13)



Now, so if L be the length of the chromosome and P1 and P2 are the 2 chromosomes, then in this technique the first task is to generate 1 number that number should be in between 1 and L and let this number be K. So, it is basically the same as single point crossover is a (Refer Time: 04:37) or point life. So, K point is to be decided first, once the K point is decided this K point is decided then is the next point, next task is we have to copy. So, K point is a (Refer Time: 04:50) point that is the point that can define the 2 parts in both parents, so left part and right part or you can say left schema and then right schema.

Then, the second step, copy the left schema of P1 into the children C1. C1 is initially empty and then left schema of P2 into C2 and then for the schema in the right side of C1 copy the gene value from P2 in the same order as they appear, but not already present in the left schema. So, if you repeat the same procedure to compute C2 from P1, then you will produce the 2 children solution. So, this is a technique or scheme that is there we have to follow it. Now, let us see how this techniques are there as an illustration, so that you can understand this technique better.

(Refer Slide Time: 05:40)



So, we assume these are the 2 solution appearance P1 and P2 and then a random K point is decided, this is the K point from where the left schema and right schema. So, this party is the left schema and this is the right schema for the parent P1, similarly this is a left schema for the parent P2 and the right schema for the parent P2.

Now, according to this technique, so idea is that, we will copy the left schema from P1 to C1. So, it is basically copy, this part is copied to this one and for the rest of the part we will copy from the P2; from the P2, but provided that values is already not present in the left part. For example, if we see E, E is already present there, so E cannot be copied and then D is also present here D cannot be copied, C is also present here C cannot be copied, then J is not present here, so, J is copied here, then I, I is not present, so for though I is there, the H is there, now B, B is also not there, so B is copied here, A is already there, so A cannot be copied, then F is not copied the F is copied and finally G is copied.

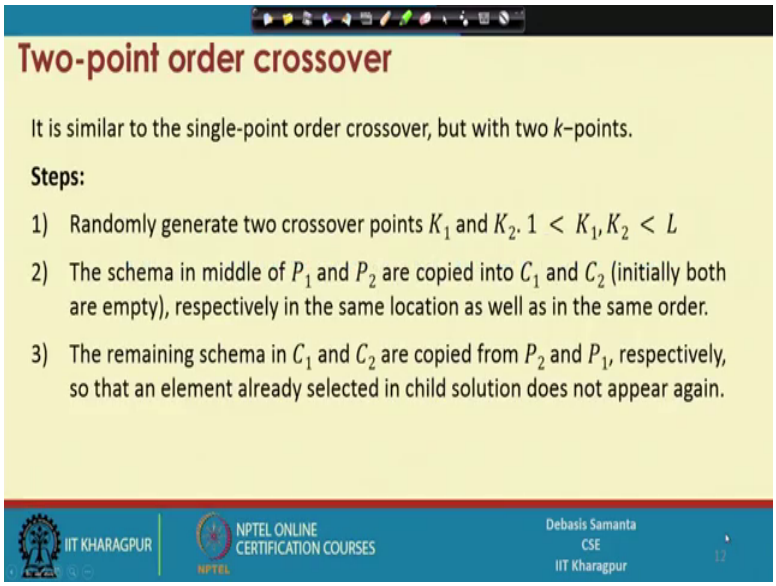
So, this way from the parent P1 and P2 and based on the kind of a the solution C1 can be obtained. Now, similarly the C2 can be obtained, in this case this left schema will be copied to this one and then for the rest of the schema we have to copy it from here right and provided that this is not copied already.

For example, A, A is not in this, so A will be selected and C is there C cannot be copied, D is there so D cannot be copied, E is here so E cannot be copied. So, then B, B can be

copied here. So, B can be copied here and then F is not copied F is copied there, G is copied there, H is there, not copied. So, I is there and so J is there, J is already there. So, J cannot be copied and I is there this way. So, this way the 2 chromosomes solution can be obtained and this is the simple technique that is called the single point crossover technique in case of order GA.

Now, we will discuss about another little bit more what is called the diversified technique we can say and it is called the 2 point crossover technique. So, the difference it is basically by it is name, in case of single point we have to consider only 1 K point, but in case of 2 point crossover we have to consider 2 points, 2 K points rather.

(Refer Slide Time: 08:25)

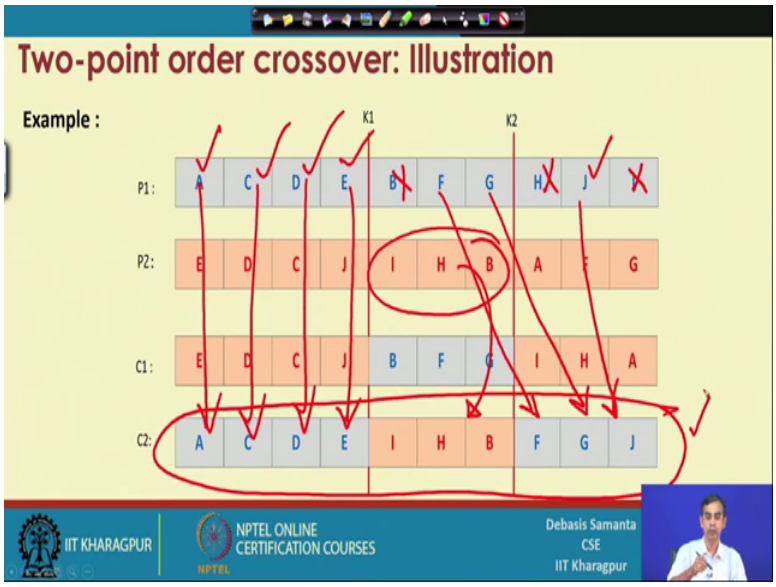


So, it is just procedure of 2, I mean deciding instead of 1 K values we have to decide 2 K values and these 2 K values are denoted K1 and K 2; the 2 values are the same as it is in case of the previous 1 scheme; that means, the values of the K values should be in between 1 and L.

So, once we decide this 2 K values, then the scheme basically says that the middle of P1 and P2 are copied into C1 and C2. So, initially C1 and C2 are empty, so on the middle part from P1 is copied into middle part of the C1 based on these values K1 and K2. Similarly, middle part of the P2 is copied into the C2 and then once this values is copied, then we have to fill the remaining portions both in the left side as well as right side in both C1 and C2. So, it will follow the same procedure as in case of single point order

crossover, so for the remaining point we have to compare in case of C1 we compare the gene values from the P2 and for C2 we will compare the gene values from P1 and fill it up.

(Refer Slide Time: 09:31)



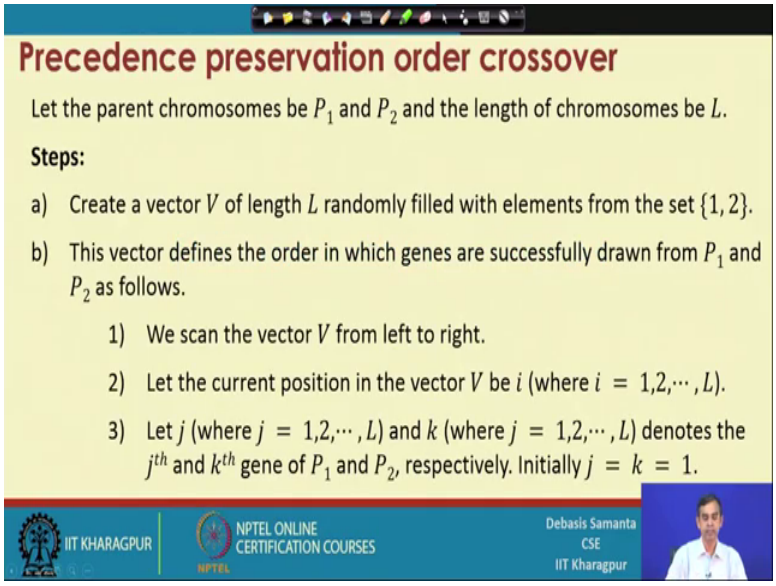
So, let us say 1 illustration to clear this idea. So, here we will consider 2 parent chromosome P1 and P2 and then 2 K points K1 and K2 are decided at random which is this one. So, in this idea, the idea first is that this is the middle part is copied to the C1 first and for the rest of the part this one will copy from the parent P2 provided that it is already not there in the parent; so, in the not there in children's C1.

For example, so B F G already copied, then we will see that E should be selected and E is there D is also not covered, so D is selected here, C is selected here, J is also selected here, then B F G already there, then come here I, so I needs to be selected here because I is not copied, now H, H is also not covered here, so H will be here, B cannot be because already B there and then A, A is not copied here, so A will be there and F and G already there, so it is there, so this way the children chromosome can be obtained.

Now, similarly the C2 can be obtained; in this case C2 will copy the I H B to here and for the rest of the part will copy from here. So, I H B is there, so A should be there, then C should be there because C has not been copied so far and D is there D has been copied and E, E should be copied here, I H B it is already there, then the next part is B, B is not there because B is already here.

So, B is rule out and then if F is copied here because F is not covered and J is copied here because J is not covered and then H, H is already there. So, it H is not covered and then J, J is not covered, so J is copied here and then I is there I already, so this A. So, this way the children C2 can be formed. So, this is the idea about 2 point crossover, it is little bit different, then because the single point crossover is pretty simple compare to the 2 point crossover, but it keeps the better I am mean diversity in the chromosome solution. So, it is more preferable than the single point crossover; however, this crossover little bit costly operation then the single point crossover; so, next will discuss about the president preservative crossover techniques in order GA.

(Refer Slide Time: 11:58)

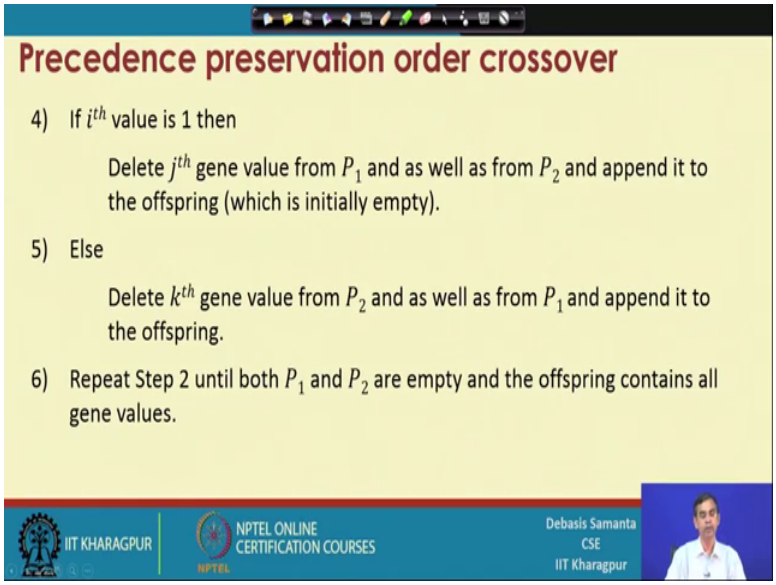


So, we will discuss about the technique here. So, it basically as it is in case of the earlier 2 crossover techniques in order GA, will follow the 2 parents P1 and P2 and we assume that length of the chromosome be L.

Now, it basically consider 1 vector; vector with the 2 different values, values they are called 1 and 2, so a vector of same size of the chromosome length L. So, that is why I create a vector B of length L and this is randomly; that means, 1 vector that can be created with its constituents either 1 or 2 and the length of the vector be L. So, this is basically is called the other pivot 1, it is just like a mask in case of binary crossover technique have a have uniform crossover technique that we have discussed in binary coded GA actually, so it is just like a mass like.

Now, then the scheme that is followed in PPX crossover, it has like this. We scan the vector V from left to right; that means, each time we will see whether the current component is 1 or 2. Now, let the current position in the vector V i that means, we are currently scanning, so it will start from i equals to 1 to the maximum up to L and then j where j is basically 1 to L it is basically a pointer to the first chromosome parent P1 and k is another pointer to P1 indicates that at what point of the P1 and P2 we are currently traversing.

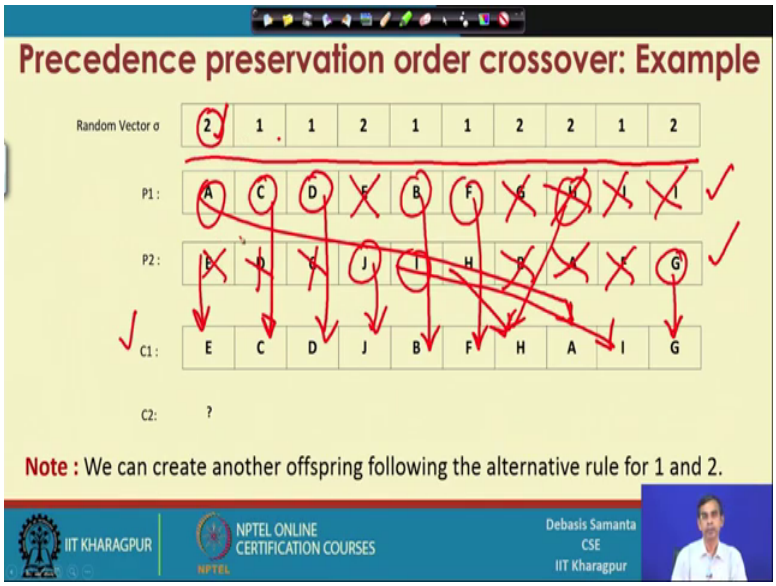
(Refer Slide Time: 13:40)



Then, this technique knowing this one, so it basically follows the 2 method. If ith value is 1; that means, currently the component that is there in vector B is 1, then it basically the idea is that delete the j gene value from P1 and as well as from P2; that means, it is select the j gene and then remove this j gene from P1 and P2 not to be copied further and then append it to the offspring, which is initially empty, so it is basically C1; suppose, we are considering the case 1 of children C1.

Now, if the ith value is not 1; that means, it is 2, then delete k gene; that means, we will just go to the Pth chromosome and as well as from P1 and append it to the offspring. So, it is basically where they were 1 and 2, it is basically will delete from P1 and P2 and then copied into the offspring actually. So, we will repeat the 2 step until both P1 and P2 are empty and the offspring contents all the gene values; so better if we can explain the concept of this technique with an example, so here the example.

(Refer Slide Time: 14:51)



So, this is the vector V, with size same as the parent P1 and parent P2 and will see how the C1 can be obtained. So, here the idea is that if P is 2, I mean if the current value is 2 then we will copy from P2 and if the current value 1 will copy from P1, so it is like this. If 2 it is there, so this is copied; here is 1 is there so will copy from this one. Now, when will copy E so, all E should be deleted both from P1 and P2, so it is been copied there.

Next, when we wants, so we will copy C and then C as it is already copied, so C will be deleted from the P2. Now, again 1, so will this one D will be copied and then this D will be removed from there, then 2, so if it is 2 then will copy this 1 from the parent P2. So, this is copied here and J, occurrence of J will be deleted from there.

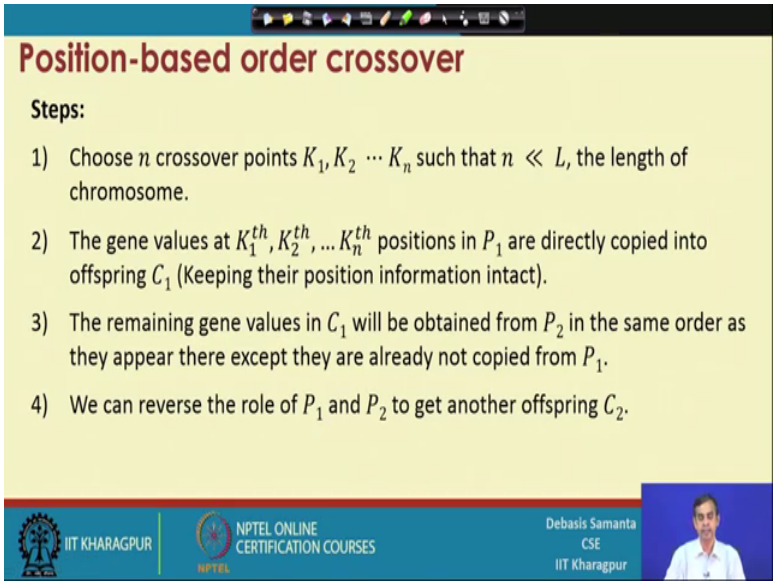
Now 1, so B is copied here and then B is deleted from the parent P2; 1 so F is copied here and then F is deleted from P2; 2, 2 G, so 2 means it will be G then, G is basically here. So, G will be copied it from the parent P2 and then other G occurrence will be deleted from there. So, 2 H, H will be copied from here and then also 2 H H means H is here. So, H will be copied here and then all this H will be deleted from there.

Now, so, then we have to see the 1, 1 means we will copy from P1, so P1 will be copied here. So, and then, so this A will be deleted. So, finally, I, so it is 2, 2 means the I is to be copied here and then all other will be removed here. So, this way the entire gene can be copied and then it will produce the offspring. Now, if we reverse the formula policy; that means, if it is 2 then copy from P2, earlier if it is 2 copy from P2 then P1, if it is 1. Now,

we revise the policy; that means, if it is 1 then copy from P2 and if it is 2 then copy from P1 and then we will follow the reverse one, so the another chromosome C2 can be obtained.

So, this is our precedence preservative crossover techniques in case of order GA and it was like this.

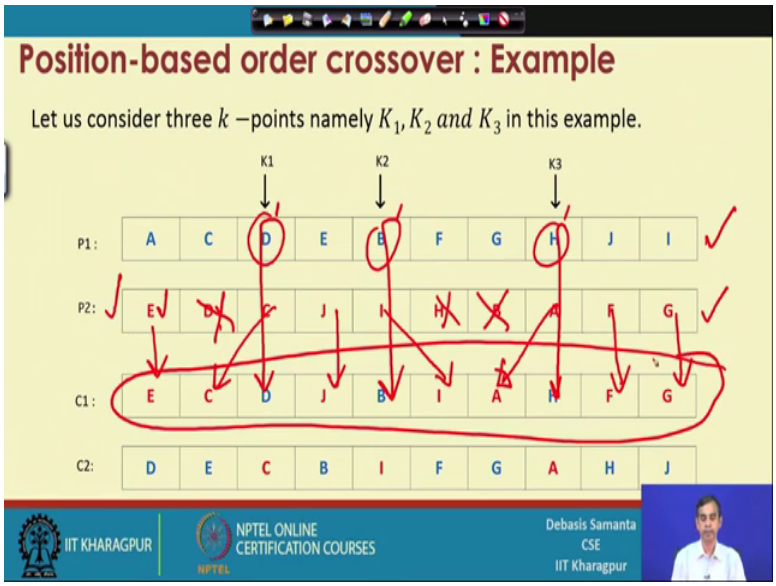
(Refer Slide Time: 17:50)



Now, another 2 technique is called the position based crossover technique, this technique it is more generalized version of the 2 point cross over technique in fact. So, here the idea is that choose n crossover points K1, K2 dot dot K n, where n will be sufficiently large than l. So, this is a crossover technique usually followed if the length of the chromosome is too large. So, that we can decide a large number of K points in fact and then, the idea is basically the gene values that K1, K2 and then K n position in P1 are directly copied into the offspring C1, keeping their position same; that means, K n value from P1 is copied to K nth position in C1, K 2's values in P1 is copied to K2 position in C1 and so on.

So, this way it will be copied and then so it will get partially filled some chromosome and then for the rest of the chromosome we have to take the confidence of P2, we have to take the we have to copy the chromosome values from P2 provide that they are already not there in C1. So, if you follow the reverse; that means the reverse of the previous procedure, then it will give another chromosome C2.

(Refer Slide Time: 19:08)



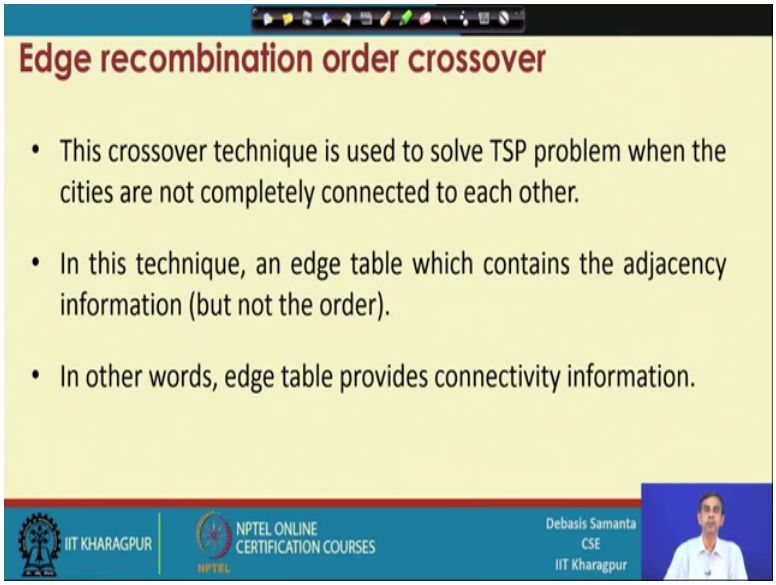
So, let us explain illustrate this technique with an example, will consider this is the P1 and another chromosome P2 and here we consider 3 points K points, these are called the K1, K2 and K3. So, account the scheme the first scheme will produce a C1, so this D 1 is copied here and then this B1 is copied here and then this H1 is copied here.

Then, for the rest of the chromosome we have to take it from the P2 provided already the chromosome which is there should not be into there. Now here, so D B A, so H, so will extract the values or copy from the P2 except D B and H which are already there. So, E it is there, so E is coming, D it is there which is not there. So, D is already there, so D should not be here, C it is there because C is not copied there, J it is there because J has not been copied and then I A it is there, so I, I it is there this I is coming here and then B already there, so B cannot be copied and H cannot be copied because H is already there and A then A can be coming here and F it is not there, so F will be copied and G can be copied. So, this way the children chromosome C1 can be obtained.

Now, if we follow the reverse procedure in the same that, if we copy the K1, K2 and K 3 points into C2 form P2, then will get another offspring for example here, this C is here and this I is here and this A is here. Then for the rest of the part we will copy from here provided that all this things are not there. So, this way you can check it, so this chromosomes can be obtained. So, this is the position based crossover technique there, then the last technique is called edge recombination order crossover technique. Now,

edge recombination order crossover is a special case, it is bit comparisonaly expensive, but very famous for the problem like travelling salesman problem.

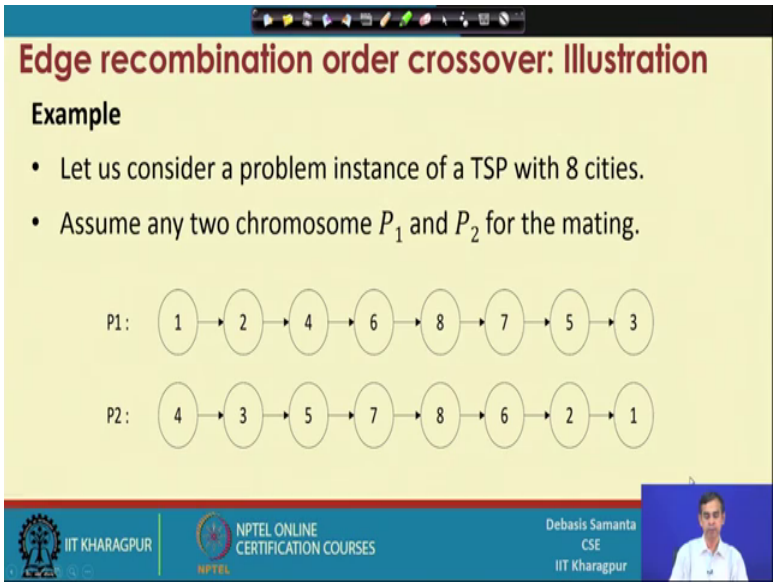
(Refer Slide Time: 21:25)



So, the crossover technique is used to solve the problem like (Refer Time: 21:29) told you and so, and also that kind of TSP problem where the cities are not well connected. So, eagerly it works better there and then it is also comparisonaly very fast because the number of chromosomes is basically n factorial and for a large values of n that is really very difficult to find because it is a comparisonaly expensive operation to find the all possible order sequence that is there in possible in the, if the all cities are highly connected.

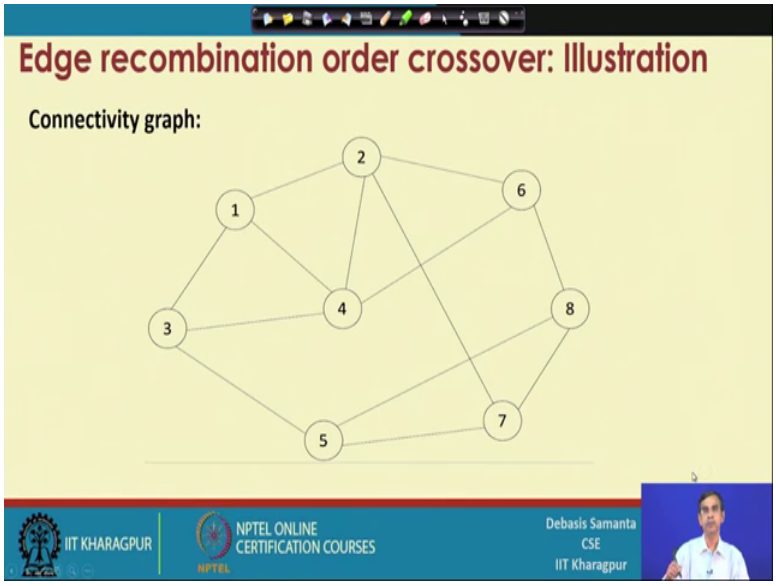
Now, so in this technique basically we will follow on lookup table it is called the edge table, which basically contains the adjacency information and then that atoms is not necessary a particular order in the random order; that means, if a city A, the city A is connected to which of the cities say it is B, C, D, E then they should be present in that edge table.

(Refer Slide Time: 22:35)



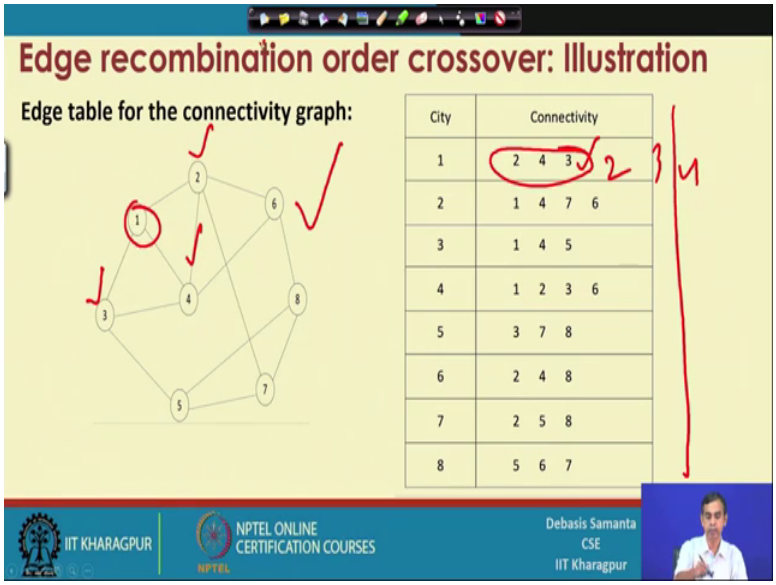
So, then once the, this edge table basically provides the connectivity information in some different form; now, as an illustration we can consider this problem like. So here, basically the idea is that say suppose these are the 2 parents P1 and P2 for edge cities problem and say you can say that these are the order sequence that is there and this are another order sequence. Now, so we want to find 1 another children C1 from this 2 P1 and P2 like.

(Refer Slide Time: 22:59)



So, idea it is basically first we have to create the edge stable and edge stable for a given instance, so this is the problem instance, this basically the showing the connectivity of a different cities there and as we see that all cities are not connected to all other cities in fact, so there are connectivities like this. Now, will see for this city map how we can produce the edge table?

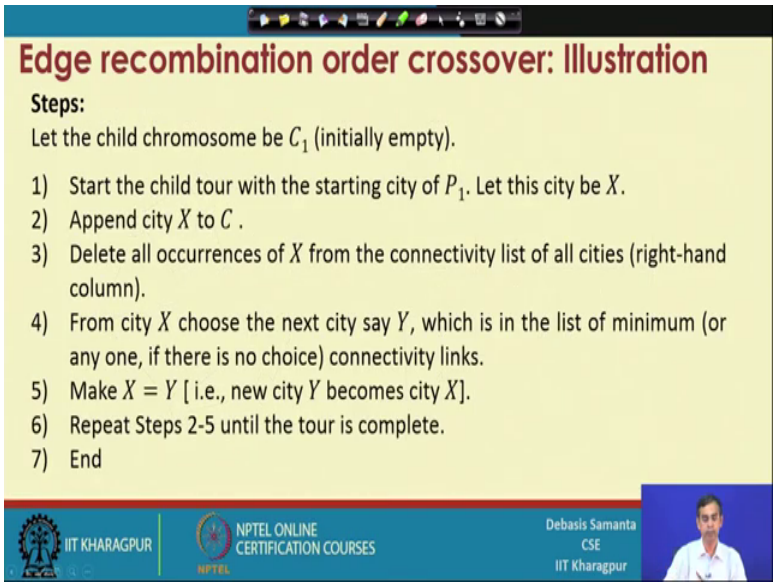
(Refer Slide Time: 23:26)



So, we will see that edge table, so these basically shows the edge table for this city map and as we see for the city 1 we have the 3 connections 2, 3 and 4. So, we have written 2, 4, 3 and this order is not important, if you say 2, 3, 4 that is also valid. So, the order is not important.

Now, likewise for the city 2 as we see the connectivity as 1, 4, 6 and 7, so it is like this. So, all the connectivities are put it there in the edge table. So, basically idea is that once the city map is known to you. So, city map to know to you, then we will easily obtain this edge table and then this edge table is used for the generation of chromosome for the children.

(Refer Slide Time: 24:21)



Now, let see what is the procedure followed, here the idea is that, so initially the children chromosome we denote it as a C1 and initially we assume that it is empty; that means, is blank nothing is there then start the children, start the tour with the for the children's tour with the starting city of P1; that means, it is same as P1 and if we take the starting city of P2 then another chromosome will be obtained.

So, let us start with the P1 first as a parent. So, we will start the; that means, both P1 and C1 have the same starting city it is called the starting city has same as the P1. So, let us denote this city be X, then we will add this city X to C; that means, this is the first city for the children solution.

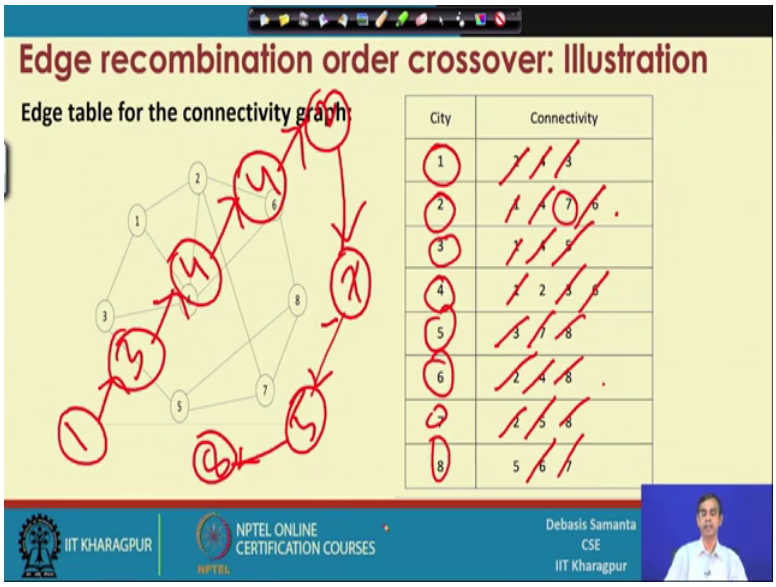
Then, once the city X is selected delete all occurrences of X from the connectivity list of all cities that mean as C if the X is selected. So, it should be removed to not to be considered for the others, for the next time, so that that X should be deleted from all the connectivity information there.

From city X, then for the city X choose the next city say Y. So, from city X we can and travels into some other city which has the connectivity Y and which is in the, so this the 1 condition that city X to city that connectivity Y and that also will select that Y because many cities are possible, we will select that Y which in the least of minimum connectivities are there. So, it is like this and then will copy this X to Y and then we

make X Y and then repeat the same procedure till we will compete the entire tour for the city; for the solution chromosome C1.

Now, here is an example that I can tell it. So, suppose the starting city of P1 is 1, so will start from 1 and then from 1 we see that, so 1 it is selected. So, this 1 will be removed, this 1 will be removed, so this is removed because city 1 is selected. So, this is the initially city 1.

(Refer Slide Time: 26:32)



Then, the next city we have to select from city 1 we can go 2, 4, 3. Now, in case of 2 the connectivity is 3, in case of 4 it is again 3 connectivity that there in case of 3 it is 4, so we should select the minimum connectivity that mean 3 in this case, so we will select the next city as 3 and then as 3 is connected. So, we will remove this 3 from every occurrences in the connectivity matrix that mean 3 has been covered.

Then, so we are in the city 3 and from the city 3 we can go to 4 and then 5. So, city 4 and 5 if we go there 4 has the connectivity 2 and 6, whereas 5 has the connectivity 7 and 8 both are same. So, we can take any orbital anyone, so let it be 4. So, from 3 to 4 we can go to the city 4 and then 4 is covered, so 4 will be removed, 4 will be removed, 4 will be removed and 4 will be removed, so this way. Now, so 4 is covered then 4 from the city 4 we can go to either 2 or 6.

So, we can go anyone, but the thing is that 2 it is a connectivity 7 and 6 and for 6, 2 and 8, so we will go anyone. So, from the 2, 7 and 6, so 7 and 6 is 2 and 8 and 7 so we can go to from 4 actually 2 and 6. So, 2 has the connectivity 7 6 and 6 has the connectivity 2 and 8. So, we can go anyone, let it be moved to 2 right, so 6 1, so 4 to 6.

So, 6 is connected and then 1 6 is connected it will remove the 6 from every occurrence it is there, from the 6 we can go move to 2 or 8. So, 2 has the only 7 and then 8 has 2, so we can go to 2. So, we can go to 2 to from 4 to 6 and then 2 will be removed from here and there. So, from 2 we can go to the 7 finally, so 7 is there, 7 has this 5 and 8.

So, it is 7, 7 and 5 and 8 out of this 5 and 8, 7 is deleted. So, 5 and 8 it is there we can go anyone, so it is anyone maybe 5 say 5 and finally, so 5 is deleted and finally 5 to 8, so the 8 into there. So, this basically gives the children chromosome according to the edge combination technique, so this way you can follow it. Now, as we say, that so total tour is completed and covering all cities there, so this is basically the idea.

Now, here we have started with the starting point of city P1. Now, if we follow the starting point of city which is P2, let it be say 4 then definitely it will produce the different 1 chromosome, so that will be considered as C2. So, this way so P1 and P2 chromosome influence to obtain the 2 children solution C1 and C2 according to the edge recombination technique.

So, we have learned the different crossover techniques related to the different kind of GA encoding scheme, binary coded GA, then real coded GA and order (Refer Time: 29:57), these are the 3 different GA techniques are very popular. So, we have learned all the crossover techniques and what is want to say is that crossover techniques is the most important and the significant operations out of all the operations are there, like in coding and selection, this is because the crossover we have to follow from the n P number of mating pools to create so many chromosome; that means, it is to be computed maximum.

Therefore, while we are choosing the crossover technique we have to choose that which takes the minimum time to compute because the overall efficiency of the geotechnics depends on how fast we can accomplish the crossover operations. So, these way crossover operations are very vital, 1 operations in case of GA algorithm and we have discuss the different operation techniques so far. Our next portion is basically the mutation will discuss in the next class.

Thank you.