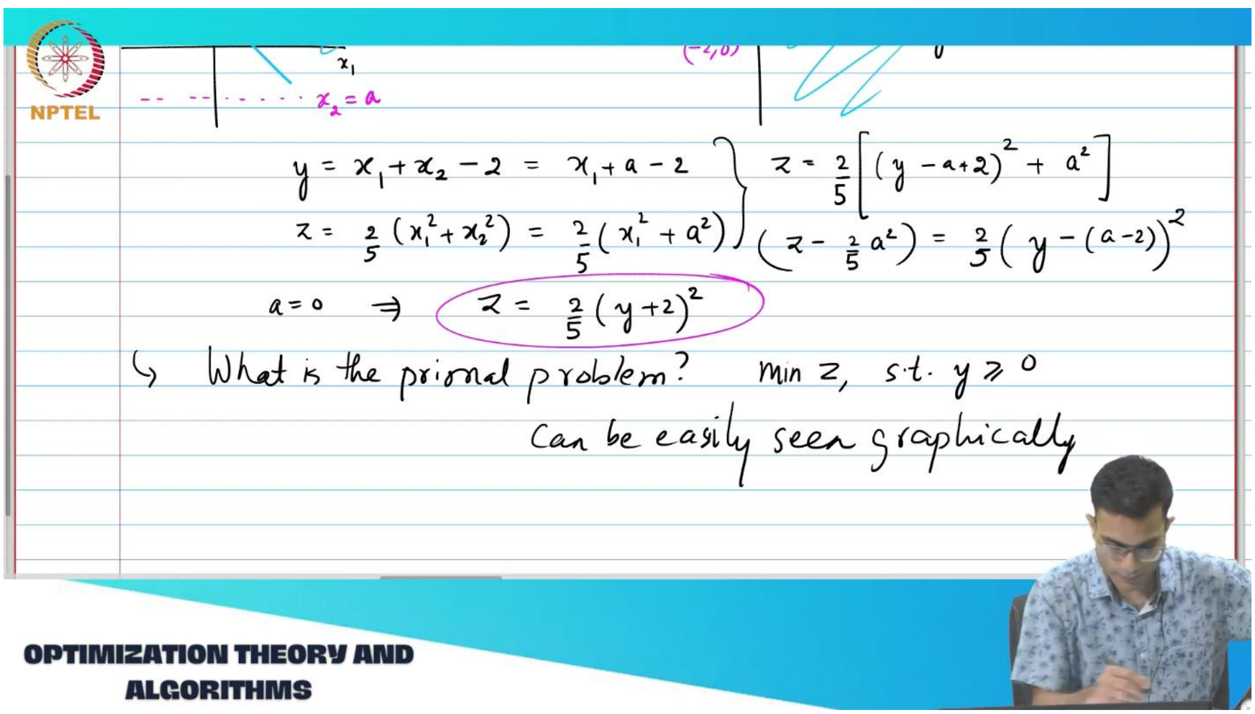
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 12
Lecture - 79

Intuition of duality and dual problem

So, now that you have this geometry in mind, let us ask a very simple question. What is the primal problem? What is the primal problem when referring to the yz-plane? So, what would you say? You have the yz-graph in front of you. So, how would you state the primal problem? And you can refer to this guy. In the primal problem, what were we trying to do? Minimize f such that $c \geq 0$. That was the primal problem. Minimize f such that $c \geq 0$.

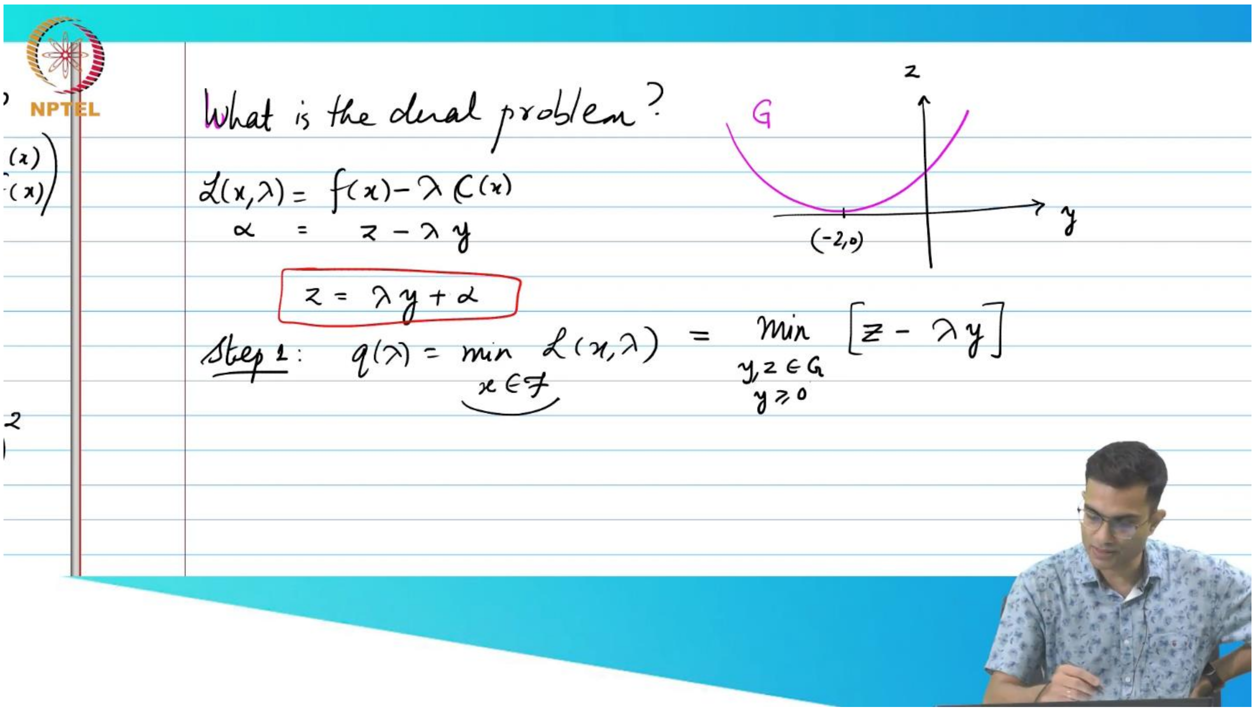
Now, can we express this in the yz-language? Minimize z such that $y \geq 0$. There is no z that is negative. I want to make z as low as possible. Don't confuse the objective function and the constraint.



So, forget this mapping. Even if you had not done this mapping, what is the primal problem in the yz-coordinate? Give me as low a possible value of z which is f, such that $y \geq 0$. Not everyone will agree. Now that we have, so let us just write that down over here. So, the primal problem is minimize z such that $y \geq 0$. So, in this yz-plane, what is this? Can we immediately see the solution? What is it? It is where this pink curve is cutting the z-point, right? It is basically this point over here, the orange point.

Because this is the smallest value of z that I can have while $y \geq 0$, while I am in the right half-plane, right? So, in this yz-plane I have got the solution again graphically. Okay, so the primal problem was simple, just right. So, do not think that the x-coordinate axis is the primal problem, yz-is the dual problem; that is not a thing. We are going to do this transformation and look at the primal and dual problem both in the yz-system. Okay, so it can be easily seen, okay? And for ease of reference, this pink curve I am going to call G, capital G.

Now, maybe I do not have enough space so I will go to the next slide. So, the next question is, okay, let's just draw this again. Okay, so for solving the dual problem, what was the first thing that I, let us write down our favorite Lagrangian?



So, this is going to be very key, $f(x) - \lambda c(x)$, okay. Now, f was mapped to what? y or z? z, right. So, this was equal to $z - \lambda y$, okay. So, just for shorthand notation, this $L(x, \lambda)$, just going to give it a shorthand notation, let us call it α. And this was my y, this is my z.

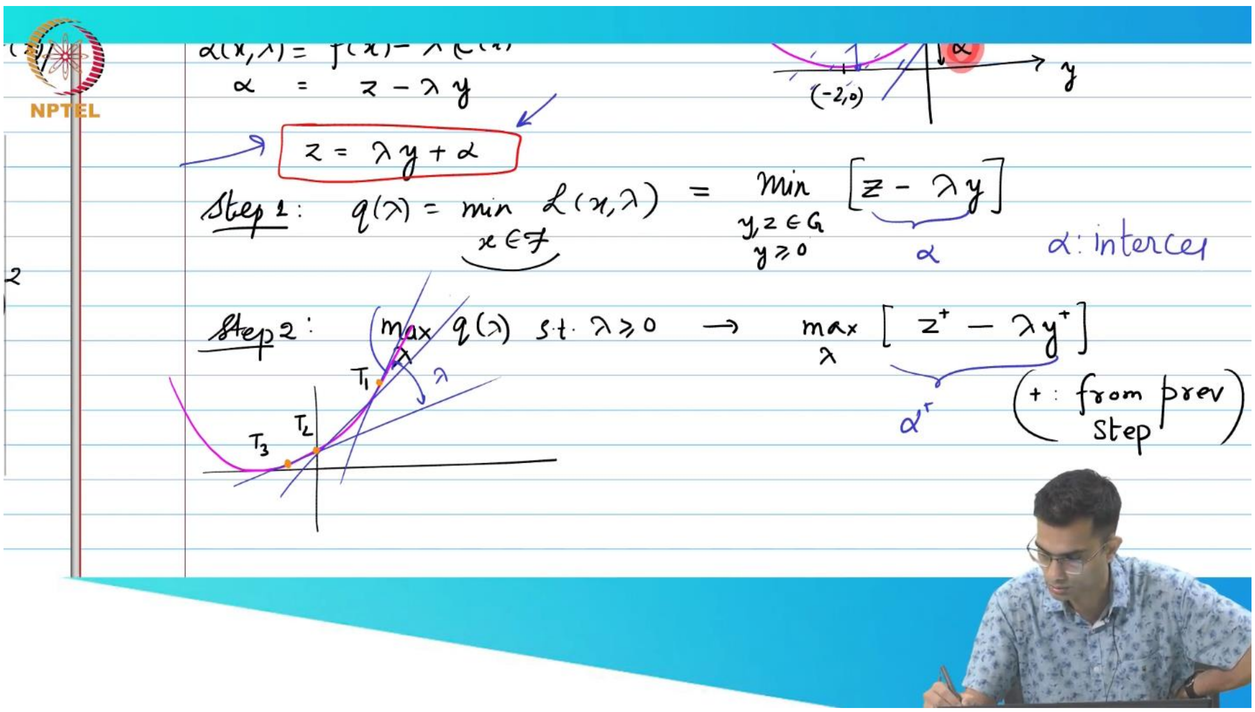
So, this equation which I have got, which is the equation of the Lagrangian, what does it say? It says that $z = \lambda y + \alpha$. This is like $y = mx + c$. I know the slope and I know the intercept. So now let us look at step 1 of our Lagrangian formula. So, I need to do what first? Write the Lagrangian dual, right? So, $q(\lambda)$ was minimize over what? x, right? And x belonging to what? Obviously the feasible region, right? So, let us say this is the feasible region and the Lagrangian of x, λ.

Now, let us write this in our new language. So, what is a Lagrangian? It is $z - \lambda y$, $z - \lambda y$, flex 2, $z - \lambda y$. Now that I am in the yz-space, I need to translate x belonging to the feasible set into yz belonging to the feasible set. Obviously, because y and z are not free to be anywhere. So, I can say this, we had already marked this region as G.

I will just say that yz belonging to G. Just a convenient way to denote this whole thing. Actually, let us correct that, it should not be yz belonging to G, it should simply be $y \geq 0$, that was the feasible thing, right? But there is a relation between y and z, right? So that is, that is, that is okay. So let us leave that here.

Because y and z are not free to be any two points on this. y and z are constrained to be by that quadratic relation. So that quadratic relation is what I am calling this. So, y and z are this and we can say that $y \geq 2$. That captures the constraint, right? So this is sort of a clumsy way of saying that respect the mapping and stay feasible.

You are not free to choose y and z as you wish. They are related by this parabola. Okay, so now $z - \lambda y$ is what I want to reduce. So, before talking about what I want to reduce, why don't we try to plot what this looks like? This is a line with slope what? Slope is λ, intercept is α, right? So, what would it, what would it look like? I have to be, let us say on this G thing. λ is fixed, in the definition of Q, λ is fixed.



You give me some λ, for that I will give you a Q. So, if I just wanted to sketch different lines of this, they would be straight lines with slope λ intercept α. So, they would basically look something like this. So, this is $z = \lambda y + \alpha$ and when I do this, what is the value of α? This much, right? This is my value of α and all of these guys go. Now, I am saying that I want x, sorry y and z should belong to this parabola, right? So, is there a further constraint on this line? It should be, it should touch this, it should be tangent, I want to be on this curve for example.

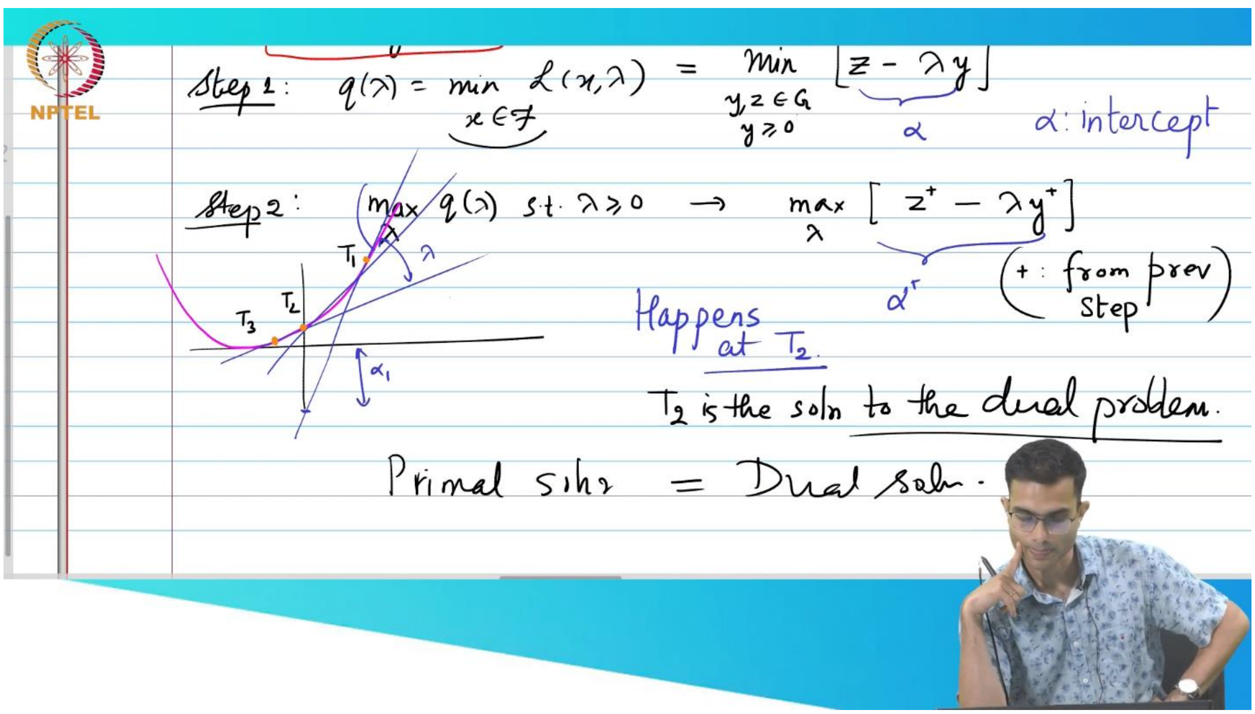
So, not necessarily, I mean I could be anywhere inside this cup region, that is okay. So, these are all, so let us go step by step. These are all the possible straight lines that respect this equation. Now, what do I want to generate the Lagrangian dual function? I want to minimize this guy, $z - \lambda y$. This $z - \lambda y$ is what? Is basically α.

So, what am I saying now graphically? Keep reducing it, right? That is how I am going to reduce my intercept. Keep bringing this line as far down as possible. But I cannot do it forever and ever because yz should belong to G. So, what is the smallest thing I could do? So, I am going to get something like, my drawing is terrible, but if I start erasing this everything gets erased. Let us assume that this is the tangent point.

So, my pathetic drawing is okay. So, these are all different lines, all same slope, and this is the minimum intercept. So, this G, this α over here is as low as I could get because if I try to reduce this intercept any further, what would happen? I would leave this pink region, I would basically fall out. So, that is my problem. Okay, so this has given me, what has it given me? It has given me a geometric intuition of what this Lagrangian dual function is.

Like give me λ, I am going to keep reducing this until I just leave the region, so that tangent place is where I, where I live. Okay, so alright, that is good. That was step 1. Step 2, what is step 2? Having got the Lagrangian rule, what did I want to do? Maximize it, right? So, step 2 was basically $\max_\lambda q(\lambda)$ such that $\lambda \geq 0$, okay.

So, in step 1 make, so basically, okay, so let us again plot it. This is my G-region, okay. For different values of λ, what is this equation $z = \lambda y + \alpha$, what is it representing? λ is controlling the slope. So, here in this optimization problem, maximizing over λ, what am I varying? I am varying the slope, right? So, they are basically all different, different tangent lines. So, I could have something like this, I could have something like this, and so on, right? These are all different, different tangent lines, which are, what are they corresponding to? They are corresponding to different λ.



As I vary λ, right? And because this q is coming from step 1, it is forcing, that is what is forcing them to be tangent. Otherwise, they could have gone inside also, right? So, the tangent line is coming from step 1. Step 2 is varying the slope of the tangent. That is what I am doing. Very clear? So if I mark the tangent points over here, I could have a tangent point over here.

Okay. So I could have a tangent point over here. I would have a tangent point over here. I mean, there are infinite tangent points. I'm gonna call this T_1, T_2, T_3. Just as an example, I've taken three tangent points.

Okay. So let's see what this says. So I want to maximize. I am going to put a special symbol now on z because this z and y are coming from the previous step, they are optimal in some sense. So, this is, they are not any y and z, they are special y and z, $y + z +$ which have come from the previous step, okay. It is not a field. They are just examples of different, different tangent points.

Now, this is what the step 2 says. Again, let us interpret this in words. This whole problem, maximize $z + \lambda y$. What is $z + \lambda y$? It is again $\alpha +$, which is. So, this is both α, is always the intercept. So, now what am I saying? Step 2, I am saying maximize the intercept. Of course, while staying feasible and all of that.

So, of these three lines which I have drawn, which have tangent points T_1, T_2, and T_3, I want to maximize the intercept. So, for example, for this T_1 guy, what is the intercept? This guy, right? This is α_1. T_3 anyway is not feasible because that point is in $y < 0$. What about T_2? T_2 is on the z-axis, so it is feasible. And of all possible tangent lines, which one do you think will have the highest value of intercept? Not mod α, α.

Who has it? T_2, right? For the line where T_2 is the tangent, that is the thing where the intercept is maximum. If I go any further to the right, what happens? The intercept becomes less, less, less, becomes 0. By the time I come to T_1, it has actually become negative, right? So, this happens at T_2, okay? Okay. Is step 1 clear? Step 1 is clear, is that give me the least possible value of intercept and I am staying on the feasible part, right? Now in step 2 what am I doing? I want to maximize the intercept because $z + \lambda y$ is the intercept of some line. What is my optimization variable in this step 2? What am I allowed to vary? λ is in my hand.

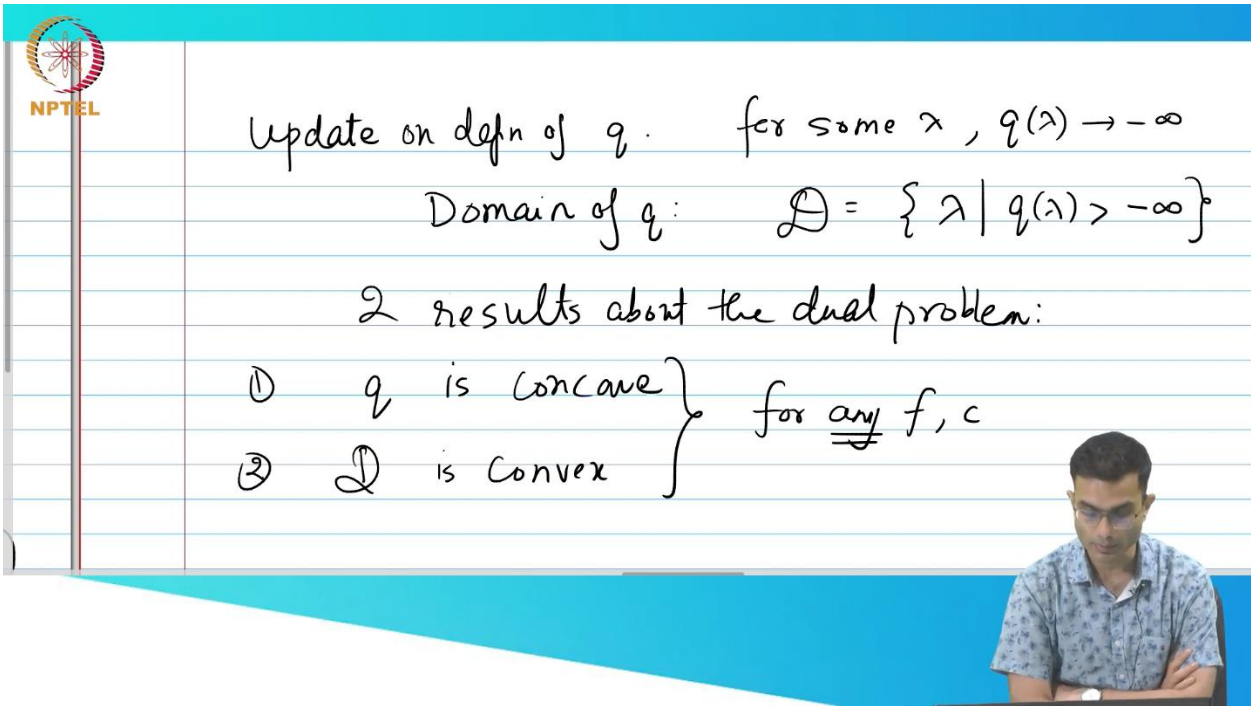
So you can quickly alternate between step 1 and step 2. q, you have to give me a λ. If you give me a λ, I will give you a q. How will I give you a q? I will basically find out a tangent line to G.

That is my q. Now that you have given me a q, that means q is basically the family of all tangents to this guy. Now in step 2 I am saying hey of all of these guys give me that tangent line which is maximizing the intercept. And at maximization of intercept you can see graphically is happening at T_2. If I go any further to the left the tangent jumps into infeasible region. If I go any further to the right the intercept value reduces.

So the best intercept that I am getting because this step is a maximization step is coming from this particular guy t_2. So you would then say t_2 is the solution to the dual problem. So, in fact the solution to the primal problem and the solution to the dual problem, remember in this case it was T_2, in the previous case we could, we saw that the true solution to the primal problem was the same point where this parabola is cutting the z-axis, yeah. So, parabola will cut the z-axis which is not equal to 0. The $y = 0$, where are you substituting? Here.

Yeah. Yeah. So, you will get $2 \times \frac{2^2}{5}$, 6,8 way. But like without using this duality, without using this y, z translation, it will be distortionally solved. Correct. We got $\frac{4}{5}$ as the answer.

We got $\frac{4}{5}$ as the answer. Have we made a mistake somewhere in the algebra? Yeah, I mean you have to.



Right. Oh, okay, okay. I see the problem. No. So, okay. So, there is a little bit of confusion. This $a = 0$ line which I took is just to show, it is the simplest parameter of a to put inside for you to see it is a parabola. This g region, because this is not a very easy thing to sketch, right? I am not marking, I am not actually marking the value at which this g region is cutting the z-axis.

It is not for $a = 0$. You have to do this carefully, the sketching has to be done carefully. This, the lowermost boundary of this is not for, need not be for $a = 0$. In fact, what you are saying proves that it is not for $a = 0$. So, you do this for different, different values of a, you will come up with some region G, which looks like this parabolic region. So, this cutting point is, yeah, you are absolutely right, this cutting point need not be for $a = 0$.

I mean not necessary. Let us just imagine that this is some, some region, pink region, ok. So, that is why I am not putting any numbers down on this orange point. At some point, $a = 0$ was just chosen so that you can immediately see it is a parabola.

That is all. Yes, they are all convex. Yeah, that is a good point, ok. So, in this case, the primal solution one moment, is equal to the dual solution. Yes. No, no, no, no.

Choice of a is not related to choice of λ in any way. What was the, what is the motivation that I had in choosing $x_2 = a$? Can anyone try to read my mind? Is to basically do the mapping. If I take different values of a from $-\infty$ to ∞, I will, if I plot it on the y, z plane, I will see what is the entire x_1, x_2 region mapping to. So it is a coordinate transformation you can think of it. And that coordinate transformation gives me this pink region.

Yes. Seems like magic. There are so many such transformations, you know, you can map a half plane into a circle. So if you take a course on complex analysis, you have all these kinds of mapping that happens, nothing very surprising about it, Z transform. So we have this situation here where the primal solution is equal to the dual solution and it is because you saw the problem and the constraint were convex. Question? We choose a tangent to the pink region.

Off slope λ because λ is fixed. Yeah. So wait, before you go to step 2, keep in mind step 1 is not one tangent. Step 1 is a family of tangents for different different values of λ. So it's like I have a database of all of these possible different tangents. Now in step 2 what I am going to do is choose a tangent which gives you the highest intercept.

Right. That is a convenient way of thinking about it. Okay, so when these guys are equal, this is called as I already mentioned strong duality. Okay. So, there is just one small update that I want to give to the definition of the Lagrangian rule because it did not come in the problem, but in general a small update on sort of more general of q ok. It can happen that for some λ, so for some λ it can happen that $q(\lambda) \to -\infty$, it can happen.

As we saw this was a concave function. So, there could be some λ for this for which this goes to $-\infty$. So, the domain of q is defined like this, λ such that $q(\lambda)$ is basically not blowing up. So we just keep this in mind, this is a small technical point but when you do your dual problem, when you are formulating your Lagrangian dual function $q(\lambda)$, there we said okay λ could be anything but we have to keep in mind that those λ's which lead to q blowing up are not allowed in the domain of q. This is one small point to keep in mind, okay. Nothing to do with positive definite over here, right? In fact, q is concave.

So, it is negative definite. Negative definite. So, a negative definite function can also blow up, right? The parabola going down at some value of λ, it will go to $-\infty$. So, what this is saying is that the domain of q should be restricted to the region where it does not blow up. Just so that it makes sense to talk about q.

That is all. As I said, this is like a minor technicality. You need not worry too much about it. So, I think this is a good place to stop. There are certain properties of q which are very useful and powerful, we will look at it in the next class. Maybe I will just state the properties, I will not prove it because Karthik had asked the question. So, two results about q or about the dual problem which we will prove the next time.

q is concave, second is this D which is the domain of q is convex and here is the sort of rabbit out of the hat for absolutely any f, c. Any f, c that means f could be non-convex. So this is, this is why this, the dual problem ends up being very very special because you can throw whatever problem you want in the primal sense. When I make the dual problem corresponding to it, I always get a concave problem, right. So if I always know that a problem is concave, is not it, it is like, it is just a negative of a convex problem, right.

So, that is why in general solving the dual problem is easier because it is, you are dealing with, if I know how to deal with a convex problem, I know how to deal with a concave problem. I just turn my paper upside down basically. So, that is why this is one of the powerful motivations behind research on the dual problem. It gives me a concave problem and the domain of this function is convex. And we have seen all the proofs that we have used about various things have used convex sets everywhere.

So, this gives us a very strong theoretical basis to think of convergence proof and all. But the flip side is that if f is, for any f or c, so if f is non-convex, you can solve the dual problem, it does not solve your primal problem because there can be a duality gap. The solution that you get to the dual problem and the solution of the primal problem, they need not be the same, they can be a gap. So, you may get a solution, but there may be a better solution of the primal problem that is missing from here. Which boundary? Yeah, but the dual problem is always concave, I mean. There is a proof, we will study the proof in the next class, not very tough, okay.