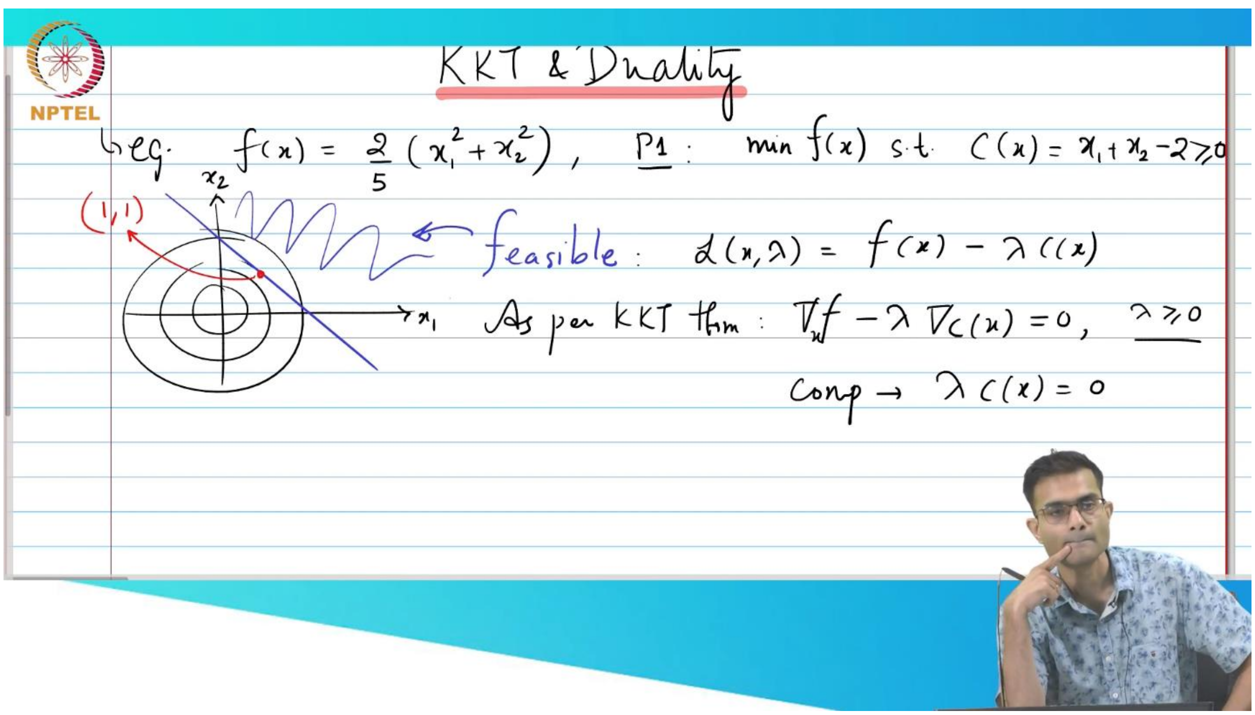
**Course Name: Optimization Theory and Algorithms**
**Professor Name: Dr. Uday K. Khankhoje**
**Department Name: Electrical Engineering**
**Institute Name: Indian Institute of Technology Madras**
**Week - 12**
**Lecture - 78**

# KKT and duality introduction

So, now we have, this is possibly the last topic of this course, also very important topic in the subject of constraint optimization. Many of you may have come across these terms as you have been reading papers. So, this is the connection between KKT and what is called duality. So, that is what we are going to explore now. This is of special significance in convex optimization, which is why a course on convex optimization would spend a lot of time on it. By the end of doing this topic, you will see why it is not as much of importance in non-convex optimization, but it is still a very important and useful concept if you want to bound things.

So following our general approach, rather than start with some dry definition, we'll take some problem, work out some geometry and get some intuition before formalizing anything. So we'll take a problem. We will solve it the way that we have usually been solving it. We'll come up to some solution.

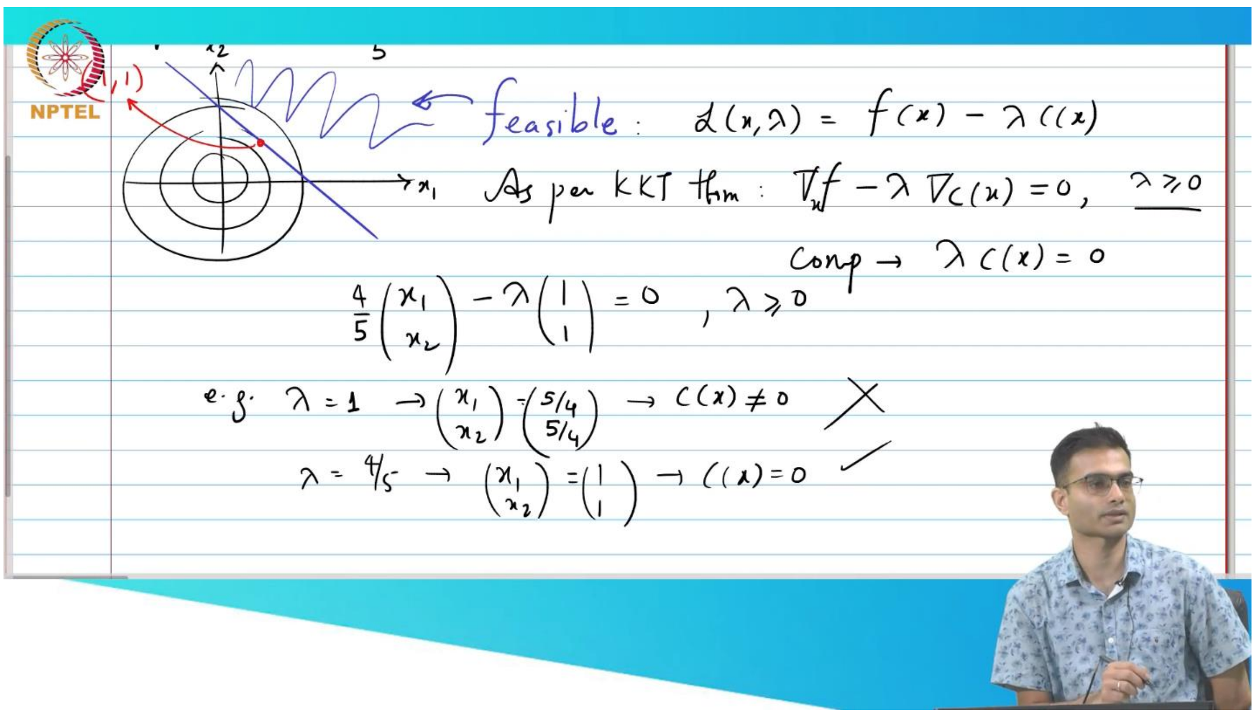



Then I'm going to introduce an alternate way of solving this problem, which is called duality. So that's the general strategy. So let's take an example. I am going to take an objective function which is actually quadratic. This is my function and problem one is I want to minimize this.

So, as before I have taken a simple enough problem that we can visualize the problem and the solution in one shot right. So, what is this, what are the contours of this cost function? Circular in the $x_1, x_2$ plane. So, if I take this as $x_1, x_2$ these are like this right and they are increasing outwards. So, in 3D this is like a bowl right, it is a parabolic equation. What is my feasible region? What kind of space is it called? It is a half space $x_1 + x_2 \geq 2$.

So, it is a line that cuts both $x_1$ and $x_2$ axis at what value? 2, right? So, it is going to be something like this. And should I be to the top or the bottom of it? Top. This is the feasible region. And we already know the solution to this. What is the solution? That circle which is tangent to this blue line and it's going to be basically this point over here, right? So this true solution, can you guess the coordinates? One, one, right? Okay.

So, it is nice to have a really simple problem because you can follow along the algebra very well. So, if we were solving this problem using our KKT business, what would we do? First step would be to form the Lagrangian, right. So, Lagrangian $L(x, \lambda)$ will be what? $f(x) - \lambda c(x)$. I just have one constraint, ok. As per KKT theorem, what would I do next? Take the gradient of the Lagrangian, ok.

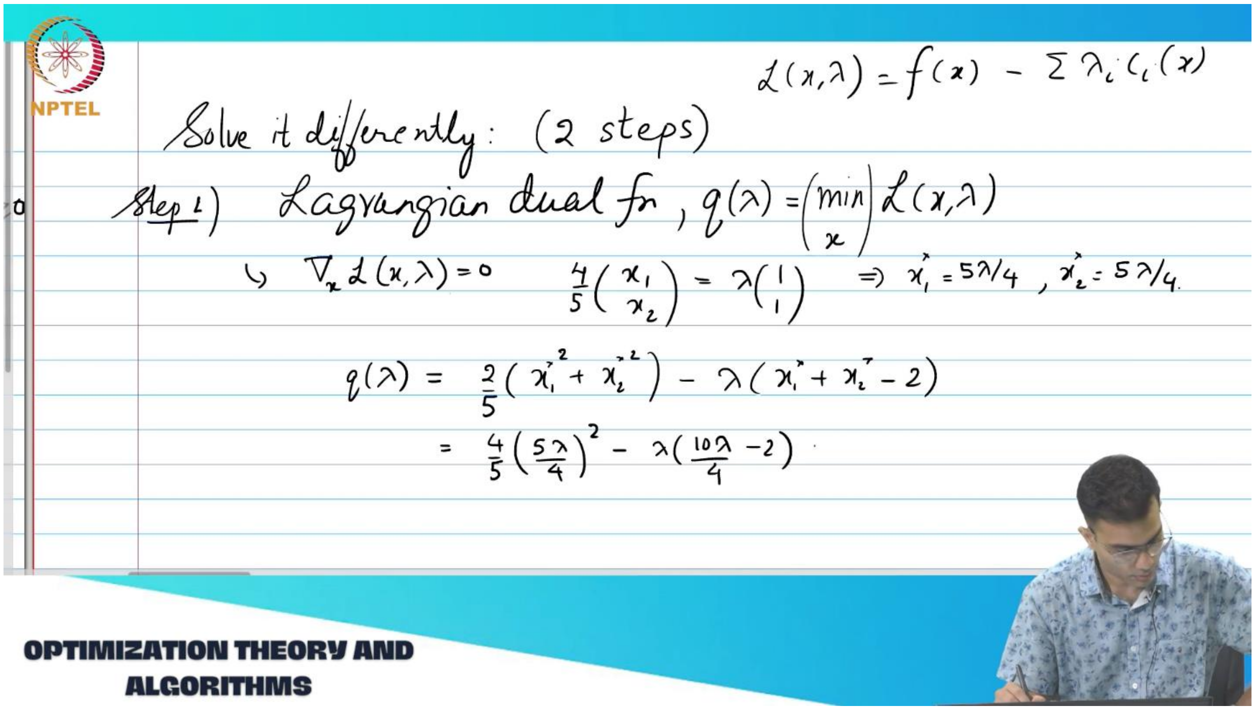



So, $\nabla f - \lambda \nabla c$ and I am going to set this equal to 0, right. What else does KKT tell me? Any condition on $\lambda$? $\lambda \geq 0$ because it is an inequality, not an equality. So, $\lambda \geq 0$. This is due to inequality. If it were equality, there is no constraint.

Complementarity will tell me, what is the complementarity condition? $\lambda \cdot c(x) = 0$. These are, this is basically what KKT theorem tells us. So, if I take the first gradient expression, $\nabla f$ is going to be what? What is $\nabla f$? $\frac{4}{5}, x_1, x_2$. So, this is $\frac{4}{5}, x_1, x_2 -$, what is $\nabla C$? 1 1 okay So, this is what we get.

Of course, this is not good enough by itself because I can take any value of $\lambda$, I can take substitute, I will get some value of $x_1, x_2$. So, for example, if I take $\lambda = 1$, what do I get? I get

$x_1, x_2 = \frac{5}{4}, \frac{5}{4}$. Is this fine? On the face of it, it looks okay or no? No, what is wrong with it? Correct, but over here if you look at $c(x)$, it is not equal to 0, but complementarity insists that $c(x)$ should be 0, because $\lambda$ is non-zero. So, this is not a good solution. I mean you can play around a little bit more, you can see that if $\lambda = \frac{4}{5}$, what do I get for $x_1, x_2$? And at this point $c(x)$ is 0, right.




So first order KKT conditions are satisfied, we have got our solution, okay. All right, so now this was the standard way of solving this problem. Let us see if we can solve this differently, okay. So now we need to add a page, okay. I am going to do this in two steps.

It is like I am giving you a two-step recipe. So we will just blindly follow the recipe to start with and see what it gives us. Step 1, I am going to define what is called a Lagrangian dual function. So it is called a Lagrangian dual function. Defined as $q(\lambda)$.

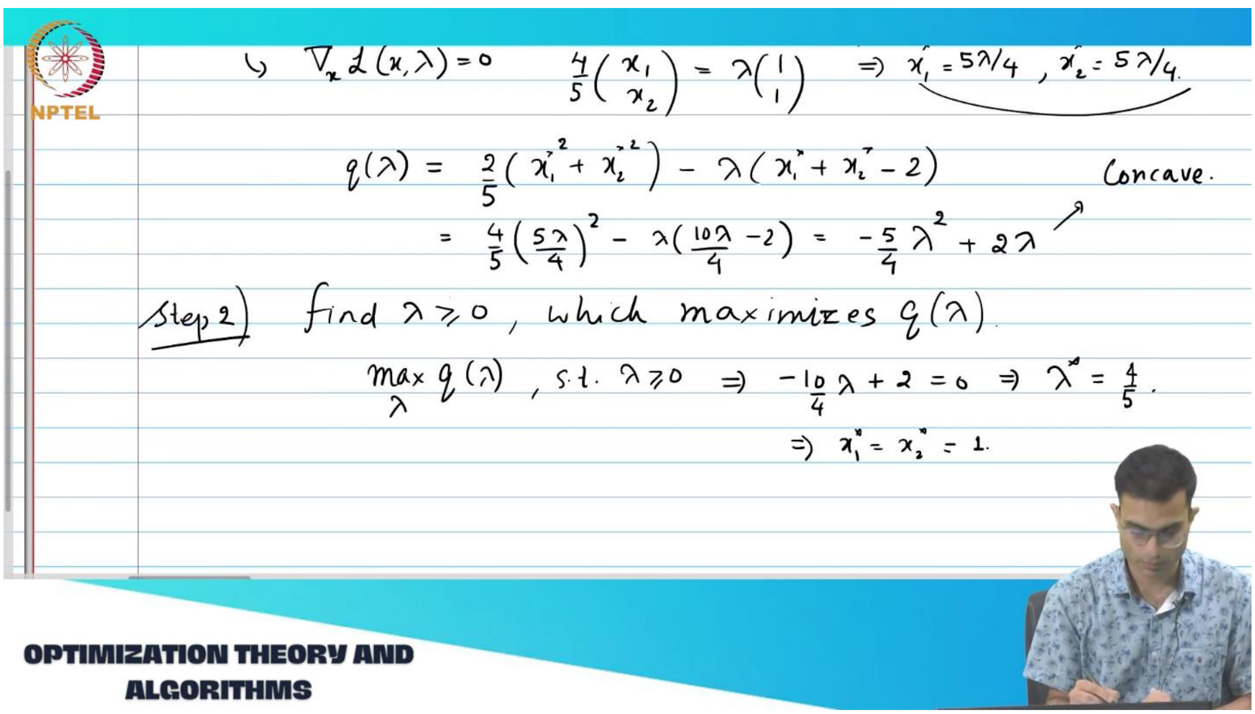
So, remember it is a function of $\lambda$ and the way I define it is I minimize over $x$ the Lagrangian. Keep in mind our Lagrangian was $f(x) - \lambda c(x)$, in general that is our Lagrangian. So, what is the first step? Is you form this Lagrangian dual function and keeping $\lambda$ fixed minimize this with respect to $x$. So, for every different $\lambda$ I am going to get something else, right? For every different $\lambda$, this minimize the Lagrangian with respect to $x$ may well give me some different expression, right? Because that is how it is, ok. So, let us try to solve this problem.

So, it is a sub problem in step 1 which I have to solve. So, let us go back to our example. So, what was our Lagrangian over here? So if I want to minimize Lagrangian with respect to $x$, what would I do? This is constraint, I mean this is almost like unconstrained optimization to start with. So I would put the gradient to 0, right? So this is what I would do. And we have already done this precise thing, so which I am going to just write this over here.

This is what I would get by setting the Lagrangian gradient 2 to be 0, ok. Now notice that this Lagrangian dual is a function of $\lambda$ which means what? I should eliminate $x_1, x_2$ from here, right? So from here I can basically write $x_1 = \frac{5\lambda}{4}$, $x_2 = \frac{5\lambda}{4}$ and I can substitute this back into my Lagrangian expression, right? So, what would I do? So, let us call this $x^* = (x_1^*, x_2^*)$. Why am I writing this as $x_1^*, x_2^*$? Because this is the $x_1, x_2$ under which the gradient of the Lagrangian goes to 0. Therefore, this expression is minimized, right? So, $q(\lambda)$.

So, $q(\lambda)$ is going to be the min part I do not need to write because I have already done it. I am left with Lagrangian, right. So, Lagrangian is $f(x)$. So, $\frac{2}{5}x_1^2 + x_2^2 - \lambda x_1 x_2^* - 2$. This was my $q(\lambda)$ with the optimum $x$.

The $x$ which minimized it is what I have put in over here, right? That is how I got rid of the min. Everyone with me on this? I got rid of the min by replacing the actual min, that is why it is an $x_1^*$, not any $x$. It is that $x$ which minimizes the Lagrangian, right? And I am going to substitute this. What kind of a function do you expect I will get in $\lambda$? Trigonometric, exponential, quadratic, looks like a quadratic, right? You can see because $x_1$ is basically $\frac{5}{4}\lambda$.




So, this is going to be $\frac{4}{5}x_1 = x_2$. So, that is how I am going to get a 2 and $\left(\frac{5\lambda}{4}\right)^2 - \lambda \cdot \frac{10\lambda}{4} - 2$. So, there is going to be $-\frac{5}{4}\lambda^2 + 2\lambda$. Before we go further, everyone agrees with this or is there a typo somewhere? Right, this is fine. Just out of curiosity, what is the shape of this curve? It is a downward parabola.

What is our immediate hint that we did not even have to graph it? Its origin is okay, but origin I can have upward opening parabola. When I take the what? The second derivative, the only one term is going to survive which is a $\lambda^2$ term and it has a negative value, it is negative curvature. So, this is a concave function. It passes through the origin and one more point.

So, and it is downward opening. So, we will just make a note this is concave. We will come back to this later. So, this was step 1. Take the Lagrangian, minimize it with respect to $x$, so I am left with a function in $\lambda$.

Now, step 2 is find the $\lambda \geq 0$ which maximizes $q(\lambda)$. So, more formally this is $\max_\lambda q(\lambda)$. Does it make sense for me to maximize a concave function? It makes perfect sense right. In fact minimizing a concave function would be disastrous.

So, this is well defined. So, to maximize this function what can I do? I can differentiate it, even for a maxima the derivative should be 0, right? So, I can take derivative. Here I do not have to take gradient, right? I just a derivative is enough because the scalar function, right? So, what will I get? $-\frac{10}{4}\lambda + 2 = 0$. This gives me what? $\lambda^*$. Now, this is a special star, right? This is the star that maximizes $q(\lambda)$. $\lambda^* = \frac{4}{5}$, ok.

And I have this relation over here for $x_1$ and $x_2$. So, when I substitute it back, I get $x_1^* = x_2^* = \frac{5}{4} \cdot \frac{4}{5} = 1$, right? Which is the same solution as we had in the regular approach, right. So, the so called regular approach which we did over here taking the KKT and setting the derivative equal to 0 blah blah blah, this is what is called as the, it is not called the regular approach, it is called the primal approach, ok. and what we just did right now in this two step formula is what is called as the dual approach, ok. So, let us just summarize this because this is very important.

So, someone is, we can call it the primal approach or we can call it the primal problem. So, the primal problem is what we have been solving in this entire course, right? Which is minimize $f(x)$, where $x$ is a variable such that $c(x) \geq 0$. This is what we have been doing the entire course. So for the primal problem, there is now a dual problem for which I have just given you a recipe without any intuition, without any proof but we will work on that.

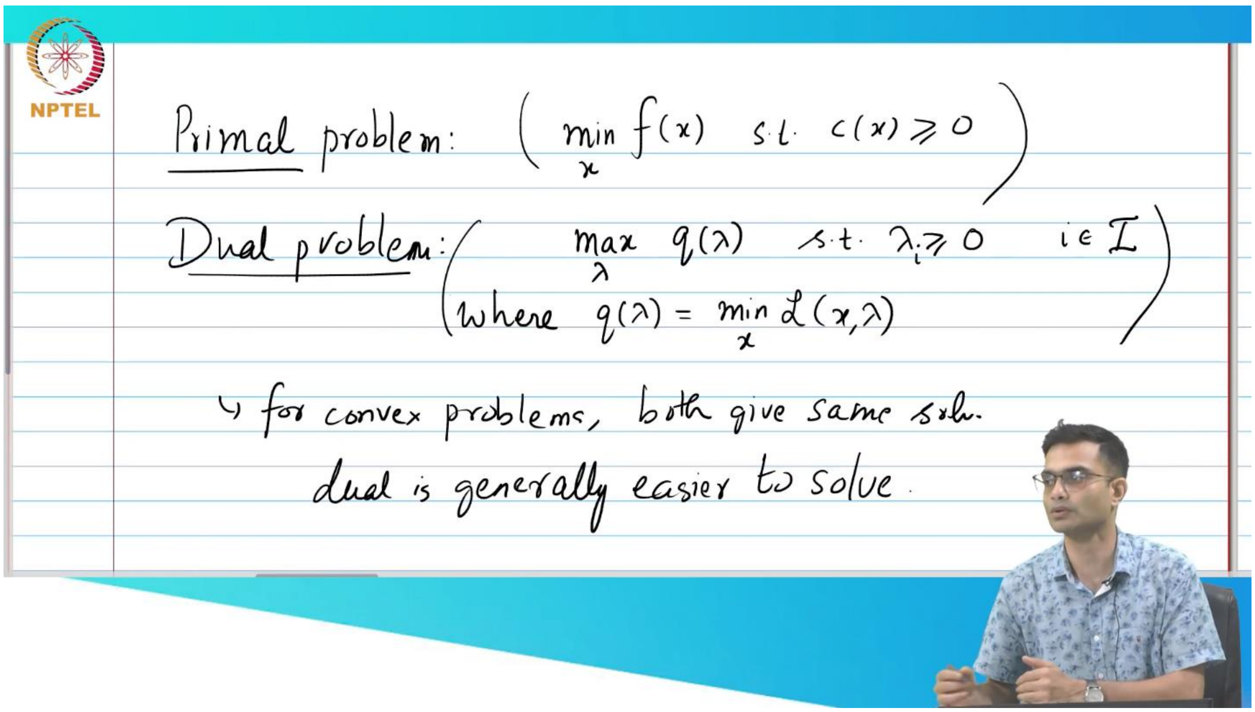
So, the dual problem is in two steps. The first step is form the Lagrangian dual. So, how do I do that? So, I called, let me write it like this. So, $Q(\lambda)$ is minimize over $x$ the Lagrangian, ok. And then once I got this, what did I do? I maximize the So, maximize the Lagrangian dual function, ok. Now this, remember this $\lambda \geq 0$ comes for the inequalities only, right.

So, whichever problem has any set of inequalities, those are the Lagrange multipliers which have to be kept non-negative. So, this is the so-called dual problem. Now you can see that in the example that we took, the primal problem and the dual problem give the same solution 1,1. So, which approach should you take? It is not at all clear. So, this switching between two approaches, as it turns out, in the world of convex functions, if your function is convex, if your constraints are convex, both of them give the same solution.

This is a very big result, right. So, let us just note it down for convex problems both give the same solution. And in many, many problems, it turns out that the dual problem is easier to solve than the primal problem. In our toy problem, both were just the same difficulty, so you can’t see it. But when you talk about more complicated problems, you will find that the dual problem is often much simpler to solve. So dual is, so I'll just make a qualified statement, generally easier to solve, and in subsequent, I mean, we'll talk about why it is easier to solve.

Okay, so this gives us a new and powerful technique of solving convex problems. The original problem, the primal problem, is hard to solve; formulate the dual, solve the dual. Dual is often supposed to be simpler. Getting a little bit ahead of myself, for non-convex problems, the solution to the primal and the solution to the dual is not the same. So, you will come across this expression in the literature, strong duality and weak duality.

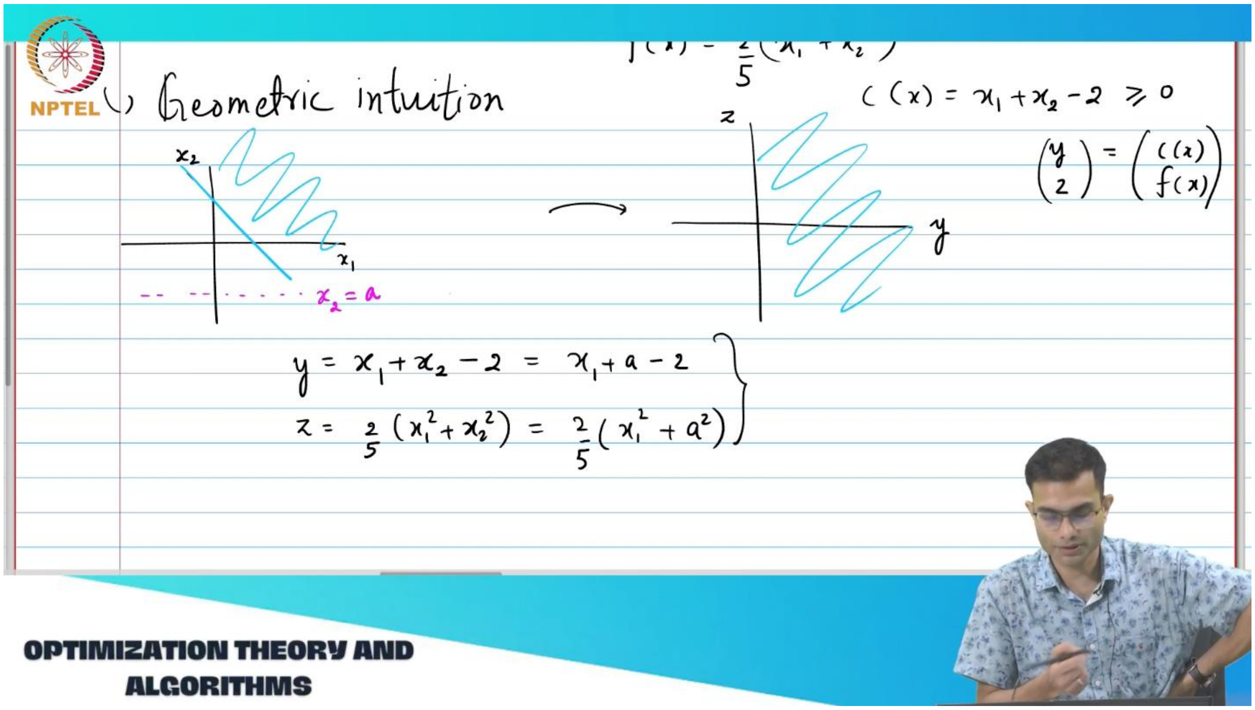
So, when the two problems give the same solution, it is called strong duality. When the two do not give the same solution, it is called weak duality. And the gap between the primal solution and the dual solution is called the duality gap. So, you will come across these words many, many times.




I will talk about it subsequently as well. But this is the basic sort of setup of a new technique of solving your original problem. As the way that we have done it, there is really not much intuition that anyone could get. It is like I gave you a recipe, we did this blindly, and we got the solution. What is the intuition behind it? What is the proof behind it? That did not come out. So, what we will do is, again rather than giving you a grungy algebraic proof, we will sketch some diagrams so that we get some intuition.

We will prove that later. We will come to it. So before we get into proofs and properties, I want to give a geometric intuition and then we will prove that $q(\lambda)$ is concave. It is not very tough. So, we are going to sort of give geometric intuition as to why this two-step process of the dual problem actually worked. On the face of it, it is not at all clear why it should work. So, I have my $f(x)$, just want to use the same problem, $x_1^2 + x_2^2$, and I had my $c(x)$ was $x_1 + x_2 - 2$ and I wanted this greater than or equal to 0, okay.

And this is my $x_1$, this is my $x_2$, so the feasible region over here was this. Everyone agrees, right? Now, what I am going to do is, I am going to create a map. We are going to go from the $x_1, x_2$ coordinates to another set of coordinates. I am going to call them $y$ and $z$, okay. And how do I define this $y$ and $z$? I am going to define $y$ and $z$ in a very simple way.



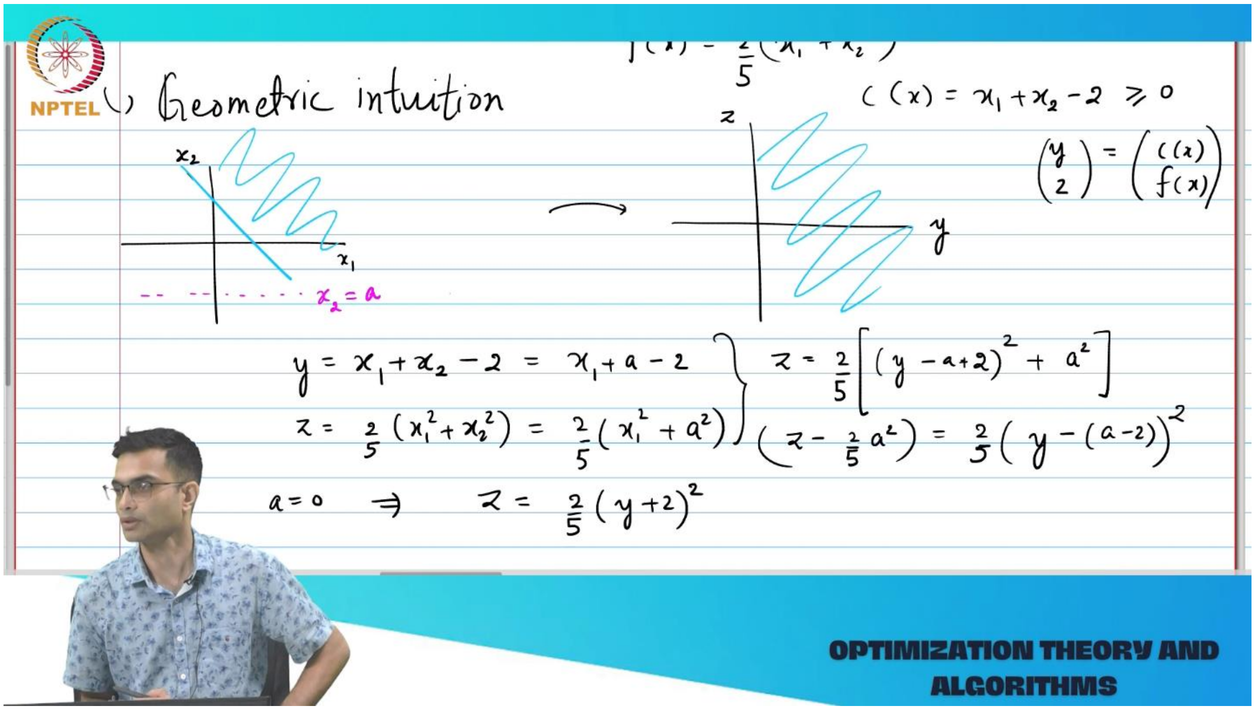

So $y$ is going to be equal to $c(x)$, $z$ is going to be equal to $f(x)$. I can do all of these tricks only in a problem of two variables; otherwise, it will be impossible to sketch this. So I have $y$ and $z$. So let us proceed, and this map, this is a map, I am going from the $x$ system to the $y, z$ system. Yeah, because it is not a very straightforward map because $f$ is quadratic, $c$ is linear.

It is not that every point is going to get mapped to a point, one-to-one kind of a thing, right? So let us proceed step by step. So let us look at the feasible region. The feasible region on the left-hand side is the blue shaded part. Everyone agrees, right? That is quite clear. In the new map, $yz$ map, where is the feasible region? Everything on the right-hand side because $y = c(x)$ and $c(x)$ should always be greater than or equal to 0.

So, $y \geq 0$ is the right half-plane, right? So, the feasible region first of all got mapped over here, okay. Then what do I have? This is not enough obviously, I need to also figure out how, you know, where is the, what are the contours of the function, etcetera, right? So, let us start that analysis. For example, let me consider a line $x_2 = a$. So, let us try to see that if I take a line in the $x_1, x_2$, what does it get mapped to? There are various ways of getting this mapping.

This is, it turns out to be a simple way. So, if I take $x_2 = a$ as a line and I now need to get some relation between $y$ and $z$. Then that will tell me a line in the $x$-plane got mapped to blah blah in the $yz$-plane, okay. So, I need a relation between $x_1$, sorry, between $y$ and $z$. So, let us take, let us proceed with that.

So, $y = x_1 + x_2 - 2$, right. What is the value of $x_2$? $a$. So, this will become $x_1 + a - 2$. What is my $z$? $\frac{2}{5}x_1^2 + x_2^2$, right? Which is equal to $\frac{2}{5}x_1^2 + a^2$. From these two equations, is it possible for me to somehow eliminate? I want a relation between $y$ and $z$, so that I can make a graph in the $yz$-plane.




How would we basically eliminate $x_1$? So, I can take $x_1$ and write it in terms of $y$ and $a$ and put this into the second equation. So, you would get $z = \frac{2}{5}(y - a + 2)^2 + a^2$. So, this is what I wanted because $a$ is a number and now I have a relation between $y$ and $z$. What does this relation look like? What kind of a graph is this? It is a parabola. For different values of $a$, what is changing in the parabola? The lowest point of the parabola is shifting.

Anything else that is shifting? The offset is also changing, right? If I take this plus $a^2$ on the right-hand side, if I take it to the left, I will have, I mean, the classic equation of a parabola, and it will look like this:

$$z - \frac{2}{5}a^2 = \frac{2}{5}(y - a - 2)^2$$

So, when I plug in different values of $a$, $z -$ some positive quantity, right? The parabola is shifting up or down? It is always shifting up. For every time I put $z = \frac{2}{5}a^2$, I am going to get 0 there. And therefore, I can choose, basically, I can get that point, right? So, I have these parabolas that are always, I mean, they are always increasing.

Okay, let us take a simple value $a = 0$. What is the equation of my parabola?

$$z = \frac{2}{5}(y + 2)^2$$

So, this is a parabola. Where is the center of it, or rather the lowest point of it? $(-2,0)$. So, you can substitute. When I put $y = -2$, what is the value of $z$? Zero, right? So, this parabola actually looks like this. Everyone agrees? This one line $x = a$ has been mapped to one parabola.

So, it is not such a straightforward transformation also where it is mapped over here. Now, when I substitute different values of $a$, we know that this parabola can never shift below. It is always going to shift up. So, it is shifting up. But along with shifting up, it is also shifting which way? Right side or left side? Depending on $a$, right? The center is $a - 2$, right? So $a - 2$ can float around this way.

But the point is it always stays above. So, when you do this exercise, it turns out that the entire, for all possible values of $a$, when you look at what all are the different possible parabolas, you would map the entire $x_1, x_2$ region into some region on the right-hand side. So, you can take my word for it that that entire region ends up being all of the parabolas somehow contained within this one parabola. You can sketch it and see, it is not so obvious, but I will just take it on faith. So, that is why I was at least showing you that the parabola never goes down.

The center of the offset is always positive. So even if I give a negative $a$, it is a squared by, $a^2$ into $\frac{2}{5}$, it is always staying up there. So okay, this is what the map looks like. So, now let us check the geometry. Is everyone clear with the geometry? Little bit, you are taking on faith that everything is contained inside this upward bowl. Intuitively, it makes sense.

So, now I am going to ask you a question that is pure geometry. So, I am going to ask. Pardon me. $A = (-2,0)$. You agree that this is, at $(-2,0)$, passes through this equation, $y = -2$, $z = 0$.

So, that is the bottom point and then it opens up, positive curvature. Correct, because as I slide $A$, it also moves up. So, all of these parabolas end up being sort of contained within this. Intuitively, because there is an $a^2$ there, okay.