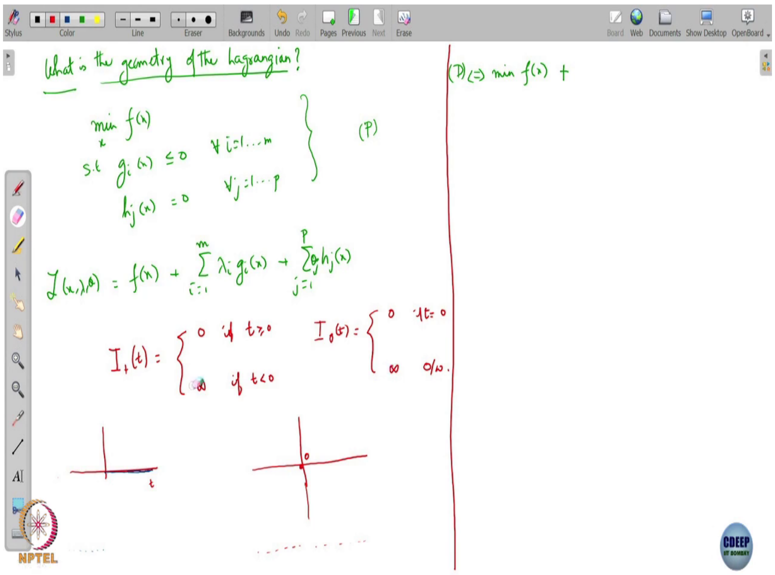
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 17C
Geometry of the Lagrangian

Let us talk a bit more about the dual function. So, what is the dual function?

(Refer Slide Time: 00:23)



What is the geometry of the dual function or of the Lagrangian? What is the geometry of the Lagrangian? So now, here is one way of, so the Lagrangian or and so on all of these appear, they make an appearance in problems that are constrained. But we can convert these sort of problems that are actually constrained to problems that are unconstrained in the by doing the following thing.

So, I have suppose a problem that I mentioned before which is minimize f x subject to over variables x subject to g i of x less than equal to 0 for all i from 1 to m, and h j of x equal to 0 for all j from 1 to p, and then I define this as my Lagrangian lambda of L of x comma theta as f of x plus right. So, this was my lambda L of x, lambda, theta.

Now, suppose I did the following, I decided I want to express this constrained optimization problem as an unconstrained optimization, how can I do that? So, to do that let me introduce this function. This function let us call this the function I plus ok, the function I plus of t. What does this do? This function is 0 minus infinity it is greater for t greater than equal to 0, it is 0; and for t strictly less than 0, it is minus infinity right.

Student: Yes.

So, I plus of t, how does this look as a function of t? So, here is suppose my t for t is strictly less than 0, this I plus is equal to 0, change the color for strict say it is this for t is less strictly less than 0, it is sorry for t strictly less than 0 it is equal to minus infinity right, so for alright.

And for t and for t greater than equal to 0 it is actually equal to 0. So, this is my function I plus. So, it is it takes value minus infinity when t is less than 0; and 0 onwards it takes value it takes value 0. Let me define another function related looking function I 0, I 0 of t. I 0 of t is defined in this way.

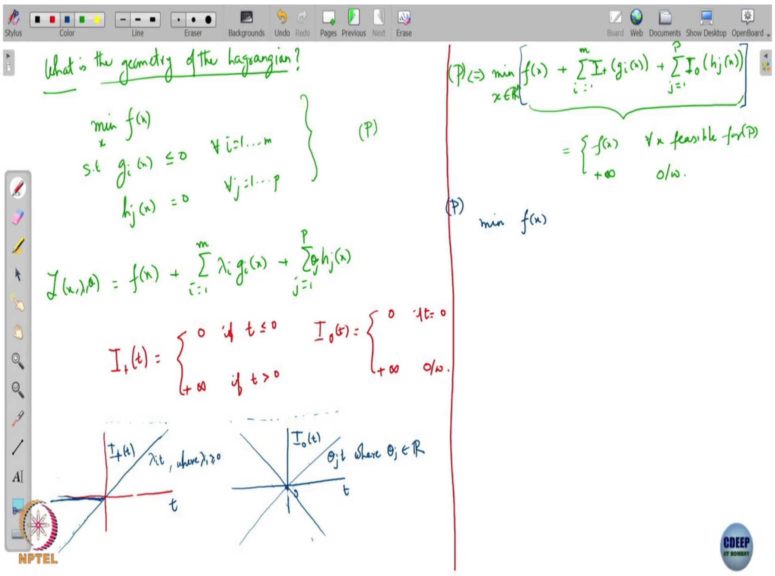
I 0 of t is equal to 0 if t is equal to 0, and it is minus infinity otherwise. So, how does this function look? This function looks like this. At 0, it is 0; and everywhere else its value is minus infinity ok. Now, these are obviously very ill-behaved functions. They are taking value minus infinity and so there is obviously a huge discontinuity, and non-differentiability here.

But in terms of these functions you I by if I include minus infinity in my calculations, I can express an unconstrained problem in terms of the I can express this constrained problem as an unconstrained problem. So, how do I do that? So, notice that this optimization ok, this problem let us call this problem P, P is actually equivalent to minimizing f x.

Now, what do I, what do I need to do here? Let me just change this a little bit. I will just change this definition a little this definition a little bit. So, actually let me make a slight change in this definition because that will be convenient for us. So, let us instead of taking defining I plus and I 0 in this sort of way, let me put this as using this as use this as plus infinity ok, and also I will change the range. So, it is plus infinity.

So, define I plus of, so let us, so define these functions. So, define I plus of t as is in this sort of way at.

(Refer Slide Time: 06:36)



So, for t for so whenever t is less than equal to 0, it is equal to 0; and when t becomes positive, it shoots to plus infinity right. And I 0 of t is 0, when t is equal to 0; and whenever t is not equal to 0, it again shoots to plus infinity right. So, what, does this look like? What sort of function does this look like? Let us draw this here ok. So, I plus of t as I said it is for t less than equal to 0, it is for t less than equal to 0, it is 0; and for t greater than 0, this it shoots to plus infinity.

So, this is my I plus of t. And I 0 of t looks like this. At 0, no problem, it is equal to 0; and whenever t is not equal to 0, it shoots to plus infinity. This is I 0 of t, alright. Now, using these functions I plus and I 0 of t, I can actually express my I can express my the optimization problem P. I can express in the following way. I can write this as minimize f x plus summation lambda i I plus of g i of x, sorry, I do not need the lambda.

So, using these two functions – I plus and I 0 of t, I can express my optimization problem P in the following way. I can write it as minimize f x plus summation I plus of g i of x, i going from 1 to m plus summation I 0 of h j of x, j equals 1 to p. So, now what does this do? What does this do?

Well, it says that look at I look at the definition of I plus, whenever g i of x is less than equal to 0 ok, I plus is equal to 0 alright. So, in that case, this term here, the term here, this term actually is equal to 0 whenever g i of x is when whenever g i of x is less than equal to 0. In particular when all the g i of x's are less than equal to 0, this entire summation is actually equal to 0.

Similarly, look at this term, whenever h j of x is equal to 0, this term is equal to 0, right; this term is 0. So, whenever, so in short if I take any x that is feasible for P, then each of these terms the I plus terms as well as the I 0 terms, each of these terms should would end up being exactly would become exactly equal to 0.

So, in short, on the feasible region, this new function that I have defined you know this new non differentiable infinity value taking function, this function is on the feasible region is actually nothing but f x. Now, outside the feasible region of P, outside the feasible region of P, what is this function?

Well, if you are outside the feasible region, then it means that at least one of these terms is going to be not 0. So, at least these if either one of at least one of these I pluses or one of these h j's I 0's one of these is going to be non-zero. And when it is non-zero, what value does it take? Whenever it is if it is not 0, if these things are equal to plus infinity remember.

So, this they take value plus infinity. So, which means that once you are outside the feasible region, this here, this expression here actually takes value plus infinity. So, what does this mean? It means that this function here let me write it like this. This is equal to f x for all x feasible for P, and plus infinity otherwise.

So, which means then if you have to minimize this if you have to minimize this this function that is mentioned here, if you are minimizing this particular function, what are you doing? You are effectively just minimizing f x over the feasible region of P, which is nothing but solving P itself right.

So, all, so this problem although you are minimizing this over all in an unconstrained way over the, in over all of R n what you are effectively doing is minimizing just f over the feasible region of P right. So, that is actually an incredible simplification because you do not really need to care about the geometry of what is happened in the constraints and so on.

But it is also very deceptive because what you have done all that geometry has actually been absorbed into these complicated I 0 and I plus functions which if you have to analyze you would need to understand the geometry of the g i's and h j's in the first place alright. Now, what is this got to do with the Lagrangian?

So, if you look at this the Lagrangian function here that is less that is mentioned here, and if you look at the function that I am optimizing here is clearly a close resemblance. Because here is your f plus a summation of something right and some other terms with the here they have inequality constraint, and then here you have your equality constraints right, and likewise here you have f plus something that involves inequality constraints plus something that involves equality constraints.

So, now, what is the connection between these two that something that we can see we can see now. So, what I have you can think of you can think of it this way that what I have done is actually in place of the I plus and in place of the I 0, I have put in some new functions here which are actually linear functions. So, in place of I plus of g i of x what I have done is put in lambda i times g i of x and likewise in place of I 0 of I 0 of h j of x, I have placed theta j of theta j times h j of x.

Now, what is that, what is that actually doing? So, let us come back to this figure. See remember I plus has this sort of form where for t less than 0, it takes value 0; and for t greater than 0, it shoots up to plus infinity. Now, if I want to approximate this linearly in a, if I want to do a linear lower approximation to I plus of t, what sort of function can I choose? Well, the kind of functions that I can choose have to be of this sort of form.

So, I have not drawn this very well, let me draw it again. So, the kind of function that I can choose has to be of this sort of form. So, what does this mean? What kind of form is this? It is a function whose slope is like this, it is positive.

Because if I take a function just this was for just for you to see, if I took a function whose slope is negative, then it would at some stage ok for some value of t, it would go above 0 above this line here. And then it will not be a lower approximation anymore right.

So, you know, so for it to be a lower approximation it is necessary that it must have a slope a positive slope like this ok. It must have a positive slope like this. Moreover, for it to since it must have a positive slope, I can look for the best lower approximation amongst the guys that have positive slope, and it is clear that there is no use having an intercept here.

So, taking a lower approximation like this is of no use. I might as well get a better lower approximation by taking the intercept by making sure it passes through the origin in short the intercept should be 0, right. So, it is of no use taking this kind of lower approximations. So, what we can take our approximations like this. We take a lower approximation that is we take a lower approximation like this which is passing through the origin alright.

So, what does this mean? The, in short a lower approximation to I, a lower linear approximation to I plus is takes the form it takes the form lambda i times t ok. So, this one here is a function of the form lambda i times t ok, where lambda is lambda i is greater than equal to 0.

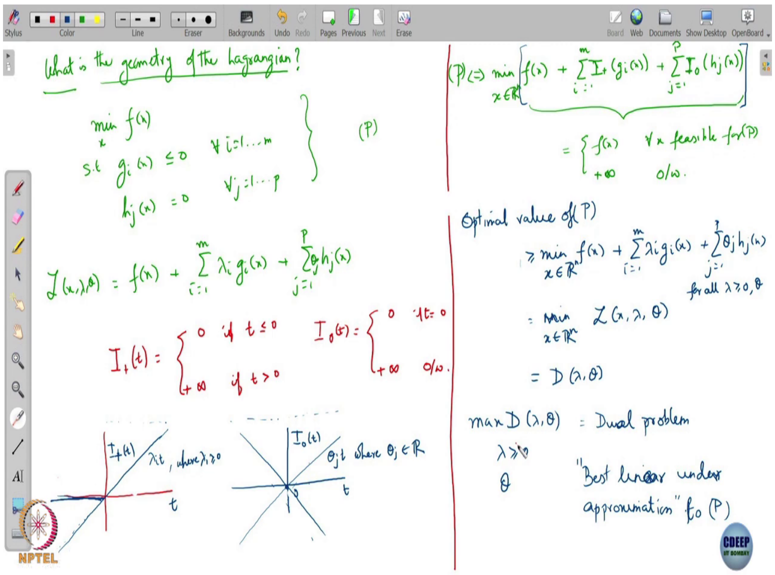
Likewise a lower let us look at now I 0. So, if I have if you look at I 0, if I 0 has this it takes value plus infinity everywhere except for 0. So, if you want to take a lower approximation to this, now there you can take a linear function like this, you can also take a linear function like this. So long as the intercept is below 0, there is no problem right because everywhere else the value is plus infinity and you will be ok you will be alright.

So, long as the intercept on the y-axis is negative, you can continue you can take any kind of linear function. But again it if you want a tighter approximation, if you want a better linear approximation, why even bother with an intercept, you might as well take a linear function that passes exactly through the origin.

And that gives you that means that the function should be of the form theta j times t either this or this or whatever right. This, these are functions of the form theta j times t where theta j is any real number right ok. So, what does this mean? This now these are both linear approximations alright.

So, which means that point wise, that means, for every t they actually take value less than the corresponding I plus or I 0 respectively right. So, which means that if I look at what is here, if I look at what is in this optimization, what I just put in the bracket that can always be lower bounded if I replace these, the I plus and I 0 by their respective linear approximations. So, I can get this is always greater than equal to write it like this.

(Refer Slide Time: 19:11)



So, P is the optimal value of P, this is always greater than equal to the minimum over x in R n of minimum or infimum whatever over x and R n i goes from 1 to m j going from 1 to p theta j times h j of x. What is and what is this? This is actually nothing but the in you are doing effectively the minimum of the Lagrangian. And this is equal to your dual function.

So, long so the greater than equal to here, this required that lambda is greater than equal to 0 and any theta. So, for all lambda greater than equal to 0 and for any choice of theta, we have that we have that the optimal value of P is greater than equal to this which is nothing but the dual function.

And now so what is effectively the dual function doing? The dual function is taking a linear approximation to these I 0 and I plus functions ok. And solve the dual function is the optimal value of that linear approximation. So, or in other words, the, what the Lagrangian is actually doing is taking a linear approximation to these I plus and I 0 functions. And the minimum of the Lagrangian is basically the minimum value of this linear approximation, and that is what we are calling the dual function.

And why is it a, it is a function of what? It is a function of the slopes that you choose for making the linear approximation. So, as a function of these of the slopes, you could have we remember we said there is no need to take the intercept and that is why we took the intercept as 0 and we got these linear functions passing through the origin. But we did not say anything about the slope.

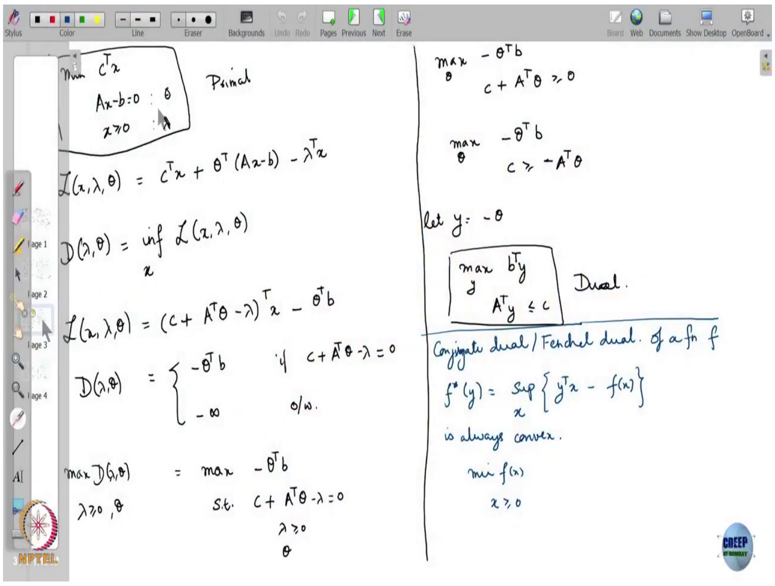
The slope is up is still up for grabs, it can it is still to be decided. So, as a function of the slope, the gap between the actual optimization and or and this can still be fine tuned. The actual approximation, the actual optimization and the and its linear under approximation can still be fine tuned.

In any case the linear under approximation is giving is captured by the dual function which gives it to you in is a function of the slope. So, maximizing the dual function, so maximizing this which is your dual problem maximizing this which is your dual problem is basically asking for, what does this ask for? It is asking for the best linear under approximation, it is basically asking for the best linear under approximation to P.

So, in this class of under in this class of approximations you can what is the best you can do right? So, you so the sequences you have you create your you write your actual problem like this, you write your actual problem like this, create a family of linear approximations using this logic.

You look for the best value of the minimum of those linear approximations and then you ask ok, what is the what is the best I can do amongst my entire family, what is the largest value of my what is the tightest lower bound that I can get using this linear approximation alright ok. So, that is what that is what the dual problem is doing.

(Refer Slide Time: 24:04)



Now this if you think about it this way it is actually nothing short of a miracle that in the case of linear programming the primal and dual actually end up being equal. So, the optimal values of the primal and dual being end up being equal. Which means, what I mean by that is see you see how grossly inaccurate this entire linear approximation is.

So what you wanted to actually approximate was this sort of function something that 0 here and shoots to plus infinity after that. Likewise what you wanted to approximate here was this function that is plus infinity everywhere and 0 here. And you are approximating it by what?

An extremely benign function, you are just taking a linear function like this. And then you are saying ok amongst this class of approximations which is the one that is giving me the best possible value that is by; that is what you are solving by solving the dual problem right.

And it is incredible, it is actually really incredible that you can in fact get back the same value as of the primal; that means, that there will be no gap between this one which involve these I plus and I 0 functions, and this problem that has been formed by looking for the best value of the linear approximation ok.

So, in the case of linear programming that is exactly what we get. We, the linear programming duality theorem taught us that the primal optimal and the dual optimal are whenever there is a solution to the primal, there is a solution to the dual and the values are equal, and that is what we are finding here ok. So, this sets the stage now for convex optimization duality. And so, we will do that we will do that in the next lecture ok.