Design and Optimization of Energy Systems

Prof. C. Balaji

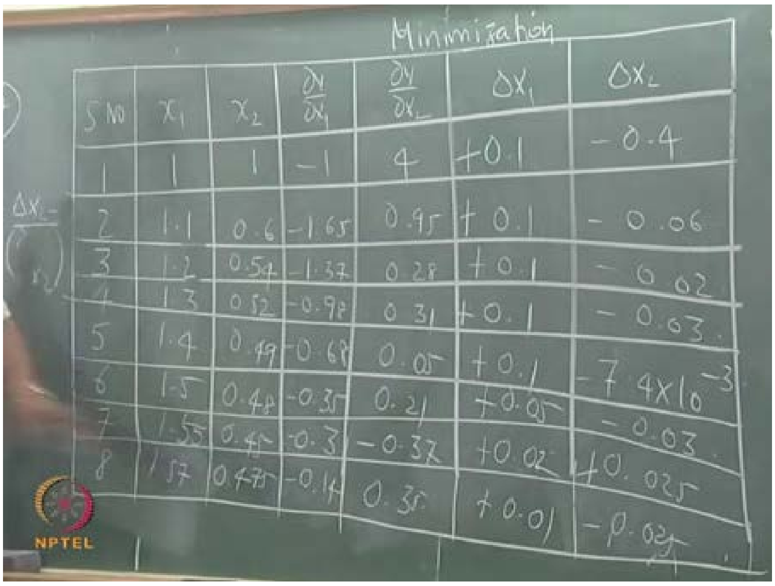
Dept of Mechanical Engineering

Indian Institute of Technology, Madras

Lecture No. # 34

The Conjugate gradient method

(Refer Slide Time: 00:13)



So, we will continue with this.

Student: minus 1.37, 0.28.

You still stick, you will say at point 1.

Student: Minus 0.2.

Minus.

Student: .02.

52? Yes.

Student: minus 0.98.

Good, minus 0.98.

Student: 0.31.

Student: minus 0.07.

What is the correct answer for X 2?

Student: 0.49.

Okay, 0.49.

But is it going in the right direction? The answer is it is fine know, okay.

Student: Minus 0.68.

Good minus 0.68, 0.05. What does it tell you? dy by dx 2 is 0.05. They are very close, but we cannot go home because this fellow has not become close to 0. We will stay with 0.1, 1.7 10 to the power of minus 4; what is it?

Student: 7.4 into 10 to the minus 3.

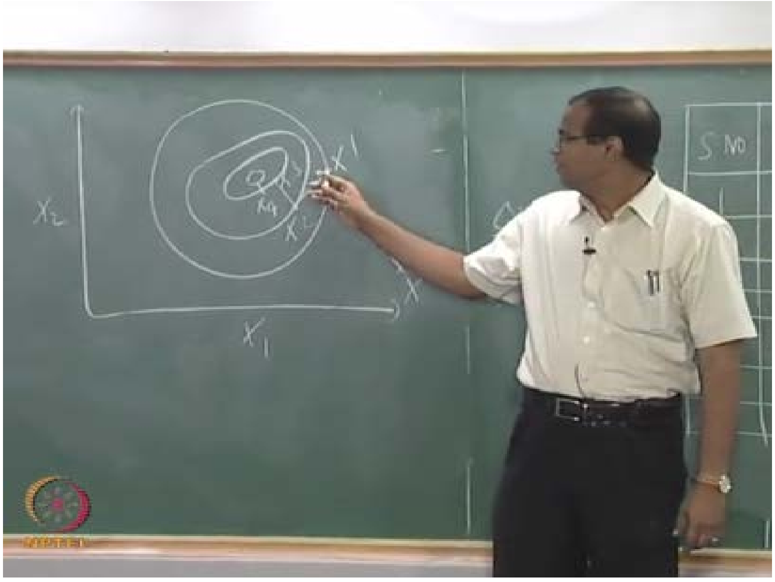
7.4 into 10 to the minus 3.

0.49. It is all minus know still? 48.

Student: 48.

See now it is becoming sluggish, right. 1 to 0.6 it was rapid, this also rapid. Now it became sluggish. The answer is 0.45. It becomes slow, that is the way it works, okay. When it is gets close then it goes like this, okay.

(Refer Slide Time: 02:31)



So, I think some people you must have got some ideas. Why are you directly going like this? If you know the two this thing why you cannot skip this and go straight?

Student: You have just beginned here sir what choice do you have?

Yeah, we have; we can do that. That is called the conjugate gradient method, I am coming to it. That is called the Davidon-Fletcher-Powell method. Our people have thought about it, okay. So, that is the advanced version of it 1.5, 0.48 and now?

Student: Minus 0.35, 0.21.

Oh, it is bad. No, 0.21it goes like mad, know. What happens if it is 0.1? Now I will make it 0.05 now.

Student: This is 0.1 that is minus 0.06.

Minus 0.06 means 0.42 it becomes. No, I will now do some cheating. So, I will make this 0.05. Yeah, I can change delta x 1, right, because I know that it is converging. Now tell me, point 0.

Student: 0.03.

0.03. How many of you are working out? Minus 0.03, okay, so 1.6, 0.45.

Student: 1.55.

Not yet. dy by dx 1 must be close to 0 now, no. How much is it?

Student: Minus 0.3.

This fellow?

Student: Minus 0.37.

Minus 0.37, wait, wait, wait, wait here. Now for the first time dy by dx 2 is changing sign, right. Please look at the board. For the first time dy by dx2 is changing sign. It is telling us that we have overshot the answer. What is the actual answer? 0.455 something.

Student: 0.46 also it is same.

Yeah, because very close to it, it will not be, so you still want to stick to 0.05?

0.05 you want to stick, okay. Now it will become so there will be a positive correction to it, correct. Then it becomes .51.

Student: Yeah.

No, I do not like it. So, I will make it 0.02. See that is why in order to adjust this, your delta x 1 will reduce. So, it will take more and more iterations. See you came very quickly to 1.5. Now, 1.55, 57, 58; 58 to reach 1.62 it will take a long time, 1.62 it will take a long time for you. What is this? We will stop with one more iteration.

Student: 0.25.

It is a positive correction, right. Yes. Now this will become 1.57, 0.475. We will stop with this iteration because all of you got a good idea of how it works, yeah. Now complete this dy by dx 1.

Student: Sir if you take as 0.2 so it will minus.

No, if it becomes minus, that means you have to change the delta X 1 because X 1, X 2 have to be.

That is all ok. X 1, X 2 should not become.

Student: Minus 0.14, 0.35.

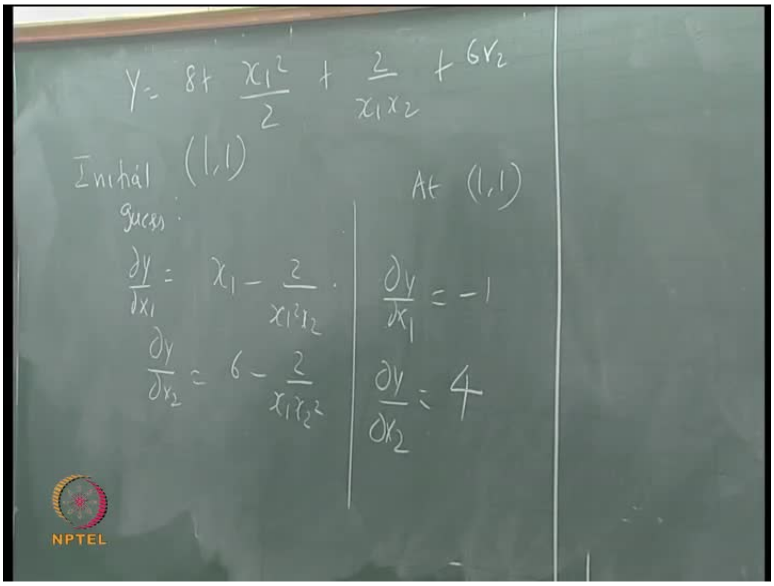
Now it rapidly oscillates, it oscillates too much. So, again I will make it 0.01.

Student: 0.025.

Yes, 025. It again becomes point negative, okay.

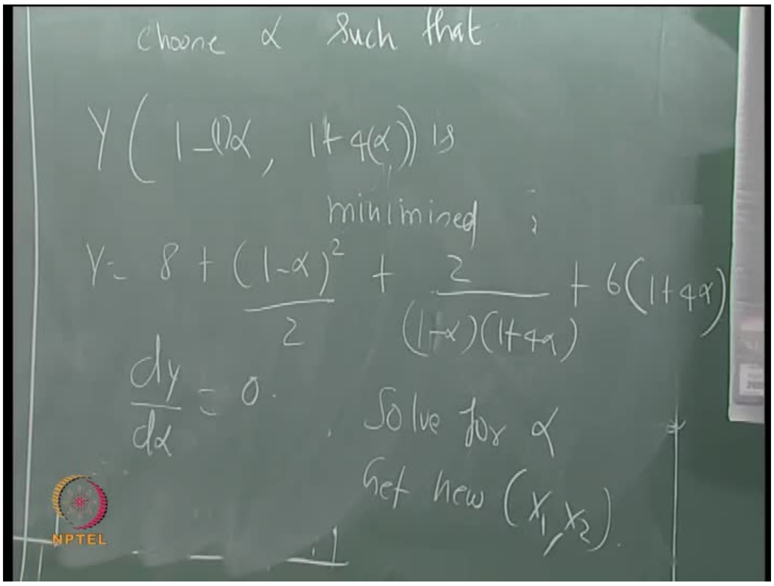
When it comes back to 0.45 it will keep on oscillating, this will slowly go to 1.64. It will take another eight iterations but this is the way it goes. Now it is going in the right direction, but it is going slowly towards the end. It will start off very fast, very impressive in the beginning and then it becomes very slow, okay. So, this is the Cauchy's method. I hope all of you got an idea of this. Now how to incorporate alpha and solve this problem? How to incorporate alpha and solve this problem?

(Refer Slide Time: 08:34)



Okay now at 1, 1, what is the value of dou y by dou x 1, minus 1, 4.

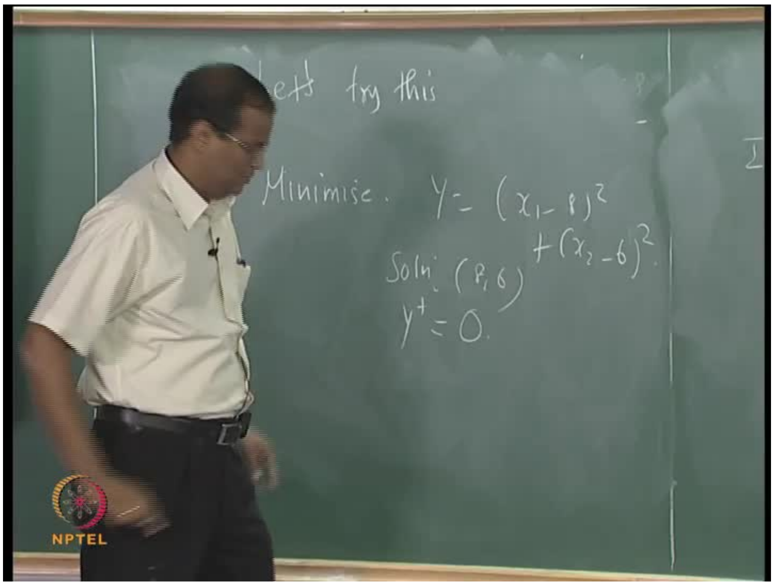
(Refer Slide Time: 09:42)



Now choose alpha such that delta X 1 is dy by dx 1 into alpha; delta X 2 is dy by dx 2 into alpha. So, X new will be X old plus delta X 1; that is alpha times dy by dx 1minus 1, alpha times dy by dx 2. Now I will say y equal to. Solve it for alpha. Once you get the value of alpha, you will get the values of new values of but see how difficult it is. Now getting alpha itself is a project now. You have to use golden section search but if you have patience, if you do it, within 2 iterations or 3 iterations you will get but what took you 15 iterations, okay. So, dy by d alpha equal to 0, solve for alpha. It is not appealing because you do not know; after getting this alpha whether we are better off compared to the.

So, we will try it on some problem where we know the final answer and at least one iteration we can calculate the value of alpha even though the problem may be silly, alright. Now you got the point? So, once you calculate alpha the new value of X 1 will be 1 minus alpha; the new value of X 2 will be 1 plus 4 alpha, you proceed. Get dy by dx1, dy by dx 2 are the new values of X 1, X 2. Get the new values. So, every dynamically, the alpha will be updated. You solve for alpha that is it basically. You convert it into a single variable optimization in alpha each and every iteration use the most sophisticated single variable search to get the value of alpha. This is the clinically correct Cauchy’s method; what I taught you before is not, is not the actual method. But it becomes very difficult; that is why I taught that first.

(Refer Slide Time: 14:00)

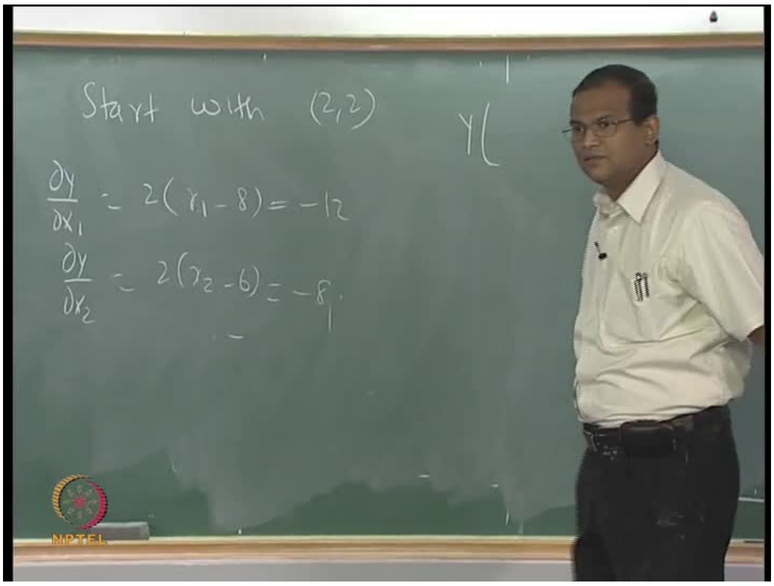


Okay let us try this. So, this is the Cauchy's method, okay. What I have told you is the Cauchy's method. If you look at any good optimization book, they should have explained Cauchy's method like this. Let us try this; we did this before X 1, X 2 greater than 0. This is a circle, right, this is a circle. What is the solution 8, 6 Y star Y plus are they optimum?

Student: 0.

Very good, so this solution we know. Now we will see how this fellow helps us, how the Cauchy's method will help us. You can start anywhere. Let us start with 2, 2.

(Refer Slide Time: 14:49)



Let us start with 2, 2. We can get dy by dx1, what is it? What are these fellows now? At the initial guess value.

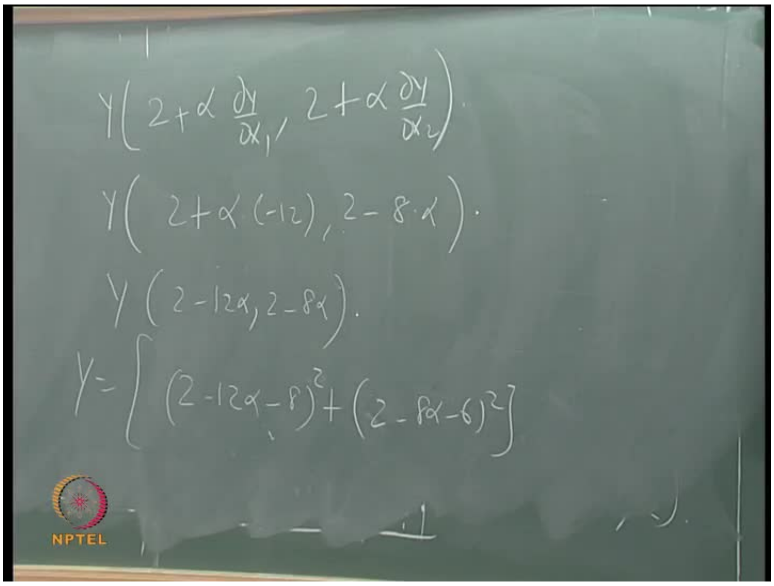
Student: Minus 12.

Minus 12

Student: Minus 8.

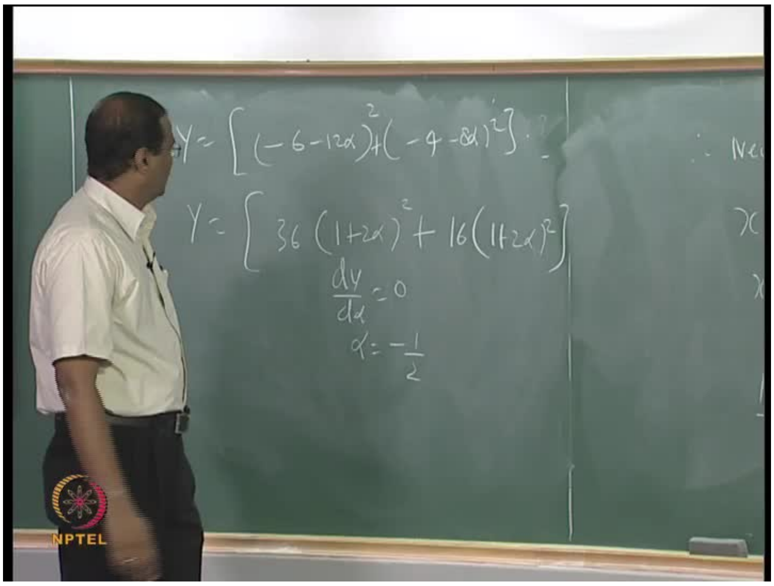
Okay, very good.

(Refer Slide Time: 15:36)



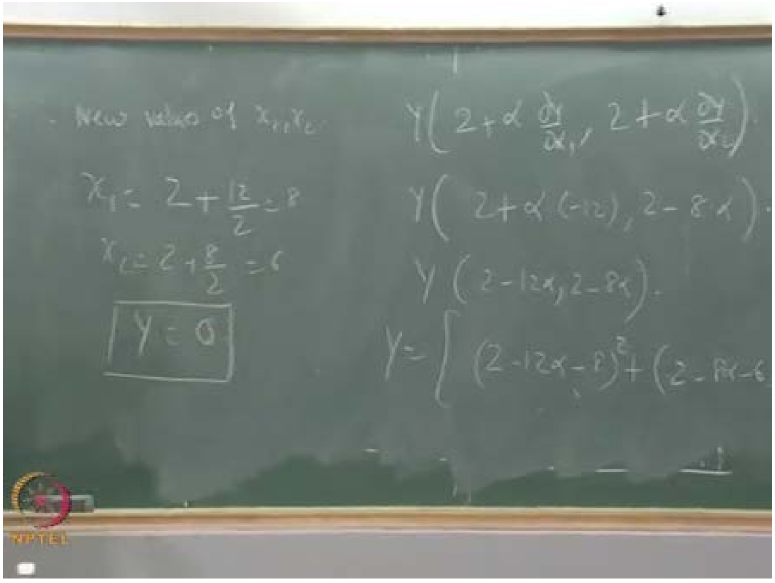
Now you have to calculate Y at the new points 2 plus alpha, correct. You have to calculate the Y and optimize, fine. So, what is this? Y of correct, Abhishek is it ok? Now you have to optimize for Y equal to. Now take dy by d alpha equal to 0 and get the value of alpha.

(Refer Slide Time: 16:49)



So, what is it now? Minus 10 minus 12, is it correct? Is that correct? So it is, correct. So 6, everything can be taken out, right. Do not worry about all these then dy by d alpha equal to 0, alpha equal to minus half. If alpha is equal to minus half, what are the new values of X? Straight, one hit, targeted, mission accomplished; straight you get 86, answer. You start anywhere; it will come because function is just a simple quadratic. So, I wanted you to believe that this works; that is why I chose this silly example, it works, right.

(Refer Slide Time: 18:23)



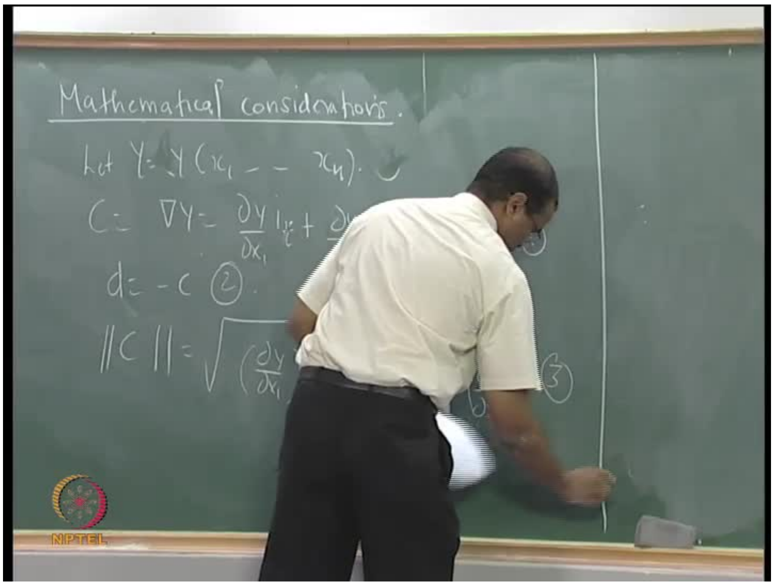
Therefore, what, what, what, is it correct? You got the answer but the previous problem getting alpha was very difficult. This is how it works but what is the logic behind all this? What this is orthogonal property and what is the vector calculus, what is the mathematical behind this; we will see that and then I will explain the conjugate gradient method and we will close. We are not solving any problem in conjugate gradient, okay. People who want to do your masters of PhD in Operations Research and Optimization you will do six or eight courses and somebody will teach you conjugate gradient method. People have worked in between 70s, 80s and 90s so many papers have been written, so many PhDs have been produced how to improve the Cauchy’s method, then came a Newton’s method.

In the Newton’s method, what is the difference between Cauchy’s method and Newton’s method? Why only the first derivative should be taken? I will take second derivative also; that is Newton’s method, okay. We will take the second derivative and make it quadratically convergent. People worked on; then somebody said modify Newton method Cauchy’s Newton method. Then somebody said I will achieve the performance of the Newton method but without taking recourse to the second derivatives I am coming up with the conjugate gradient method. So people say conjugate gradient method gives performance which is very similar to Newton's method, but second derivatives are not required. Like that people have developed, developed, developed and developed this and now it is nearly perfect many of these. Lot of optimization is there in chemical engineering, are you aware of this? Chemical engineers solve lot of these problems; lot of these NLP problems. What are these NLP problems?

Student: Non-linear programming.

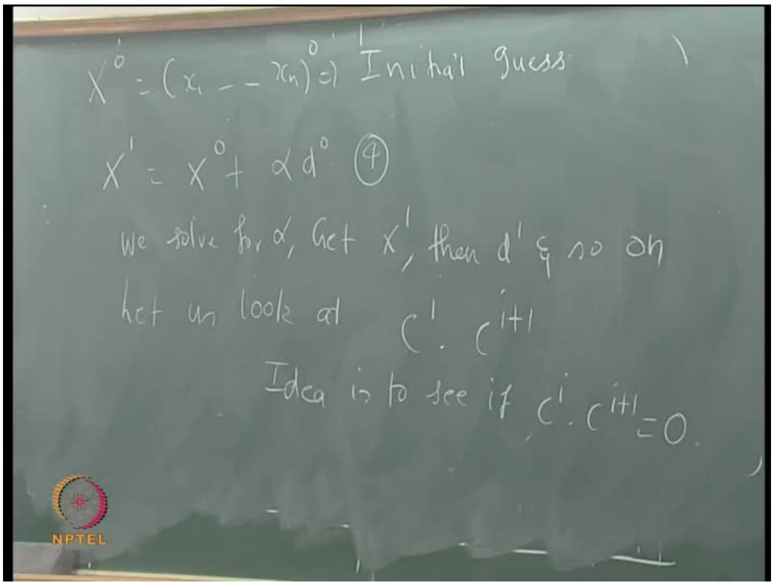
Non-linear Programming, so, now you are convinced that this alpha business works, right. Now we should, yeah any doubts? We require a proof of this, right. But why does this conjugate why does this steepest descent method work or where does it come from?

(Refer Slide Time: 21:21)



So, let us say mathematical consideration. So, we have first I gave the algorithm we solved a problem, we saw two variants, two variations and we solved problems in both. Now what is the mathematics behind it? Let Y be Y of x 1 to x n and C is delta Y, okay So, I said d is minus of C, you can put the equation numbers. You can call this as 1, this is 2. So, d is actually the minus of the gradient vector; d is the minus of the gradient vector. Now if you say what will be the modulus of the C vector? Root of dou Y. Yeah, so we got this, right.

(Refer Slide Time: 23:06)

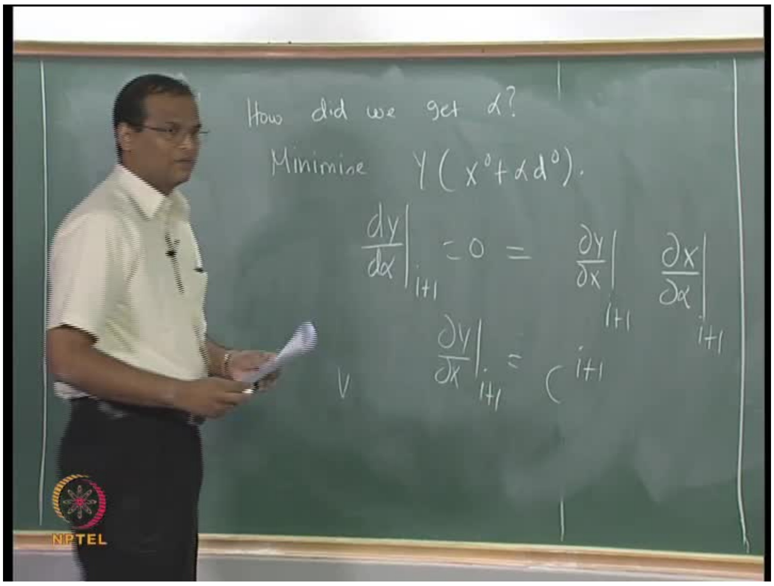


Now, X naught. What are we trying to do? X1, yeah can you tell me? If you tell me properly, then I will have the satisfaction that you have understood the algorithm. So, X 1 equal to X naught plus alpha times.

Student: Alpha times d naught.

Alpha times d naught, correct. Of course your treatment is for a minimization problem. That dy by dx 1 minus let us treat it for a minimization problem; otherwise, we can adjust alpha, do not worry about the sign, okay. So, this we can call it as 4. With this, this is how the algorithm works. So, this is the mathematical notation, I have written it in the mathematical notation; we have already solved the example. Now we are curious to look at let us look at C i. What is the product of C i and C i plus 1 where C itself is given by C is a del of Y. Rather I am trying to see whether C i into C i plus 1 equal to 0, so that the new direction is always orthogonal to the old direction. It has to be so, but now we will prove it formally, okay. We want to see, idea is to see, okay. How do you do that?

(Refer Slide Time: 25:41)



So, how did we get alpha? Minimized, correct. So, we are trying to do dy by d alpha Y plus 1 is equal to 0. dy by d alpha i plus 1 equal to dou Y by dou x at i plus 1 into dou x by dou x; one is a capital x one is a small x, agreed. So, d is equal to into dou X by, sorry dou X by dou alpha, dou X by dou alpha, okay. This x is the capital X, correct. What is dou Y by dou x at i plus 1?

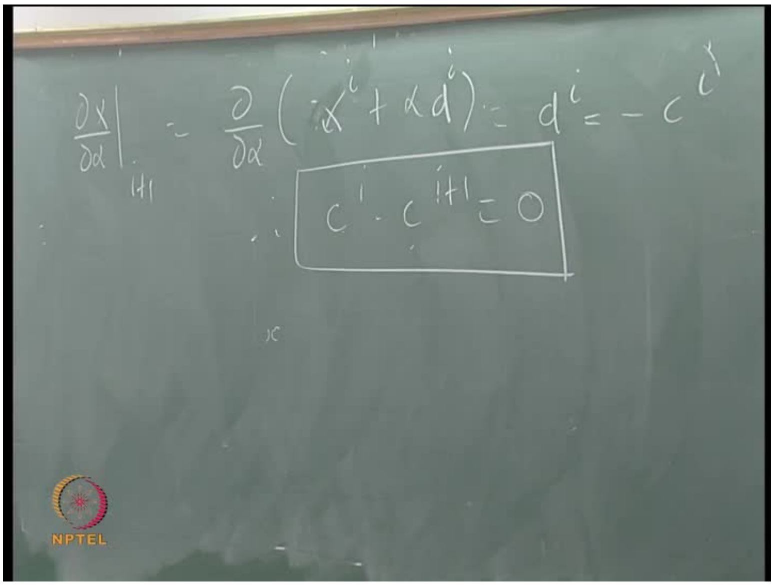
Student: C.

C. At what iteration?

Student: C i plus 1.

C i plus 1, okay.

(Refer Slide Time: 27:41)



What is dou X by dou alpha at i plus 1? Is equal to dou by dou alpha, I cannot put 0 I have to put i, right. For the general notation I have to put i. So, if you take derivative with respect to alpha, what do you get?

Student: d i.

d i, what is the relationship between d i and C i minus C i. So, if d y by d alpha has to be 0, dy by d alpha is itself a product of two partial derivatives dy by dx into dx by d alpha; dy by dx turns out to be C i plus 1 whereas dx by d alpha turns out to be C i. Therefore, C i into C i plus 1 is equal to 0. Therefore, at i plus 1th iteration the direction of movement is orthogonal to the previous direction QED, okay proved. Therefore, so the new direction of movement is orthogonal perpendicular to the old direction of movement. So, that is steepest descent.

Student: Can you repeat?

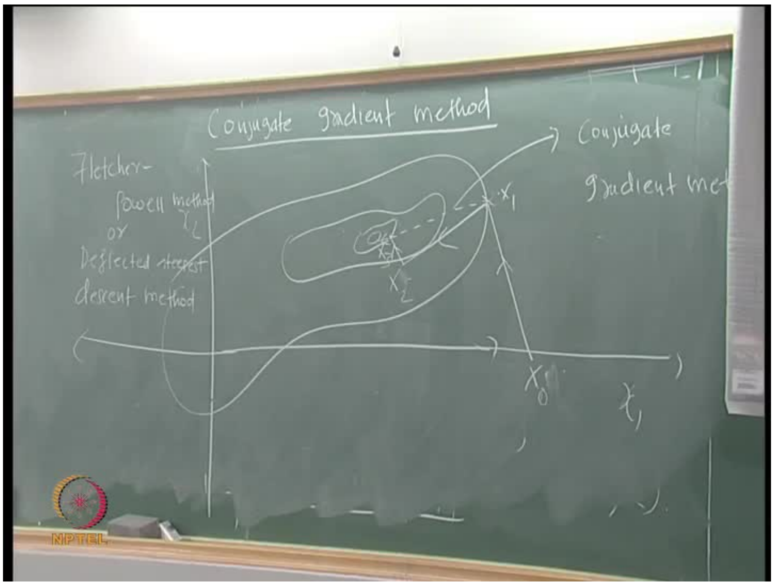
Which one?

Student: Last part.

Last part, so I did not do any cheating here; that is why I gave an example X 1 minus 8 whole square plus X 2; that is how you got alpha. You put dy by d alpha is equal to 0; I am writing dy by d alpha as a mixed derivative, I mean as a product of two derivatives dy by dx into dx by d alpha. dy by dx is nothing but C at i plus 1th iteration but if I take d x by d alpha, I can write X i plus 1 as X i plus alpha d I because that is the way you have written it here. If I take the derivative everything vanishes except this d. So, this d is nothing but minus C. Therefore, C i into C i plus 1, that is what we started out with. So, that is why I said we go like this, ok.

So, this is the proof of the steepest ascent method. So, last 5 minutes we will just wind up our discussion on this Cauchy's method by introducing a very interesting variant namely the conjugate gradient method. Conjugate gradient method can be used in your project; it can be used in your research. If you say I have solved a problem with conjugate gradient method, reviewer will not say anything. If you say I have solved it using C plus descent, he will say no good, okay. And you cannot get out of a 600 levels course without at least hearing about conjugative gradient, ok.

(Refer Slide Time: 30:40)



So, conjugate gradient method. A very crude form of conjugate gradient method we already saw the Levenberg-Marquardt algorithm; when we tried to do non-linear regression where there was a damping factor I introduced, right; that gives you a rudimentary idea of this conjugate gradient. So, it is better understood if you actually, that is okay. So, you start with X naught and hit X 1. So, this is small X 1, X 2, right. From here you go like this, X 2. What did I do X? Either use a subscript or superscript consistently, right. So, you started with X naught went to X 1. So, what are we doing here? What method?

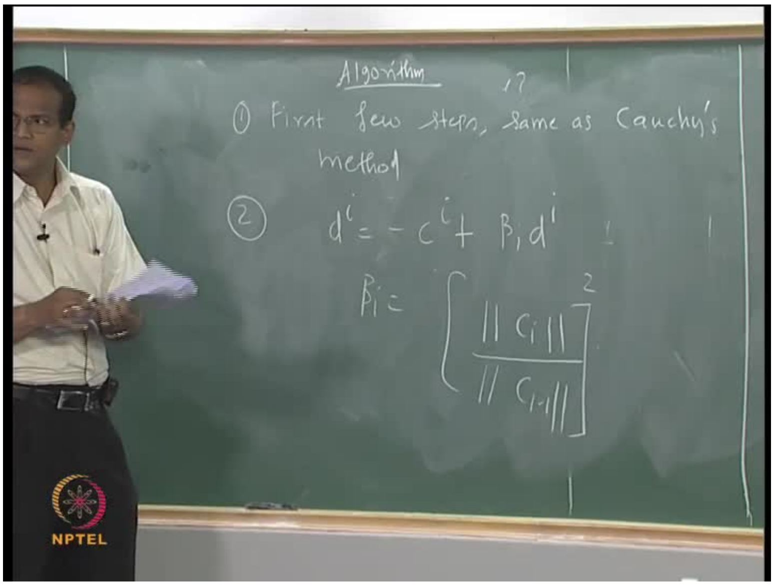
Student: Conjugate.

The Cauchy's method the steepest, now some fellows thought that from X naught, X 1, why go to intermediate step; from here if we do this, can go we go directly from X 1 to X 3 bypassing this? So, consistently if you do this, then you can jump some steps. Can you deflect the direction in which you are moving; instead of deflecting only in the orthogonal direction to the previous, so can you come up with a deflected steepest descent method? So, this deflected steepest descent method is called the conjugate gradient. Who will tell me what that deflection is we will work out the mathematics. It will be just 4 lines, but somebody has already thought about it 30 years back, okay. So, if you apply this funda you start from X naught first two steps are same. You start with X naught, go to X 1 using steepest descent and then X1 directly you go to X3. Each time you skip some steps.

So, this is the conjugate gradient method, it is also called the Fletcher-Powell method, deflected steepest descent method. It is one of the most powerful algorithms ever used. It was the most powerful algorithm, it has come to stay it has withstood the test of time. Even in circa 2010, papers are written with conjugate gradient. If you say conjugate gradient, people say it is ok acceptable, okay. So, because suppose you say, suppose I write a paper I have solved a heat transfer problem with genetic algorithm simulated annealing and I am getting this advantage, genetic algorithm is blah blah blah, it mimics the processes of nature, I write my story. Reviewer will say how does it compare with conjugate gradient method? Can you demonstrate with a simple problem that it is better than conjugate; in terms of memory requirement, in terms of the speed, how does it perform with respect to conjugate gradient method? Yes, it is some sort of a benchmark.

Then one Davidon also joined them, then there is a Davidon-Fletcher-Powell method called the DFP method; that is also very powerful, right. So, it is endless. Suppose somebody wants to work on the theory of optimization also, so much is there but invariably it will be either in a mathematical department or in an industry engineering setting; somebody who is very mathematically inclined who wants to probe further and further or he wants to do hybrid of evolutionary optimization and calculus based then you can, okay. So, then what they will ask you to do is all these algorithms have to be tested on certain standard optimization problems, okay. We will talk about this in a later class.

(Refer Slide Time: 35:35)



Now first few steps is so the algorithm. Now that you do not have any quizzes and you have plenty of holidays I would expect, I am not going to ask; I would expect at least half a dozen of you to code it in mat lab and check with a simple problem which you have done in class and you can show it to me. I will be glad to review it on a voluntary basis, okay. At least some curiosity should be there know. Some of you may not be interested, okay, 'oh, why did I take this course', but I think a few of you who are really mathematically inclined, this can be easily programmed and I have given you an example problem.

You do not have to search for example problem. You have seen how the steepest descent struggled, that problem you can apply. So, if somebody can do it and demonstrate in the class, we will allot some 5 or 10 minutes in one of the classes for one example problem, right. So, d i equal to, yeah please see this. d i minus 1, no, it is fine, okay. You calculate the C i but now I introduced the beta i. You calculate the length of del Y at that current iteration and the previous iteration; take those ratios, square it, that becomes your beta.

Student: Same notations. Notations are the same. So, minus a is nothing but a.

Yes, yes. Where is that Srinivasan? Yeah ok.

There is no problem, right.

Yeah, fine correct. Senthil what he is saying is fine. d i into 1 minus.

Student: 1 plus.

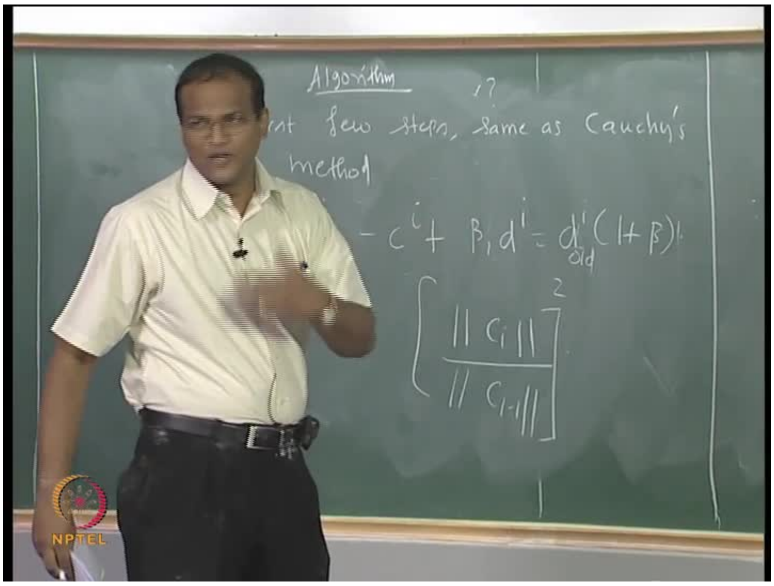
Okay, that is correct. You want that, should I do that? 1 plus, you are sure?

Student: Yes sir.

Okay. Please check whether it is 1 plus or 1 minus.

What happened?

(Refer Slide Time: 38:56)

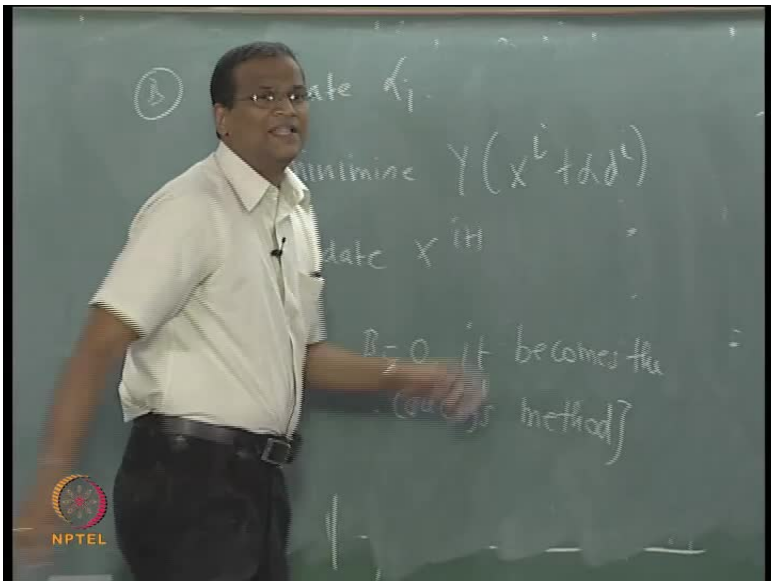


Student: Both sides d i new.

No, no both sides d i new or d i plus.

This is the way it is. Once you write a program you will write like this, okay.

(Refer Slide Time: 39:01)



Now evaluate if there are any changes required for maximization problem, please you can rework, right. Update i. If beta equal to 0, it becomes the Cauchy's method. There are rigorous mathematical proofs as to why the conjugate gradient method is quadratically convergent. I have already told you what is quadratically convergent, right. There are rigorous proofs available; people have written papers, they have demonstrated with test examples and so on. Conjugate gradient method is really one of the powerful methods you can see that, okay.

Student: Sir how did you get that value of beta sir? We did not get that.

What? That is the algorithm. I do not have time to prove that.

Student: What is d i old sir?

No, some people objected, if you leave it like this also it is all right. You can say X equal to X plus 3, right, in a program; you can say X is equal to X plus 3.

Student: Sir that is C i by C i minus 1.

Yes.

Student: Is that C i minus 1.

C i by C i minus 1, we are always looking at the ith step. If it is i plus 1, you can put what you are saying; be careful about that. See for each iterations to get the value of beta you require at two other places, i and i minus 1. You require two values of C to get beta, right, current and the previous. So, you can make it beta i plus 1, C i plus 1 and C i, whichever way but C has to be calculated at two steps. So, that is why you will start with the Cauchy's method and then only after two iterations, one or two iterations you can switch over to the conjugate gradient. Okay, I will take attendance. We will stop here.