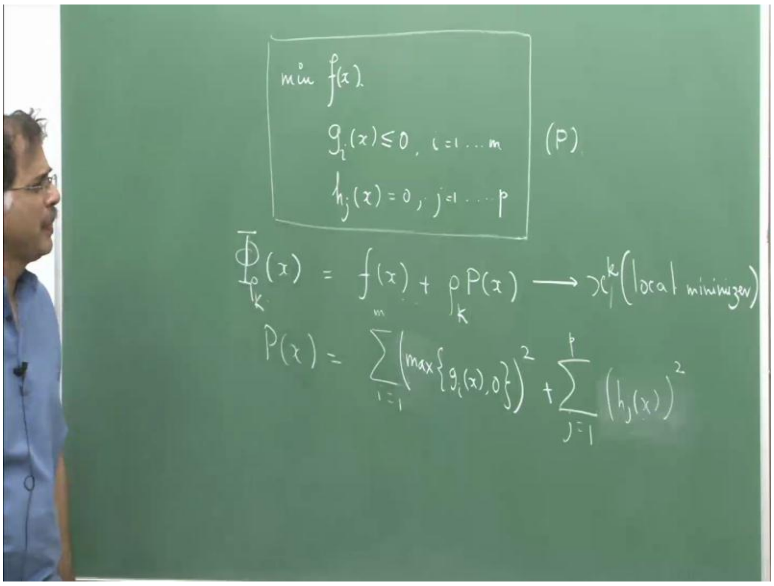
Foundation of Optimazation
Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 30

In the last lecture, we were concentrating on a particular problem, that is, a very standard inequality, inequality constraint optimization problem.

(Refer Slide Time: 00:20)



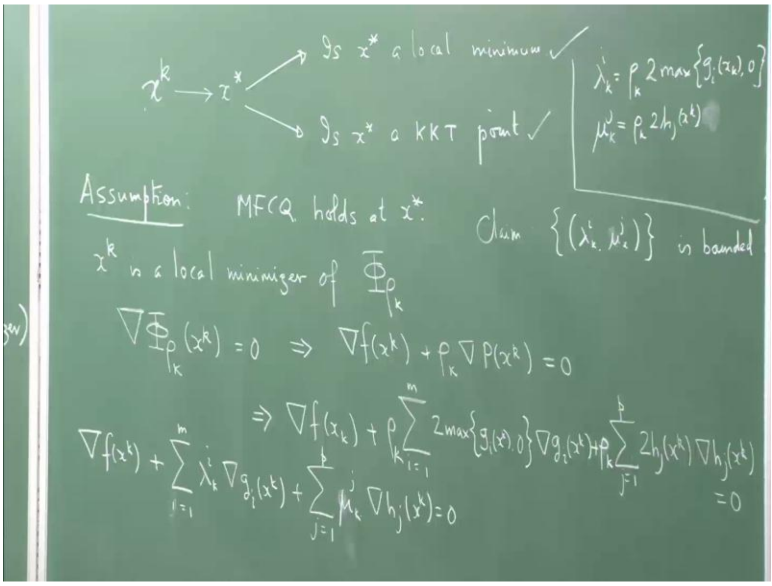
And, we devised a penalty function for this, and we computed the gradient of the penalty function, which is essentially a part of it. What we have learned before was a, if you can find a global minimum of these penalized problems, then that sequence of global minimizers would go and hit the global minimizers of the original problem provided the original problem has a global minimizer. I want to warn you and again remind you that there could be problems, which would have local minimizers, but not have global minimizers.

Now, what I want to stress here is that in general, when you look into an unconstrained problem; say in this case, our problem would be… Our penalized problem would be like this P of x as we have discussed in the last lecture… So, P of x is now this… I would just… I am sure you would remember most of them; maybe this could be k or P or whatever. So, do not bother about the index number in depth; just some finite number of equality constraints and finite number of inequality constraints; that is it. So, this P x is what we had here.

Now, see when you have the… If you have even smooth functions continuously differentiable functions; this and this that this is a differentiable function. So, if you had this, this becomes a differential function provided f is so. But it is not so easy to find a global minimize; even if you run certain algorithms, even if you make an application of direct search method about which you will talk very soon, you cannot really get the global minimizers so easily. Finding a global minimize is a very, very difficult thing and is a topic of very (( )) of current frontier research in optimization. Now of course, if all these are convex, then of course, this is true.

But convexity is something we will not assume here at this moment; that is, if you have all these f, g i’s an h j’s; f to be convex, g j's to be convex, this to be (( )). Then of course, this would be a convex function, which is differentiable and whatever solution you get of the gradient can be considered as a global minimize. Now, in general, if I am trying to solve this problem, what I will obtain is… Suppose I take rho k. And, for each k, I am… k equal to 1, 2… That goes on. I am evaluating the local minimizer, which I term as x k.

(Refer Slide Time: 04:34)



And, now, let me assume that the sequence of local minimizer actually converges and converges to some point. So, let me assume that x k goes to some x star; maybe you can consider x k as the bounded sequence and the whole sequence need not go to x star. And, there are subsequence of x k, which goes to x star. But that does not matter. So, without loss of generality, we can always assume that x k goes to x star. What sort of property you expect out of x star? Is x star a local minimum? A question is – if you cannot really confirm this, and the next best thing that you can confirm is – is x star a KKT point – Karush-Kuhn-Tucker point or a fridge-on point with a normal fridge-on multiplier. So, these are the two things; so this is what we expect. And this is what you would accept. So, there is something called expectation and something called your acceptance. In real practice, we would just even accept this one.

Now, let me tell you, it does not come for free that… you cannot just say that, whenever x k goes to x star, I would freely have this condition, freely have this fact. This fact you have to forget for the time being. This fact also you cannot say that, you will freely have it. So, what sort of condition you need to satisfy at x star? So, our fundamental assumption is MFCQ holds at x star; there is a Mangasarian-Fromovitz constraint qualification holds that x star. So, if this holds that x star, then you can guarantee that, this is the KKT point. So, what are we now supposed to do? We are supposed to show mathematically that, if this happens, then the sequence of local minimizers – these are all local minimizers – local minimizers of this problem. This is an unconstrained problem.

Here you see the attempt is always to convert a constrained problem into an unconstrained problem and then proceed. Of course, you can say if you have an abstract constraint, then this will also remain a constrained problem, but as a slightly lesser degree of work with a… But that does not actually take off the issue. The issue is to make it an unconstrained problem. But in real applications, you would basically have this sort of thing. Sometimes you have to look in to this inclusion. But, in many cases in engineering applications, this x element of u, this x would be nothing but bounds on the variables say x i between a i and b i. So, you can again put them into equality-inequality form and work with it. Theoretically, you can do it. But when it comes to actual practice, you really (( )) an unconstrained problem; and, that is something we need to keep in mind. And, that is why I have taken this problem as our general problem, which maybe you should write as P.

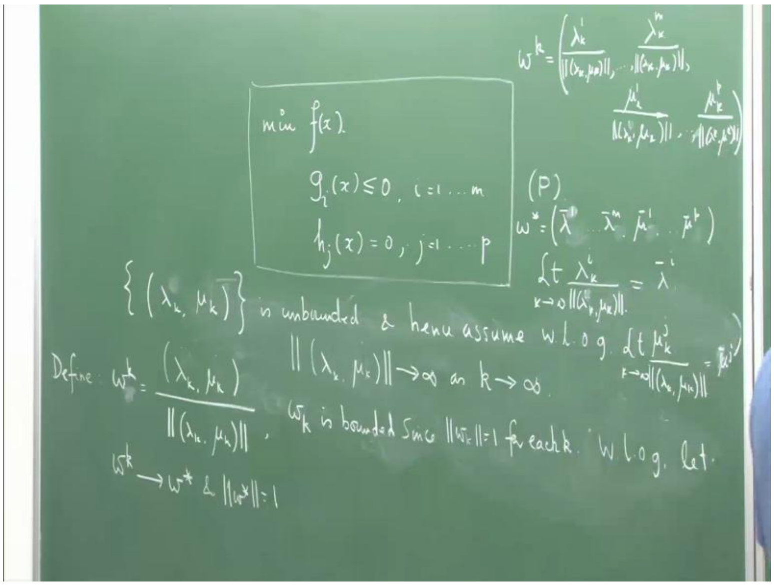
Now, if x k is a local minimizer of this problem, then from the very basic optimality condition, we know that this fact is known to us. So, what would that imply in our particular case? In our particular case, it would imply grad of f of x k plus rho k is a constant grad of P of x k. And, that so this is nothing but this and this is 0. Now, what is the grad of P of x k – we know already from yesterday's descriptions. So, x k in general, is not feasible. So, it will basically become g i x k in most cases. Here also it is not feasible. So, it will become this. And, this whole thing must be equal to 0. We have already computed the derivative of this yesterday – derivative of this part. And, this part you can compute easily; just a chain rule. So, I did not compute it.

Now, what is important is to do some little bit of manipulation. Once we do the manipulation, we have less boards to write. And that is the whole point. Maybe I will… Let me do a little bit of manipulation and write from here. Now, I am coming here and I am writing grad of f x k plus summation lambda k i grad g i x k; i is equal to 1 to m plus summation j is equal to 1 to p mu j k grad h j x k equal to 0. Of course, you know that how did I get this lambda i k and how did I get this mu j k. So, you can obviously note this fact that lambda i k is nothing but rho of k, is nothing but rho of k… Here it has to be rho of k – into this thing; that is, rho of k into twice max of g i x k and 0. And, this one is this one into rho k.

Here also have the symbol rho k. Excuse me for the slips. Here of course, mu j k is given as rho k times twice h j x k. So, this part is the rho k into this is mu j. So, what we have to prove; we have to prove is that, these sequence of vectors are bounded; that is, this sequence; they form this sequence; claim is following; this sequence is bounded. Actually, what we will show that, if it is not bounded, then the Mangasarian-Fromovitz constraint qualification will be violated. In its equality in its more functional form, in its equality form, it will be violated actually. So, you see Mangasarian-Fromovitz constraint qualification is a key idea in whole of non-linear programming. And it came out of a very, very… It came through very simple steps.

I think it is a very important idea to go and read the original paper of Magasarian-Fromovitz, which appeared in journal Mag analysis and applications in 1967. So, let me just rub this part. Let me see what happens, if I assume that this claim is false and we will assume that this is unbounded. So, this is the mathematician's most biggest weapon actually. The biggest weapon is a proof by contradiction. I am assuming something, which is contradictory to what I should do. And then I am trying to resolve the matter by coming to some absurd conclusion. Hence, my assumption be wrong and whatever I have assumed earlier, what I have claimed has to be true.

(Refer Slide Time: 16:19)



Let me assume that this is unbounded – lambda k, mu k – this sequence is unbounded. So, it means there is a subsequence among these in this sequence, whose norm would go to infinity. So, how do we come to such a conclusion? Because if a sequence is bounded, if a sequence has a limit, then every sequence has to have every subsequence; or, if a sequence is bounded, then every subsequence is bounded. If one of the subsequence is unbounded, then the sequence cannot be bounded. Thus, p implies q and negation of q implies negation of p; that is what we applied to conclude that, if this is unbounded, there is a subsequence here, which is unbounded. So, without loss of generality, means, without trying to (( )) and making the things to fit complex, we assume that… And, hence, assume – which mathematicians use the symbol very much – without loss of generality that this norm of this vector – this is going to infinity as k tends to infinity.

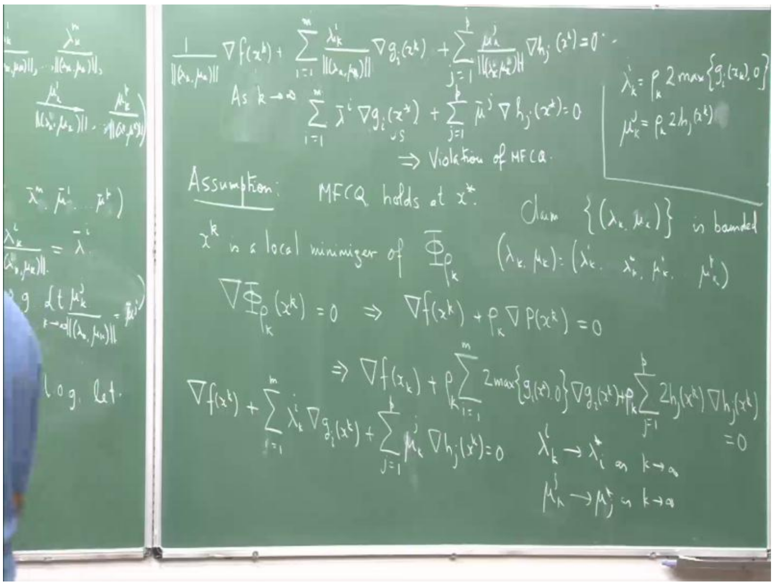
Now, once this idea is known, then here we can start making sequences, which are bounded. Now, let me define a sequence of this form. Let me define a sequence – define… Now, define w k to be of this form for every k. So, w k is… Obviously, this is a very big vector; rather I should be… I think the symbolism is not really correct. I think I should say… I think the symbolism has to be bit better, because you see what I want to claim is that, I am making a vector out of these. So, basically, I should write lambda. See when I am writing this, I am meaning lambda 1, lambda 2, lambda k… Let me… Maybe that will be too bad a shorthand. So, how do I write? Maybe I should write… Maybe I just write this; let me see; that is much more better. And, I will make the symbolism much more better.

So, what I here mean is this fact is that, lambda k mu k. And, I am pretty efficient that, making that; is lambda 1 k, lambda m k, mu 1 k, mu p k. So, that is exactly what I mean by this. (( )) I should not have kept r e and j. I think that is not a good way of writing things. So, this sequence of vectors that I have – lambda 1 mu 1 of this form, lambda 2 mu 2 of this form, lambda 3 mu 3 of this form. This sequence I am claiming to be bounded, which are known assumed to be unbounded. So, this is so I now construct a new sequence w k, which is of this form. Each element of w k is of this form.

Now, if you observe, w k is a bounded sequence. With each norm, w k is of norm 1. So, w k is bounded, because since norm w k is 1. And, observe that this bounded sequence would have a convergence subsequence. So, without loss of generality, again, we can assume that, w k converges to some w star. So, without loss generality, let us assume… So, without loss of generality, let us assume let w k converge to w star. So, w k converges to w star. And, norm of w star is 1, which is obvious. And hence, this w star is a nonzero vector; it cannot be 0. Now, what I do; how would I handle this thing now? Let me now go and remove the…

Here I divide all the things by 1 by norm of this. Basically, if you look at w k; if you look at the vector w k, how does it look like? The w k vector looks like the following lambda 1 k norm of lambda k mu k lambda m k norm of lambda k mu k going on mu 1 k divided by norm of lambda k mu k and mu p k divided by norm of lambda k mu k. This is w k. And, this goes to w star. So, w star – these are individually converging to some num – numbers. So, w star can be written as lambda star lambda bar 1 k lambda bar m k mu bar 1 k mu bar p k. And, this lambda i k… So now, limit of… where limit of lambda i k divided by norm of lambda k mu k as k tends to infinity, this is going to lambda bar i k. Similarly, the limit of mu j k divided by norm of lambda k mu k; so, as k goes to infinity, this is going to mu bar j. Here there is no k; there is no k here. We have taken the limits. So, k is gone. So this is what is happening. Now, once this happens, it means that here I can do something. So, let me look at this equation. And, I am not rubbing this, because I want you to remember that actually this is what it is; this lambda k and mu k if you divide norm, etcetera or whatever. Now, what you do is very simple in the sense that you divide all of these by this number.

(Refer Slide Time: 26:21)

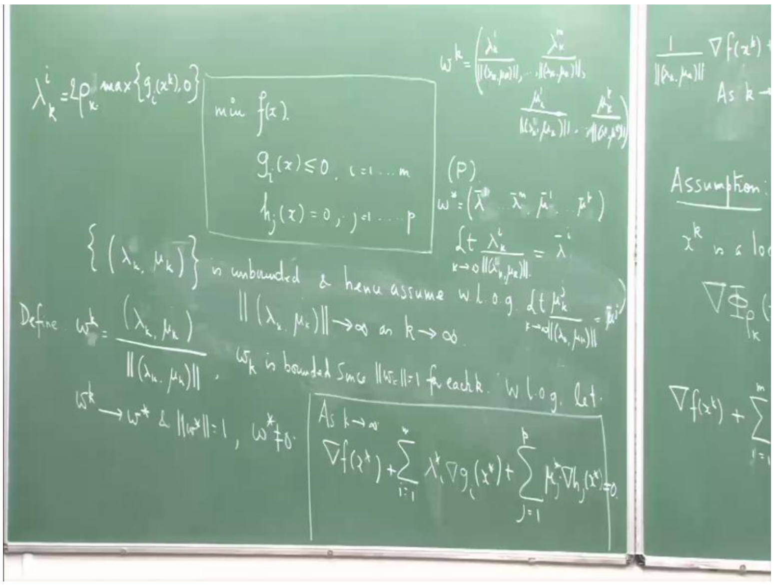


So, you have 1 by norm of lambda k mu k grad of f of x k. The summation i equal to 1 to m lambda i k divided by norm of lambda k mu k is that k; it is not very clear to you – and, that into grad of g i x k. Of course, I am assuming that all the functions are continuously differentiable; they are smooth; that is, their gradients also are continuous – plus j is equal to 1 to p mu j k by because this is going to infinity; this norm of this vector is going to infinity; this vector is going to infinity. So, it cannot be all 0. For k sufficiently large, (( )) it has to be nonzero. So, that means that this division is making sense. And so lambda k mu k. In fact, we can assume that this is always a nonzero vector again and again (( )) the symbolism grad h j x k.

Now, suppose I take a limit. Now, if this is going to infinity, this is going to grad of x star. So, this is the bounded sequence and here is the sequence, which is going to 0. Hence, the whole thing here goes to 0. What I am left? So, as k tends to infinity, what I am left with it is the following. I am left with the fact that summation i is equal to 1 to m lambda bar i grad g i x star plus j is equal to 1 to p mu bar j grad h j x star. So, this is equal to 0. Now, you know that this vector w star, which consists of these vectors is a nonzero vector. So, one among them is nonzero. And, that you know – violates the Mangasarian-Fromovitz constraint qualification; that violates the Mangasarian-Fromovitz constraint qualification. So, if MFCQ holds that x star, then this expression cannot hold unless all of these are zero.

Here we must have at least one of them nonzero, because w star is of norm 1; that is, w star is a nonzero vector. Hence, this implies violation of MFCQ; which means that this sequence – they form a… This is a bounded sequence. So, you see the importance of MFCQ. It makes the multipliers bounded. But, this sort of sequence of multiplier is bounded. Now, what happens is the following; that once I know that these are bounded, then… Once I know that these are bounded… So, there I know that lambda i k is going to some lambda star i and… Once I know this is bounded, then I can assume that lambda i k goes to some lambda star i as k tends to infinity. Obviously, there is a subsequence, which goes. But without loss of generality, we will always take a bounded sequence just converges. And mu j k goes to some mu star j as k tends to infinity.

(Refer Slide Time: 31:09)

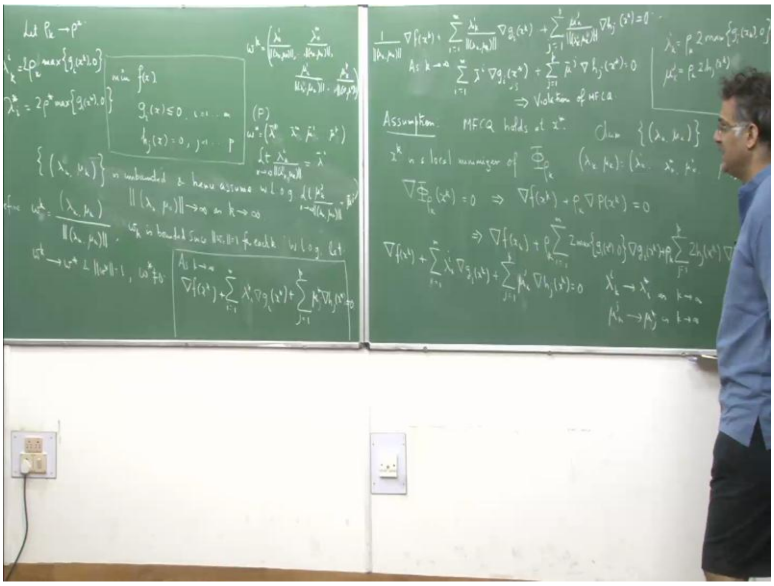


So, as k tends to infinity, what I get is the following. Finally, as k tends to infinity, what I get here is the following. We get grad of f x k plus summation i is equal to 1 to m lambda star i grad… No, this will become x star, because x k is going to x star – grad g i x star plus j is equal to 1 to p mu star j grad h j x star is 0. So, one part of the Karush-Kuhn-Tucker condition is done. What we need to show is a complementary slackness condition. So, let us see how do we look at the complementary slackness. So, what is now more important is to note the following that, we now have to prove the complementary slackness condition. If I look at the expression – this one, here 2.

Now, when lambda i k is going down you know, this rho k cannot be unbounded. If rho k is unbounded, then lambda i k cannot be (( )) If rho k… because if I am assuming that lambda i k is converging to lambda star i, then I cannot take that rho k is an unbounded sequence. We cannot assume that rho k is unbounded, because suppose if I am assuming that rho k is unbounded, then things can just get out of track, because suppose there is a convergent. If rho k is unbounded, there is a subsequence among rho k, whose norm is going to infinity, which immediately means that among the lambda i k. The corresponding subsequence will go to infinity, which means that sequence is also unbounded, because if the sequence has to be bounded, then all the subsequence has to be bounded by the same bound.

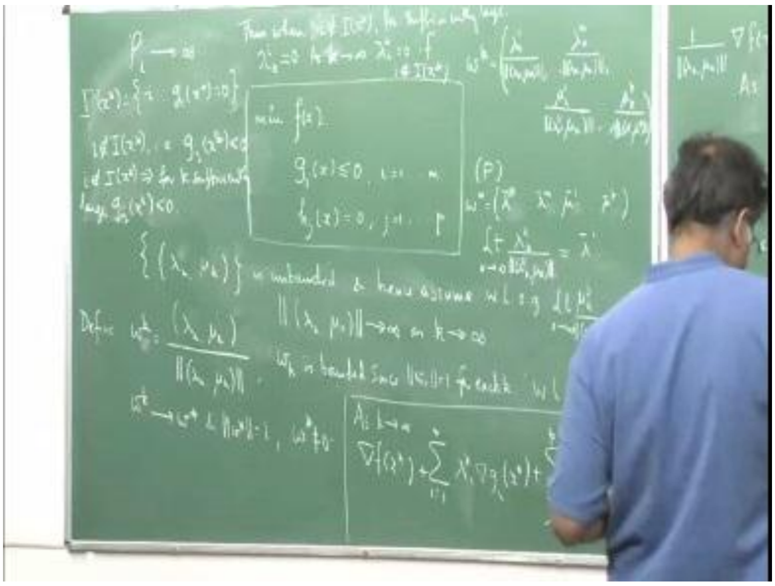
Now, all the subsequences have to be bounded by the same bound. All the there cannot be any subsequence, which is unbounded. If there is there is one subsequence, which is unbounded, means the sequence is not bounded. So, this is something very important to note that, if I have rho k to be unbounded, then all of the subsequences… There must be a subsequence among rho k, which is unbounded; which means that lambda i k – all the lambda i k… Among the lambda i k, the corresponding subsequence corresponding to rho k being unbounded that would also form an unbounded seq… In looking at the structure of the sequence, that will also be unbounded; which will tell me that here I have a subsequence in rho k, which is unbounded. So, the sequence cannot be itself be bounded, then that subsequence also has to be bounded. So, that is the very important thing that, if a sequence is bounded, then all the subsequences have to be bounded inside; there cannot be unbounded subsequence inside a bounded sequence; which means rho k is bounded. So, rho k…

(Refer Slide Time: 35:21)



Let me assume that rho k goes to rho… Let rho k goes to some rho star. So, lambda star i is twice of rho star max of – max been a continuous function; let us take that inside – so, g i x star. So, whenever g i x star is strictly less than 0, then now what we have proved is that these sequences are bounded. We are telling that these are not unbounded sequences. If the sequence is unbounded, then we will lose MFCQ at x bar.

(Refer Slide Time: 36:19)



But remember when we run a penalty method, then rho k goes to infinity. Rho k itself is an unbounded sequence. But the product etcetera makes a bounded sequence. Now, what

I have to really prove is a complementary slackness condition – the last part of the Karush-Kuhn-Tucker condition. Now, rho k is unbounded. It looks very strange. And, lambda i k is bounded. So, it might look very strange to many people. But let us now argue how do we look at it. What we need to do to get the complementary slackness condition? We need to know that… If I say that this is the set of active indexes for the inequality constant – set of i for which g i x star is equal to 0. If I know this fact, then what I have? If I know this fact, then I know that if…

Let i is not element of i x star, that is, g i x star is strictly less than 0. Then, around its bar; by continuity of the g i's, there is a ball; and, in that ball, for all the x, g i x would be strictly less than 0; which means for k sufficiently large, whenever i is not in i x star, that is, this; whenever i is not in i x star, it implies that for k sufficiently large, g i x k is strictly less than 0, is also inactive; which means for k… When i is not element of… which means thus, when i is not in i x star, for k sufficiently large, we have… because this is strictly less than 0 for k sufficiently large, because i is not in i x star. So, this max is 0. So, this whole thing is 0. We have lambda i k is equal to 0. As k tends to infinity, lambda star i is equal to 0 if i is not element of i x star. And, that would give you the complementary slackness condition; you have to think a bit about it.

Now, I would like to end up with a question here (( )) these are the very pretty involved argument. Now, the question that I want to end up here is the following; that do you think that this is true that for most cases, for most situations, when you use this penalty function P x; this penalty function would not in general provide you a solution of the penalty function for rho k very large, because rho k is tending to infinity, may not give you in general a solution of the original problem. That is, you can keep on increasing rho k, but still the x k that you get may not feasible and may not be actually the solution of the original problem. What would you, do you think that such a statement would be correct? The problem is that, this issue is a major issue in understanding optimization; that if you use penalty method, the x k that you would choose at the end – when you stop and choose an x k that may not be at all feasible to the original problem. And that (( )) leads to the study the approximate solution with a certain degree of relaxation, all those things.

But what is important that, can we devise a penalty function, whose minimization for some large value of k would give me a solution to the original problem under certain very simple natural conditions. That would lead us to a study of what is called exact penalty function that we start tomorrow. But in the exact penalty function, we need to have an additional property of the function that the functions have to be twice continuously differentiable, because we need to use second order sufficient conditions for constant optimization to understand that.

But we have not yet done anything about second order sufficient condition. You know in unconstrained case, you know that, the easy one is possibly definite; then, you will have a local minimum, rather very strict local minimum. But, for the constrained case, we have not told you how to know a local minimum. Suppose you have a Karush-Kuhn-Tucker point. How do you know that is a local minimum? So, that second order condition is something which we will also discuss after giving you a brief idea of how to construct exact penalty function and what can it give us.

We will first derive the second order conditions for sufficient conditions for optimality and then will use that idea to show that the exact penalty function can indeed provide us what we really want; that if I am solving that problem, I will get the solution of the original problem. But, exact penalty function is sometimes also called as Zangwill-Eremin exact penalty function. Zangwill is an American optimizer, while Eremin is a Russian mathematician. The problem that lies largely with exact penalty function that we will see tomorrow is that they would be non-differentiable especially at the solutions of the problems. So, this non-differentiability would create a lot of issue. But somehow, we would be able to manage through.

You will see that, nowadays, there is lot of tools to handle non-differentiability. So, that would not become too much of an issue. But theoretically, it is a very, very powerful thing, because it tells me that, you need not actually bother about the constrained problem, you can bother about a unconstrained problem, but which is of the non-smooth type. But if you can solve it successfully, then you lead at – after certain large k, you land up with the solution of the actual problem, which is a very, very fascinating result. That we will discuss in the next two classes.