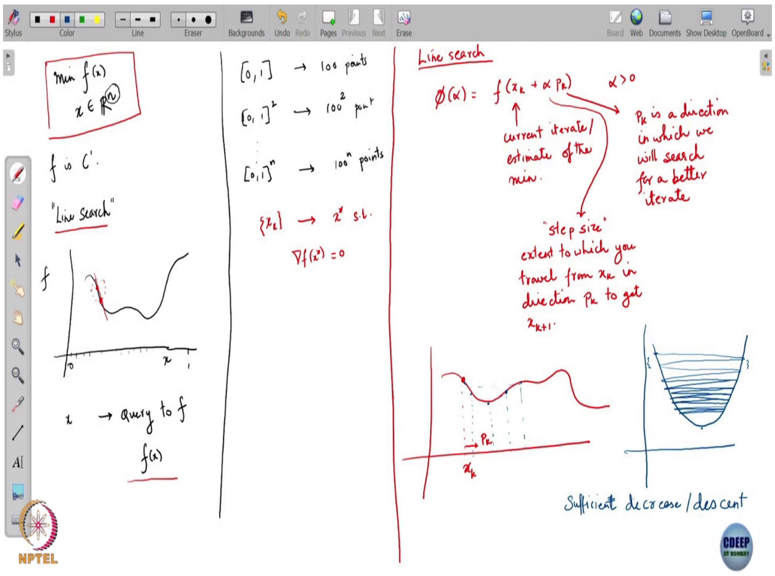
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 19B
Wolfe conditions

(Refer Slide Time: 00:16)



So, there are couple of things you need to be mindful of in this whole in this whole of effect. The first is you know you do not want to make you do not want your decrease to be too little. Because say for example, if you are decreases too little say let me give you an example like this.

So, suppose I draw a function like this; suppose here is your function and here is your iterate you take a very large step, say and you go here. But then the decrease is very little the decrease is just decrease in the function value is just this much.

Then you take another step looks like a very large step. Again the decrease is not much the decrease is just this then decrease this, then decrease is this, then the decrease is this, this. What is happening here? What is happening here is that although you are decreasing your function value at each step, your decrease is probably becoming progressively lesser.

And now it depends on the kind of iterate that you have designed it could very well happen that your decrease by doing this, you in fact your decrease becomes so insignificant; that you end up sort of stagnating at some so at this sort of level without actually reaching to the minimum here.

So this so the, first thing you and one looks for when you want to choose the right sort of alpha in a line search. By the way this is called line search because what we are doing is a unidimensional search along the direction, along the line or along the direction P k yeah.

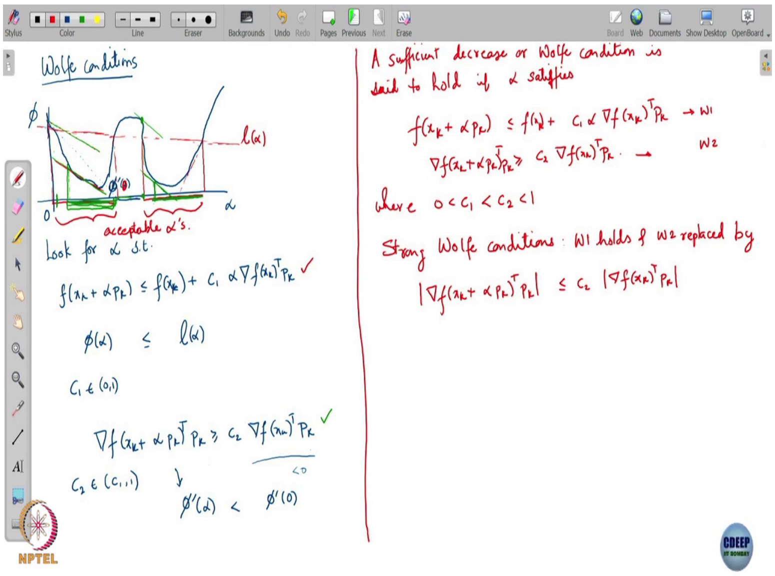
So, the first thing one looks for when one does a line search is what is called sufficient decrease or sufficient descent. So the, that means, that when we are doing line search the function must every step should actually give us a sufficient amount of descent.

Now how much is sufficient? So, is 10 to the minus 2 sufficient, is 10 to the minus 3 sufficient, is 10 sufficient is 1 sufficient; so there is no there is no we cannot put a number on what is sufficient what is what is sufficient descent. So, what we can do is we can try to define sufficient descent, in terms of the in terms of the function in terms of properties of the function itself ok.

So, there is an entire set of there is a there is a there is a there is a way of defining what amounts to sufficient descent, which I will tell you now. Which basically tells you what sort of alpha ranges are good for you in terms of the in terms of the properties of the function itself ok.

So that so, this leads us to what conditions that are called Wolfe conditions. So, this is these are not the only known conditions, but these are very popular. So, hence I am hence I am mentioning them, so suppose once again let me draw a function like that.

(Refer Slide Time: 03:55)



So, this I what I will plot now is just phi and here is my here is my alpha ok, what are suppose we ok. So this is suppose my make this a little smoother at this point, this is my alpha. Now remember we do not know the value here ok, we do not have a picture of this kind of a function here. What we can try to do? Is as I said feel our way through ok, so let me also make this a little.

So, one popular condition which is which are called Wolfe condition they impose the following. So, you look for an alpha, so you look for the condition basically says look for alpha equal to alpha k, such that or rather look for alpha such that.

We look for an alpha such that this the quantity on the left here which is, What is the quantity on the left? Well the quantity on the left is simply phi of alpha. So, we want phi of alpha to be less than equal to the quantity on the right. Now what is this quantity on the right? Let us look at this as a function of alpha; this is actually a linear function in alpha, so let us call this l of alpha.

So, remember we are fixing a k fix fixed for a fixed k we are looking for an alpha, such that this inequality hold and let us see what this inequalities ah. By the way what is C 1 here C 1 is just some C 1 is a constant is some constant between 0 and 1. Usually it is taken to be something like 10 to the minus 2 or some something like that. So, what is, so C 1 is some constant between 0 and 1 and now for that fixed constant we are looking for an alpha that satisfies this. The right hand side is a linear function of alpha is a linear function of alpha.

So, if C 1 was 1 what would it what would this look like? Well this l of alpha then would try to track in that case, the slope in that case would try to track the slope would be this sort of line. It could be a linear approximation to it would be the linear approximation to phi of alpha.

However because so, this is this slope here is phi dash the slope here is phi dash alpha. However, because C is between 0 and 1 what this does is it actually flattens the slope it tends to make it a little more horizontal, so l of alpha actually looks more like this.

So, this is l of alpha this line here the red line that I have drawn that line is actually l of alpha. And what is this condition saying, the condition is saying that we should look for an alpha such that the function phi lies below l of alpha. So, the acceptable alpha according to this condition, then become this interval and this interval these become the acceptable values of alpha.

Now, if I just simply, so if I impose this condition effectively what it is saying is well, I am going to accept any alpha once if my if, the way I impose this condition is it says that I am going to accept my alpha once this inequality here once the Wolfe condition here is satisfied. Now the, what this does is it does one good thing it prevents regions like these, it is prevents regions like this green region.

If you look at this green region what is happened here is that the in this green region, what is happened is that the function has as it you have come to the point, where the function is now has is certainly increasing and it makes no sense to pick that large an alpha right. So, in a in other words where if the, so what it does is it at least make sure that you are not you are not taking too large an alpha.

You are not taking two larger steps, so that you over; so that to such a large step that you your function then starts going into an increasing regime. So, in the in that sense it is a it this is actually useful that it prevents this green region.

However it does not also it does still does not guarantee that you will be taking where you can you will not be taking very small steps. Like for instance if you look here this you know you these little steps that are close to alpha equal to 0, these are also these are also still acceptable values of alpha.

So, if it could well be that these you know your algorithm will come back to you with you know if based on the way you are actually searching for alpha. It may well be that it will come back to you with an alpha that is too small, say some alpha here; this green point that I am right.

So, to prevent to make sure your alpha is not too large, but at the same times not too small you actually you need another condition. So, another condition which is imposed in this sort of wave, so this condition says f of x k plus alpha P k transposed P k is greater than equal to C 2 transpose f of x k transposed P k. Where now C 2 is another constant that is between C 1 and 1 and C 1 is the constant that we have used above.

Now, what is the meaning of this? Why does this make sense? So, the left hand side what is the left hand side here. So, this condition is effectively let us try to first intuitively understand what this condition is doing, this condition is trying to prevent these kind of this sort of extremely small steps.

So, the way to prevent these extremely small steps is what by the way it is doing it is by saying I would like my curvature to be sufficient, I want to get to a point, where the curvature is sufficient ok. Now, why does how is this condition capturing curvature.

Well what it is doing is? It is you look at what is the left hand side here? The left hand side is actually, the left hand side here is the derivative of phi at alpha. And what is the right hand side the anybody? Yeah. So, what is the right hand side? Well the right hand side is the derivative of phi at 0.

So, what this condition is actually doing is that it is asking that the derivative of phi at alpha should be at least C 2 times the derivative the initial derivative which is phi dash of 0, so alpha equal to 0 is this point. So, what its basically saying is that the curvature that you get for the acceptable alpha should be such where the curvature is at least C 2 times the initial curvature or the initial or the slope should be at least C 2 times the initial slope.

Now, why does this make sense? Well, so suppose if phi dash alpha was negative was suppose phi dash alpha was negative ok. Now if phi dash or if phi dash alpha is negative then this condition will not if phi dash alpha is you know say strongly negative; that means, it is in that case phi dash alpha is going to be less than phi dash 0.

If it is strongly negative means it is much less than phi dash 0 then in that case this condition would not be satisfied and it would, you would be looking for an alpha that you would be looking for an even greater alpha. So, when phi dash alpha is strongly negative it effectively is alluding to the fact that you could decrease the function even further, because your phi dash alpha is negative.

If and if you are if the function can be decreased even further then you mean you know it makes no sense for you to settle for the alpha that you have already, you would want to search even further and get an even better decrease right. So, when phi dash alpha is less than phi dash 0 this condition will not be satisfied.

So, it would ask you to increase it would ask you to you know search further to get an even better alpha, because there is scope for an even better alpha right. So, on the other hand if phi dash if phi dash alpha is positive, one usually has that phi dash alpha is phi dash of 0; that means, this initial one this one is usually negative right.

Because you are getting to you are choosing a direction in which your you are function decreases. So, this quantity is usually negative and. So, if phi dash alpha itself is positive then this condition is satisfied and or even if mildly negative it is condition would be satisfied because it is a it is pointing to the pointing to the case where you cannot decrease your function any further. And you have potentially reached your sort of the amount of decrease you could get in this direction the optimal amount of decrease you could get in this direction.

And hence it is now this time to stop searching and freeze the alpha right. So, geometrically the, what is happening here is that well. So, let us say this here was this here was phi dash at 0, if that is your phi dash at 0 then what we are good. What we are doing is again you are multiplying this by C 2, so that will make that will flatten the slope a little bit further. So, is we are, so you are looking for a slope that is a little flatter than that say of slope like this.

And so the kind of, so the desired slope looks a little bit like this maybe I should use a different color here; so this is what the desired slope looks like. So, the as a result what happen what has happened is from the acceptable alpha range has now become from all the way from here, till actually this sort of point till this point. But then if you also impose the first condition if you impose both of these together, the acceptable range then becomes something like this intermediate range right; and from here it goes from here till here yeah.

So, the green range that I have that I have drawn here that becomes then the acceptable range of alpha. So, as you can see why this is the reason all this is complicated is because we are we have to sort of feel our way through and choose the right alpha by based on some sort of you know educated estimates or educated guesses about how the function is going to behave in the you know as we change the alpha right.

So, coming put to summarize we have for these are what are called these are the Wolfe conditions. So, a sufficient decrease or and so a sufficient decrease or Wolfe condition is said to hold if alpha, satisfies these two equations where C 1 and C 2 bit have this kind of inequality between them.

Now, C 1 and C 2 are constants that are chosen that are fixed, they you do not vary them across iterations alright. So, these are the this is yeah, so this now choosing it takes usually a bit of trial and error to figure out what the right values of C 1 and C 2 are. But any C 1 and C 2 between that satisfy this are sufficient are alright for us.

Now, there is a stronger version of these of these Wolfe conditions in which the second one what are called strong Wolfe conditions, the stronger Wolfe; the strong Wolfe in the strong Wolfe conditions you change this here is replaced by. So, let us call these equations W 1 and W 2, W 1 holds and W 2 replaced by the following, which is you replace it by the absolute value instead of taking the gradient, oh sorry I missed the transpose P k here yeah.

Instead of simply taking the derivative you actually take the absolute value of it less than equal to C 2 alright. So now, in this case what happens is under the strong Wolfe conditions effect you the strong Wolfe conditions will if will are stronger in the sense that, you are not we are all we are not; the difference is that we no longer allow the derivative to be to be too positive right.

So, we do not want we, so the derivative of this the derivative at alpha of phi at alpha k should not be too positive, so we do not allow for instance slopes, where the yeah. So, say for example, a slope like this where the derivative is now is too positive we do not allow those sort of things. So, that even further constraints your further constraints your choice of alphas.