Distributed Optimization and Machine Learning

Prof. Mayank Baranwal

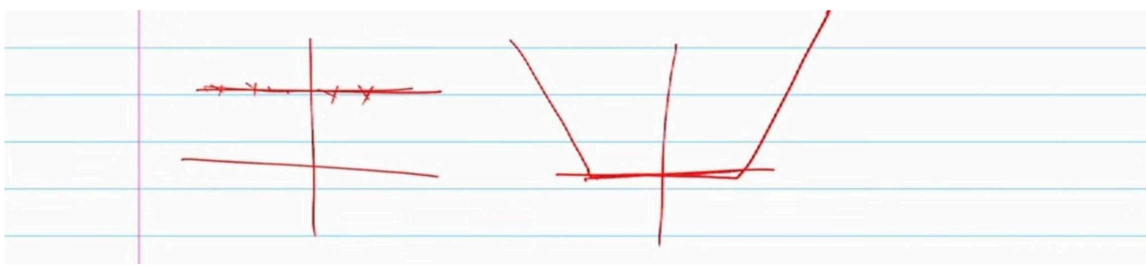
Computer Science & Engineering, Electrical Engineering, Mathematics

Indian Institute of Technology Bombay

Week-1

Lecture - 3: Course Outline

How this course will be structured? So, we will first start with the just a brief course outline. So, we will start with convex optimization. And for the major part of this course we will just focus on even for the distributed setting we will largely focus on convex optimization except for functions which satisfy PL inequality where you can generalize it to some slightly non-convex function like x square plus 3 sin square x. We will largely be focusing on convex optimization. So what is one specific property of convex functions? Why do we care so much about convex functions or convex optimization? unique global



minimum right. So, is a constant function a convex function? Is this a convex function something which is a function which is just constant? What about this function? Is this a convex function? So all of these are convex functions.

But what is the property of a convex function? So in this case you have multiple minima. So if you have a local minima, it also happens to be a global minima. So that is the property of a convex function. It's fine that you do not arrive at x equal to 0.

Maybe you arrive at x equal to 1. But if that happens to be a local minima, it also happens to be a global minima. So at each point, the optimal value remains the same. So all these are globally optimal solutions. Even in here, like every point on this line, it's a globally optimal solution.

Not just locally optimal but also globally optimal. So the global optimal value remains the same. Every local optima is also a global optima. So that is a property of convex optimization. So that means something like this for instance is not a convex optimization.

So this is a local minima, this is a local minima, this is a local minima but only this particular local minima is your global minima. right. So, in this case you cannot say that like let us say you have found this particular local minima, you cannot say that we have arrived at the globally optimal value of the function. In this case you can say right, here you cannot say that you have arrived at the globally optimal value of the function. So, that is why convex optimization is important because you can guarantee like if you have arrived at a locally optimal solution, you can guarantee that it is also going to be globally optimal.

Whereas with non-convex functions we can only provide local guarantees that this is I mean we have arrived at the locally optimal solution cannot say much about the global optimality is this clear? So, after convex optimization we would move to. So, basically we would look at both constrained as well as unconstrained optimization. So, we will look at the constrained convex optimization. and after this is when we would have a slight departure and we would start studying Lyapunov stability theory. So, this would be important from the point of view of analyzing and designing faster algorithms ok.

So, we are going to look at largely going to look at something called fixed time stability theory. It also happens to be one of the areas that I work in. So, where you can guarantee that no matter where the where you start you are guaranteed to converge to the optimal solution in a fixed amount of time. So, that is initialization independent guarantees on how quickly you can converge to the optimal solution. So, all of this again is going to be in continuous time since we are largely going to be invoking Lyapunov stability theory for continuous time dynamical systems.

We would be designing new optimization algorithms which are again going to be in continuous time, but then we would look at some guarantees that we can provide when we try and discretize those continuous time dynamical systems. we would then move to the distributed aspect of this course. So, we will start with some basics of graph theory. So, things like Fiedler Eigenvalue or connectivity of a graph and so we will study those concepts and we will then conclude the course with designing algorithms for distributed optimization and depending on the time remaining we may or may not study federated learning. So, that that is something that we will have to look at ok.

So, this is something that we will decide towards the later half of the course. So, but this is how the course is going to be structured and most of the content that you are going to be seeing in this course has been developed over the last 6 to 7 years. So, it is pretty topical. So, there is no single reference for this particular course. So, it will largely be the lecture notes and some research papers that I will be pointing you guys to.

But then the advantage is that whatever you're going to be learning, it's going to be I mean, basically help you a lot if you are planning for grad school or otherwise. I mean, large-scale learning or distributed learning is anyway a very hot topic these days. And so much of the tools is you're going or at least a theoretical aspect of those tools you're going to look at in this course. Thank you.