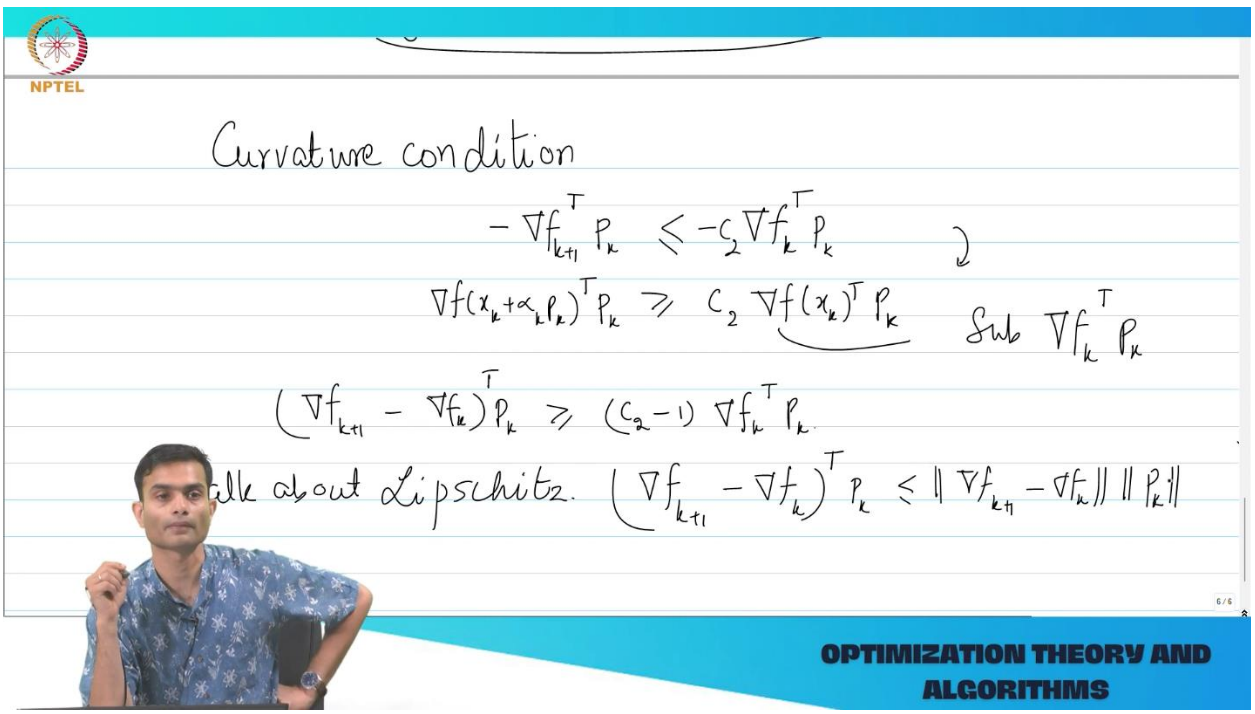
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 04
Lecture - 29

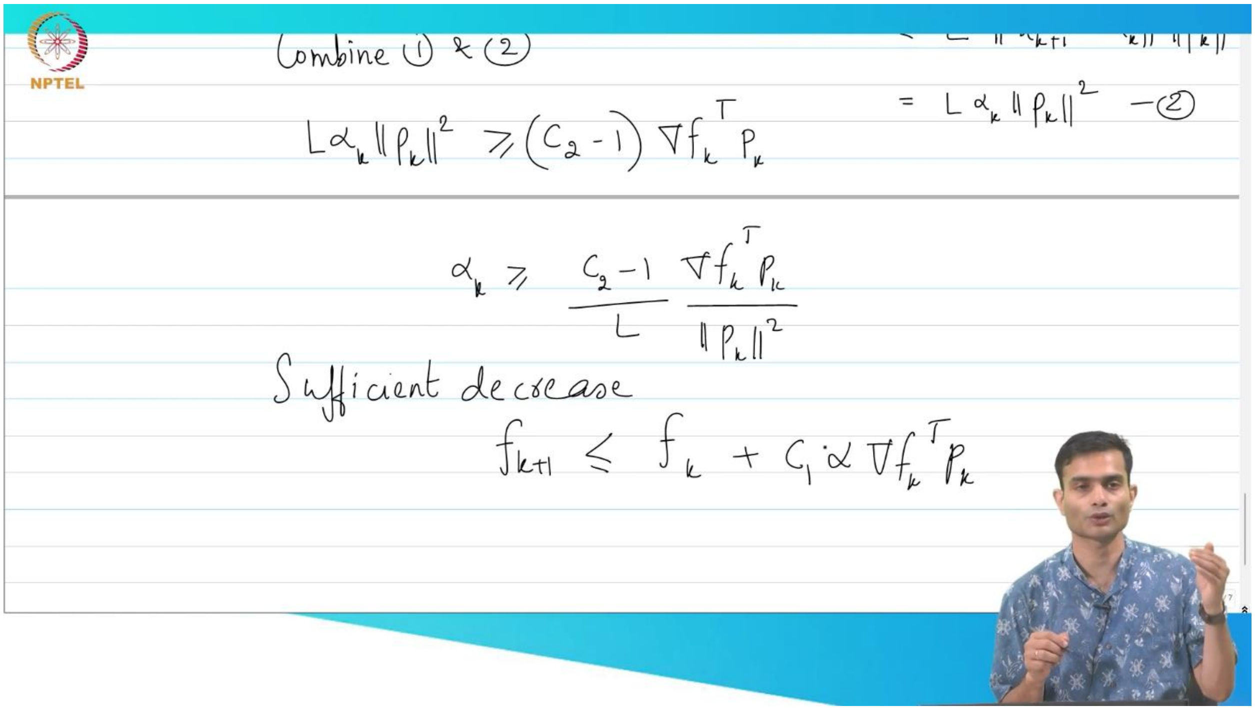
Line Search - Convergence and Rate - 2

So now when you look at this summation over here, I mean this is what we want to prove. So let us work towards it. Now one kind of basic hint is that whatever is stated in the prerequisites or the assumptions for the theorem, you will find that all of them get used. Without it, I mean it is not possible for you to prove it. So what are the things that we have in mind, what are the nuts and bolts that we have? Descent direction that means $\cos\theta > 0$ that is one thing that we will have to keep in mind. Wolfe conditions what are the two Wolfe conditions? Sufficient decrease and curvature conditions these then f is bounded from below these are the four ingredients of the recipe that we will have to kind of use ok.



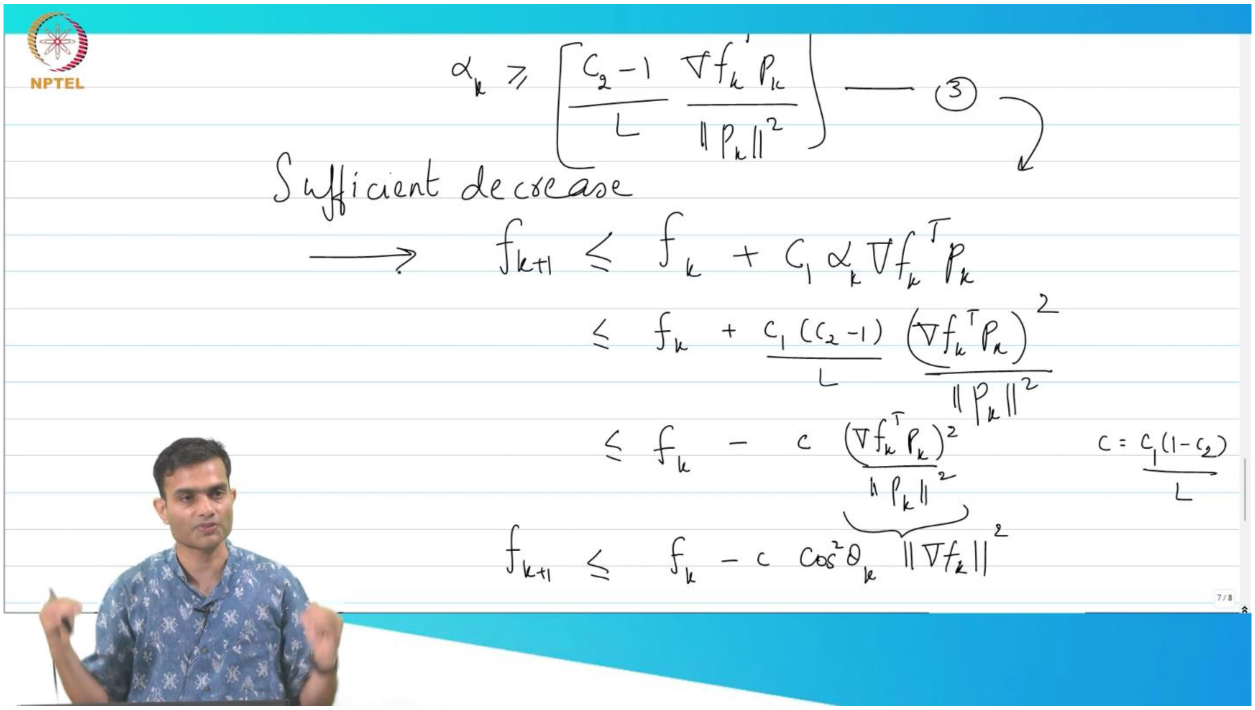
Notice that what I want to prove has a summation of $\| \nabla f \|$ right now has everyone heard of telescoping a series? So, when I telescope a series that is when I can add a lot of terms right. So, your hint is when you see a summation happening in over here that there is going to be some kind of a series which I telescope and I am going to get a summation that is going to give me some hint, but if you did not know that that is ok. Let us start with the curvature condition Yes, which is why it is very powerful not curvature condition.

So, what is the curvature condition telling us? again it is better if you recall the geometric flavor of it rather than the exact thing. What was it saying that the slope of the linear approximation is smaller than the slope at $\alpha = 0$, right? And the way we wrote it what was the slope at why minus because otherwise slope was a negative number I made it positive I said this was this should be less than or equal to the slope at 0 right. So, the slope at 0 was and we relaxed it a little bit right C_2 this is how we wrote. Now let us get rid of these minus signs. So, what will I have and let us open it up.



So, there is a minus sign I am multiplying on both sides. So, this which curvature condition versus the strong Wolfe condition or the regular Wolfe condition. In the strong Wolfe condition we put a mod, but there is no mod here. So, this is the weaker one, but it is still good enough. Now, what I would do is, let us subtract $\nabla f_k \trans p_k$ from both sides.

So this what will I get this is going to be. Now, let us see we want to use we also want to use this Lipschitz condition. This Lipschitz condition is telling us what that if I see $\| \nabla f \|$ differences I can replace it by $L \| \mathbf{x} - \mathbf{y} \|$ that is something that we want that we can potentially use. Let us see if I can retrace the steps. Ok Now let us just talk about Lipschitz over here. and to talk about Lipschitz I have to talk about the difference of two gradients right.



So, that I see there are two gradient expressions being subtracted over here right. So, let us write those down. So, I have ok. Can I relate this to the norms of both of these vectors and inequality in a product of **a** and **b** can I relate it to the norm of **a** multiplied by norm of **b** less than or equal to right this is less than or equal to $\| \nabla f_k + 1 - \nabla f_k \| \cdot \| \mathbf{p_k} \|$ why? cosines between.

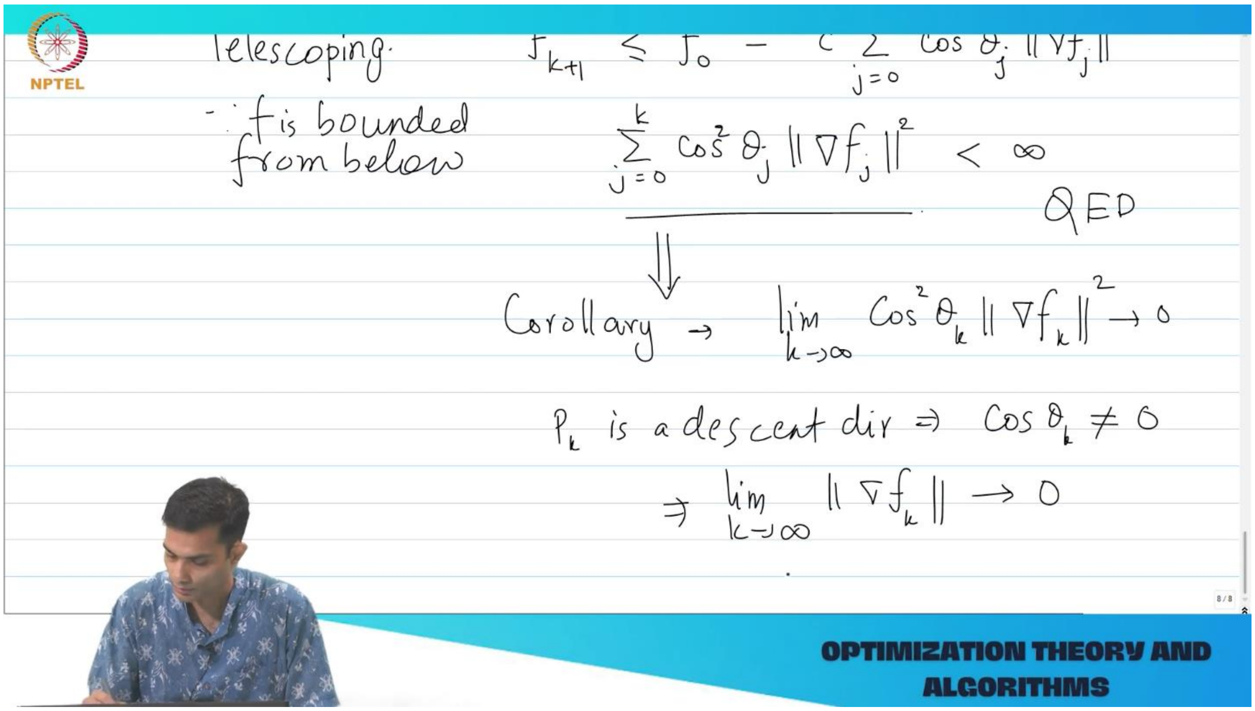
Cos has magnitude between -1 and 1 right. So, this the term on the right is always going to be greater right. I got this difference of gradients over here right. Initially can I apply Lipschitz? Lipschitz is saying that this $\| \nabla f_k + 1 - \nabla f_k \|$ should be what? Less than equal to $L \cdot \| \mathbf{x_k} + \mathbf{1} - \mathbf{x_k} \| \cdot \mathbf{p_k}$. Is there something nice I can say about $\mathbf{x_k} + \mathbf{1} - \mathbf{x_k}$? $\alpha_k \mathbf{p_k}$, right.

So, this is equal to $L \cdot \alpha_k$ which is a positive number. Therefore, I have $\| \mathbf{p_k} \|^2$, right. So let us call this 1, let us call this 2. So can I combine can I combine 1 and 2? this term over here is greater than something, but this same term over here is less than something. So, can I combine these two inequalities here? So, what am I going to get? I am going to get that

So this is actually giving us a very interesting inequality on α_k. α_k we thought we could take any really small number, but what this is saying is α_k is actually like this. So we have used curvature condition, we have used the Lipschitz condition. What have we not used? bounded from below we have not used and sufficient decrease we have not used. So, of the Wolfe conditions we have used only one condition which is curvature condition what is left to use is the sufficient decrease.

Now sufficient decrease simply says that the function value at $k + 1$ should be what? Should lie below the linear approximation of the function at k. So this is what it would be $f_k + \alpha \nabla f_k \trans p_k$ and when I throw in the relaxation what will happen? C_1 right. So, always remember these conditions just by their English statement it is you will you will you will not go off right function value below the linear approximation curvature is slope at that point is less than the slope at the starting point right that is what gets us this. Now, actually this is the guy where we can use the idea of telescoping a series you see that because I see $f_k + 1$ and f_k on either side now if I write this out from f_1, f_0, f_2, f_1 like this over here what will happen I can cancel off these guys that is going to be that is going to be the intuition over here and the term that is left the second term on the right hand side is what is going to give me the summation but we are let us let us get there I have α I have an inequality here on α_k right can I substitute this in here Will the inequality still hold? I want to use this is condition 3. Can I substitute it in here? Yes, right.

So, this is going to be ok. α is greater than this yeah. So, this whole expression is even greater. No why? So, I here is my α_k and I if I substitute this k over here I mean I have this will continue to hold this way right. Yes, but this implies α_k is less than that.



What is the smallest value α can take? This guy, right? This is the smallest value. And α can take any other value. Does anyone else have a problem here? Did we make a mistake in the third inequality then? it looks correct to me right I when I combine 1 and 2 this is exactly what I got and I got this actually one quick question this term on the right hand side of 3 is it a positive number or a negative number C_2 is less than 1 that is ok that is not enough right if the overall term is positive because it is a descent direction. So, is actually a negative number right. So, this is telling us α_k is above some positive number right.

So, is this fine? The direction of the inequality is fine over So, one common thing that is done is that this C_2 as we just observed that C_2 is actually less than 1 right. So, what people will do is we will write this as like this. We just bunch up all of those constants into one number C. $C = \frac{C_1(1-C_2)}{L}$ right. Now, we have a better sense of what we are doing. So, when I do that what happens to α? My α is greater than or equal to C is still positive number and since α is increasing then since I have these inequalities like this that implies that the left-hand side must converge.

Now, I am essentially telescoping and my sum is bounded. I want to make sure that I know the index terms right. So I will keep using k as my index for the summation that I will keep building up. So, here is where I can use it to show my summation will converge. Right now what I can do is I can actually take $\| \nabla f_k \|$ on this term this will get normalized into that right. Now this telescoping series is just going to make that very explicit. So, when I add all these together, it is like adding all these terms from 0 to k for the final step. This is the intuition. But remember, so we have used curvature and Lipschitz. The way we have done this, we have not done it in the right order. We could have done the sufficient decrease before the Lipschitz condition.

So, once I do that, I will have the result that I am looking for. Because $\| \nabla f_k \|^2$ will continue to keep diminishing. And since it is bounded from below, this is telling me the overall sum will converge and is going to be bounded.