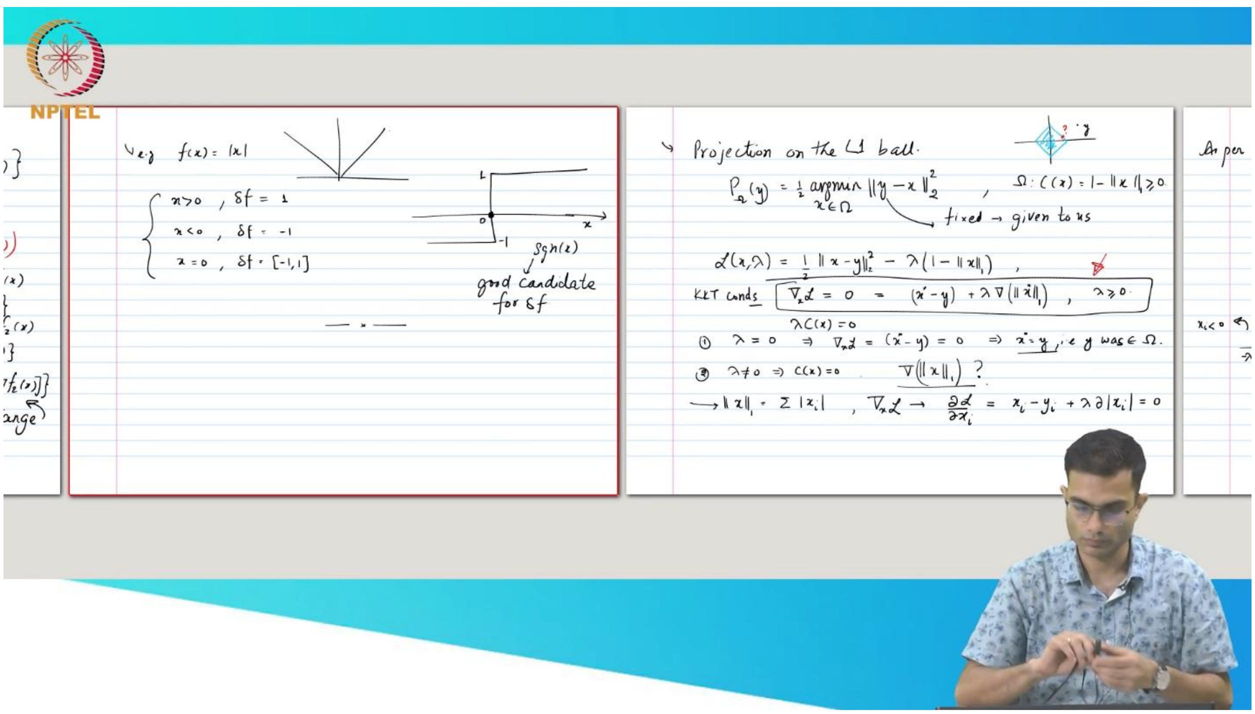
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 11
Lecture - 77

Recap of Projection onto l1 ball

I thought that before we start today's topic, I will just go over yesterday's L_1 projection stuff in case things are not clear. We were looking at sub-gradients, then sub-differentials, and then the L_1 projection. So, the concept of sub-gradient was, I guess, fine to everyone, right? Supporting hyperplane. I think the trouble possibly was when we began to solve the projection on the L_1-ball, right? So, let us just go over it step by step. I want to make one small correction over here.

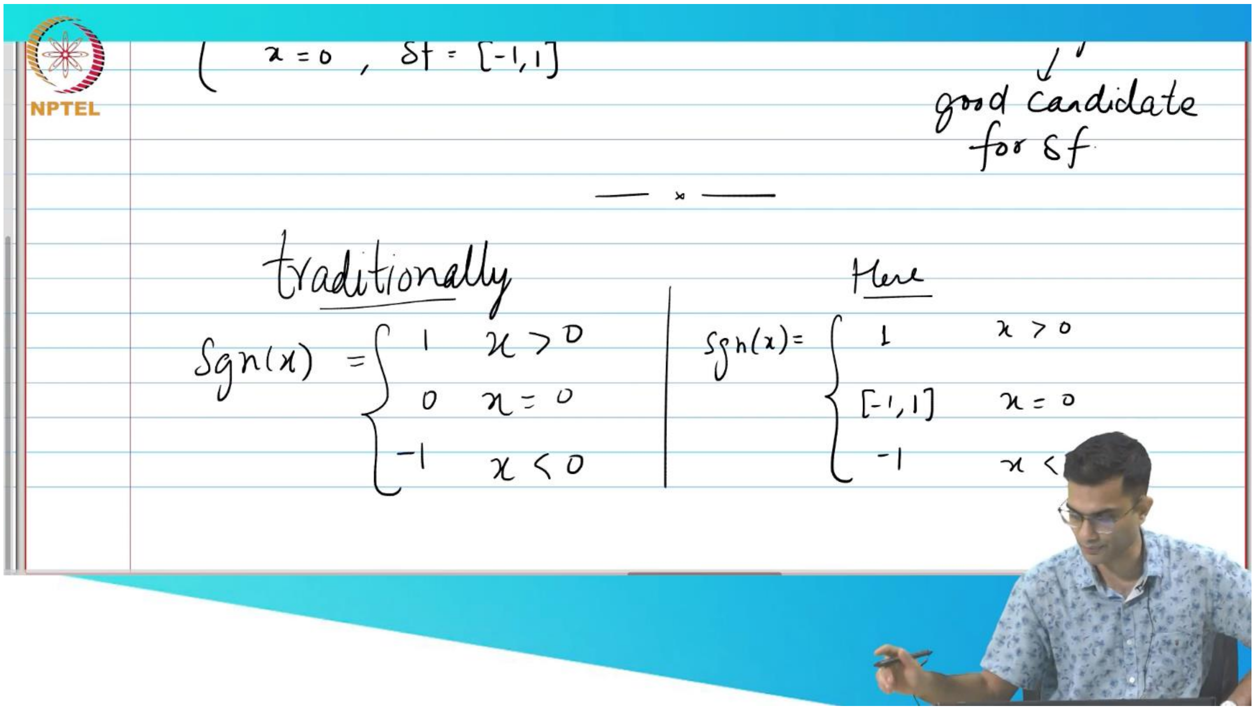
So traditionally the signum function, let me just make a note over here. This is what I mean, this is when people say the signum function, this is what they usually mean, value of zero. Now in the derivation which we have done yesterday, the signum function has a different, what should I say, meaning. It is much, much closer to the sub-differential, which is -1 to 1. This will clear a lot of the confusion that we were having.



So, because of this, the graph of the signum function looks like, I will use a highlighter. This is what the graph of the signum function looks like. So, this is different from the graph of the signum function which all of you are used to from signal processing where signum of 0 is equal to 0 by definition. Here we are making, we have modified the signum function to have a different behavior at 0. Technically speaking, this is not a function.

Why is it not a function? Because it does not have a unique value at $x = 0$, it takes a range of values, right? So, it is a generalization of a function. So, this is what we mean by the signum function for this discussion. If it bothers you, you can call it by some other name, the sub-differential version of the signum function, okay?

So, then we looked at the L_1 projection, the cost function, the Lagrangian became this, that was fine, right? And then when we began to take derivatives, that is where these various options come up, right? So, if I write the L_1 norm in this fashion, take the Lagrangian, $\frac{dL}{dx_i}$ is what I have to calculate.

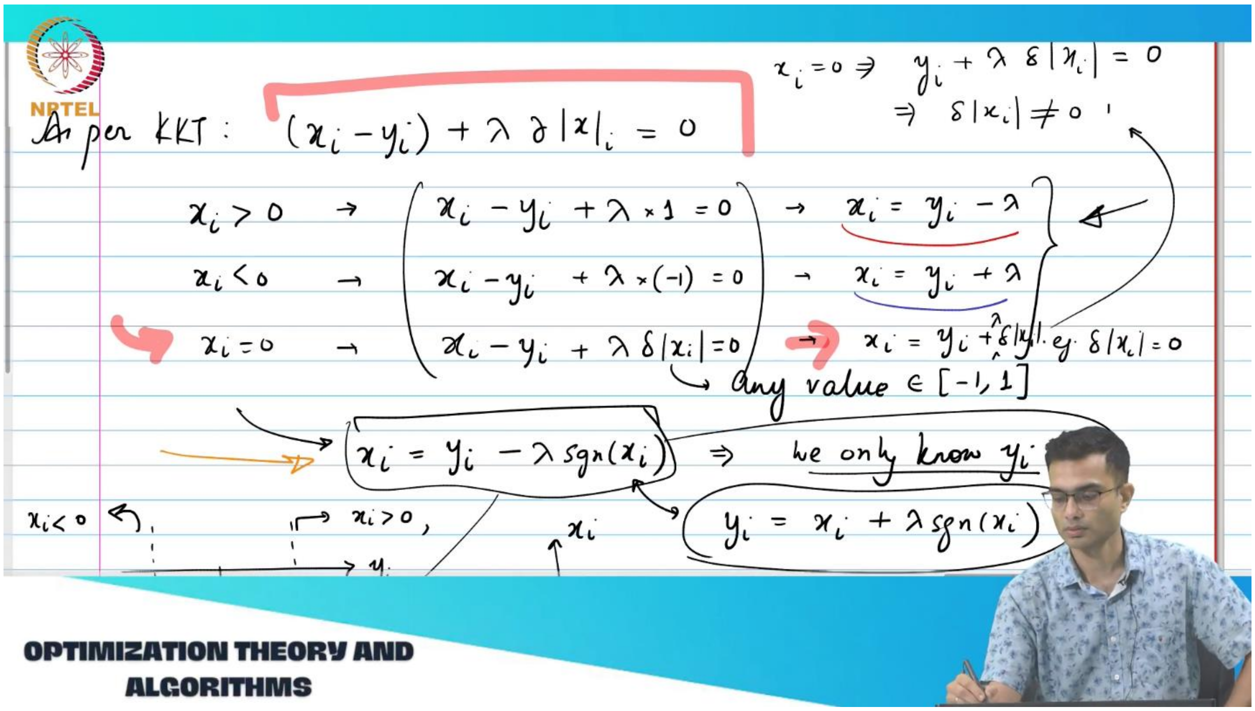


So I am left with the sub-differential of $|x_i|$, okay. And here is where the various cases come into play, okay. So, all I am doing is I am taking this expression and I am taking different cases for x_i. When x_i is greater than 0, one by one, these cases come up. The controversial or the slightly tricky-to-understand case was when x_i was equal to 0, right? Now $x_i = 0$, this is the third expression over here that I get, $x_i - y_i + \lambda \times$ sub-differential of $|x_i|$.

And the sub-differential, as we said, is any value between -1 and 1. So when I move, keep x_i to one side, move everything else to the right-hand side, remember we have to always keep in mind that $x_i = 0$. That is the only case when this is so, therefore I am left with $y_i + \lambda \times$ sub-differential of $x_i = 0$. We are just blindly substituting and seeing what comes out. So, therefore this implies what? That my choice of the sub-differential is not, I cannot choose 0. Because if I choose the value of the sub-differential to be 0, then I cannot, y_i is being given to me in the problem. y_i could be like -6 or something, right? Then this thing will not evaluate to 0, right? I will get $6 + \lambda \times 0$, which is not equal to 0.

So, therefore, I have to choose the value of the sub-differential to be what? $\frac{-y_i}{\lambda}$. So, I cannot choose 0. So, therefore, let me write it explicitly in general. So, that is what we have to keep in mind. And so, when I wrote this in compact fashion over here, in compact fashion over here. This, if you look at it a little bit more carefully, this use of signum x is the updated definition of signum.

Then it makes perfect sense. Because now signum, when I substitute $x_i = 0$, signum x_i is not 0. It is the sub-differential, and whatever it is. Then we sketch the graph, right? And then this entire range that I have over here from, what are the values over here? This was from λ to $-\lambda$. In this entire range, x_i remains 0 and y_i basically was, I chose the value of the sub-differential to match y_i.



That is basically it. So, this function looks like this. For $y_i < -\lambda$, I have the blue line, for $y_i > \lambda$, I have the red line and in between, x_i remains 0, okay. Once I have done this, the rest of it all, the rest of all is just algebra. I just looked at this graph and I rewrote this in expression b.

I need not have, by doing so it just makes my algebra a little bit more convenient, okay. And then we looked at a few just examples of how to use this expression that we got. So, in this part, are there some points which are not clear? So, the only update is that keep in mind that this is our definition of the signum function that we are using here. At 0, we are going to let it be free, between -1 and 1. Okay, so let us switch.

Yeah. I mean, it is a set of all vectors in between those two gradient values. Yeah, you have to make sure basically what is the test? Whatever linear combination you take, it should still be a supporting hyperplane. So, the definition we had for sub-gradient, right, if you look back at this definition here. This definition, this should hold. So, here you can substitute your value of the gradient, right, and verify that this is happening.

So, for example, you can take $\alpha_1 \nabla f_1 + \alpha_2 \nabla f_2$, substitute into this, come up with some relations on α_1, α_2. That is the way to do it. You should always go by the definition. The definition is supporting hyperplane, make sure that your candidate g satisfies that.