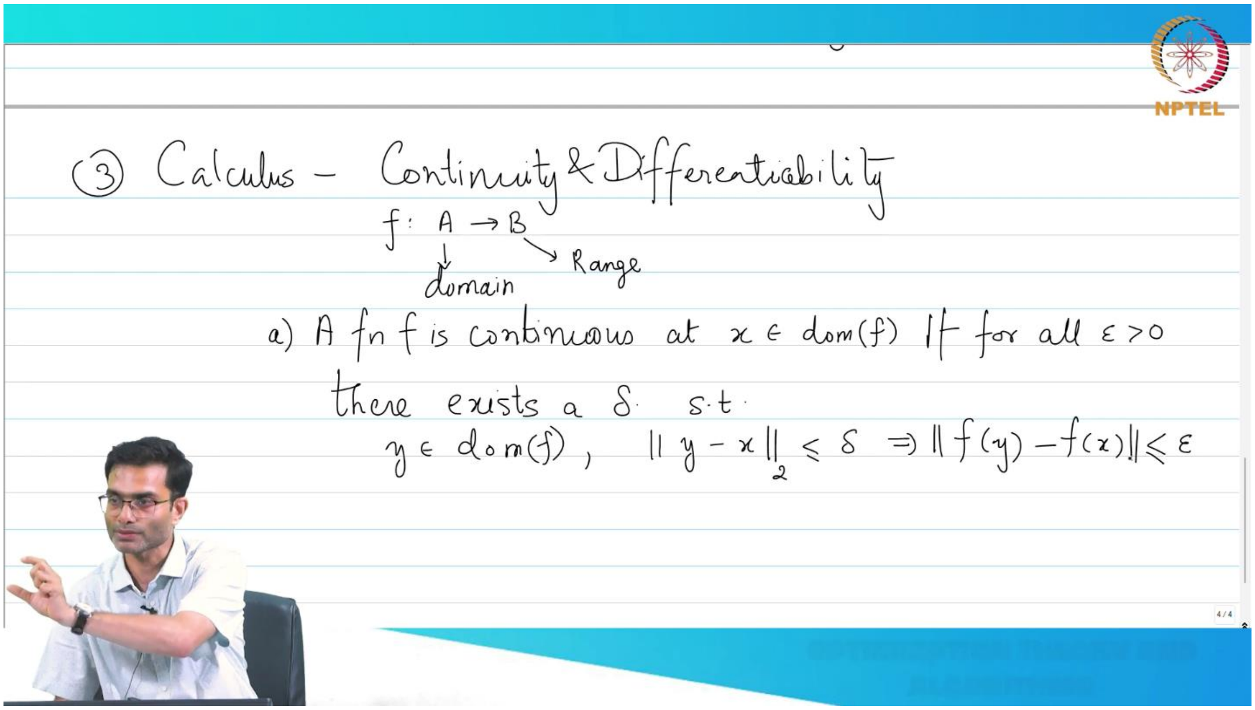
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 02
Lecture - 09

Summary of Background Material - Calculus 1

Yes, question. You have to check it. In this case, the function is defined over the entire real $\mathbb{R}^n$. So, it is already convex, right? But in general, your function may have a limited domain. So, you would have to first check because when I wrote down the definition of a convex function, point A was that the domain must be a convex set. So, for example, if I take this function which we tried to describe and I chose this domain over here, you do not even have to evaluate that convex combination; you can look at the set and say that this is not going to be a convex function.

In this example, I assumed over here x and y the entire $\mathbb{R}^n$ was where they lived. So, it is already convex, but you have to check it, you are right. Yeah, in this case or in general? No, you have to prove it for all points in the domain.



Yes, but okay. So, he is asking a good question here. I have taken the convex combination of two points x and y. Now, he is saying what if I take a convex combination of n points? Do I have to prove it for that also? Any thoughts on this? Is the question clear? I have I am only showing it for convex combination of two points; the question is, but what about n points? Let us see where I should write this as long as

$$\sum_{i=1}^{n} \alpha_i = 1$$

He is saying it is a convex combination of n points. So, well, I can suppose I have

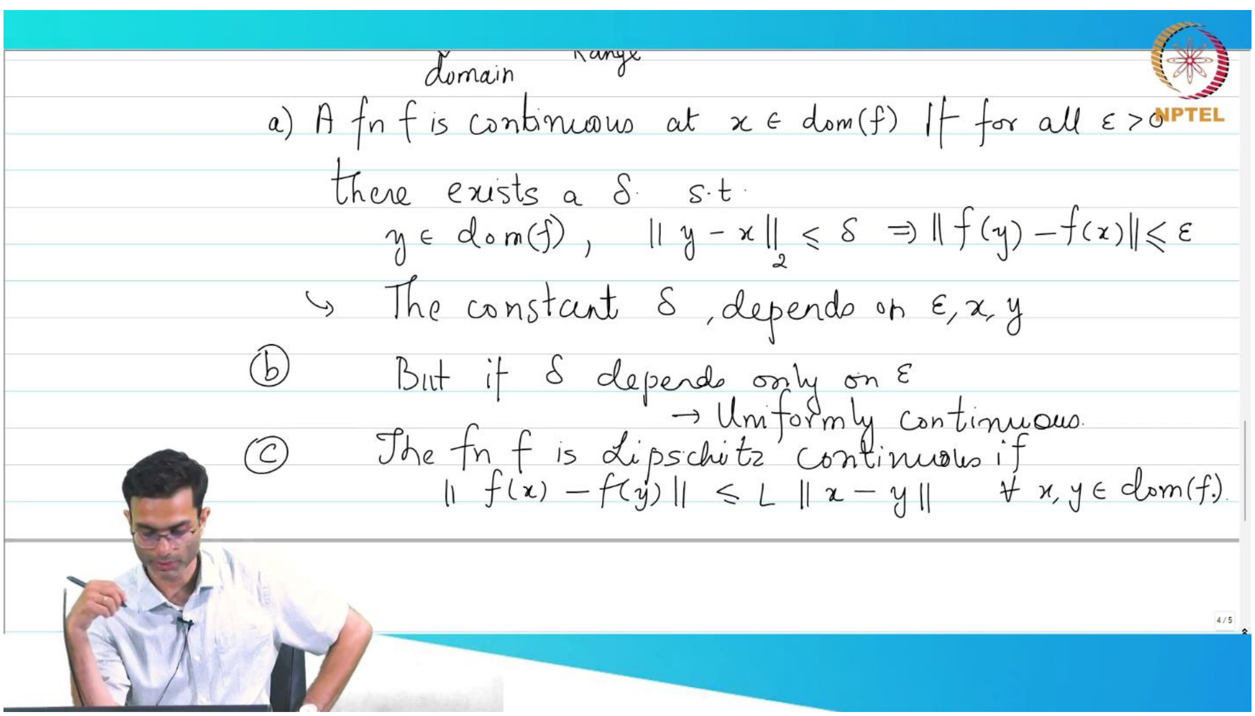
$$\sum_{i=1}^{n} \alpha_i \, x_i$$

I can write this as

$$x_1 + \sum_{i=2}^{n} \alpha_i \, x_i$$

from here. Right? Now, is it true that so, I can take this as one new point, right? And this is another point, right? Now, all that remains to be proved is that the coefficients still satisfy the convex combination, and you will find that it does. So, two points is enough, okay? It is a good point. Did everyone follow this? n points I can split as 1 and $n-1$, and it works for that and any combination like.

Again, we will start with something very basic. Everyone is familiar with the notation of a function. So, if I write f like this, what does it mean? It is a map from domain A to what do we call it? We can call it co-domain or range. So, this is the domain. What entity is it? Is it a function, a combination, a set? What is it? Just a set.



So, just as we had here are things you know things get a little bit more broadened and generalized compared to what you studied in high school. So, in high school, basically class 11,12, you had a definition of continuity that was quite straightforward. When we come to n-

dimensional space, there are actually three different definitions of continuity that we have to talk about. Okay? So, let us list those down.

So, a function f is continuous at x that belongs to the domain of f if for all $\epsilon > 0$ there exists a δ such that y belonging to the domain of f, the distance

$$\| y - x \| < \delta$$

implies that...

This sounds like a complicated looking expression, but it is quite straightforward. If you can find two points that are close by, then the function values that they map to also are close to each other. So, that is our intuitive sense of continuity, right? That if I change x a little bit, the change in the f value is kind of, it is not exploding; it is bounded, right? But it is done in this formal language that for all $\epsilon > 0$ there must be some δ such that this inequality is satisfied. So, that means the distance, the norm of $y - x$ simply refers to the distance between two points. So, the distance between two points x and y if it is less than δ, I should be able to map it to a distance between the function values. So, this is the standard definition of continuity.

This is where you would have ended in your high school kind of course, right? Now, things get a little bit more interesting. So, now this constant that we are talking about δ, okay? The constant δ does it depend on $\epsilon, x,$ and y? That is the distinction, okay? In general, δ depends on $\epsilon, x,$ and y; this is the most general case. And when this happens, we have called it continuous, but if δ depends only on ϵ, then this is another type of continuity; this is called uniformly continuous. I will take an example in a few moments to clarify this, okay? But at a high level, do we follow what is going on, right? Does this constant depend on where I am evaluating the inequality? Yes or no? Yes, it is continuous. No, I am saying I am making a stronger statement; I am calling it uniformly continuous.

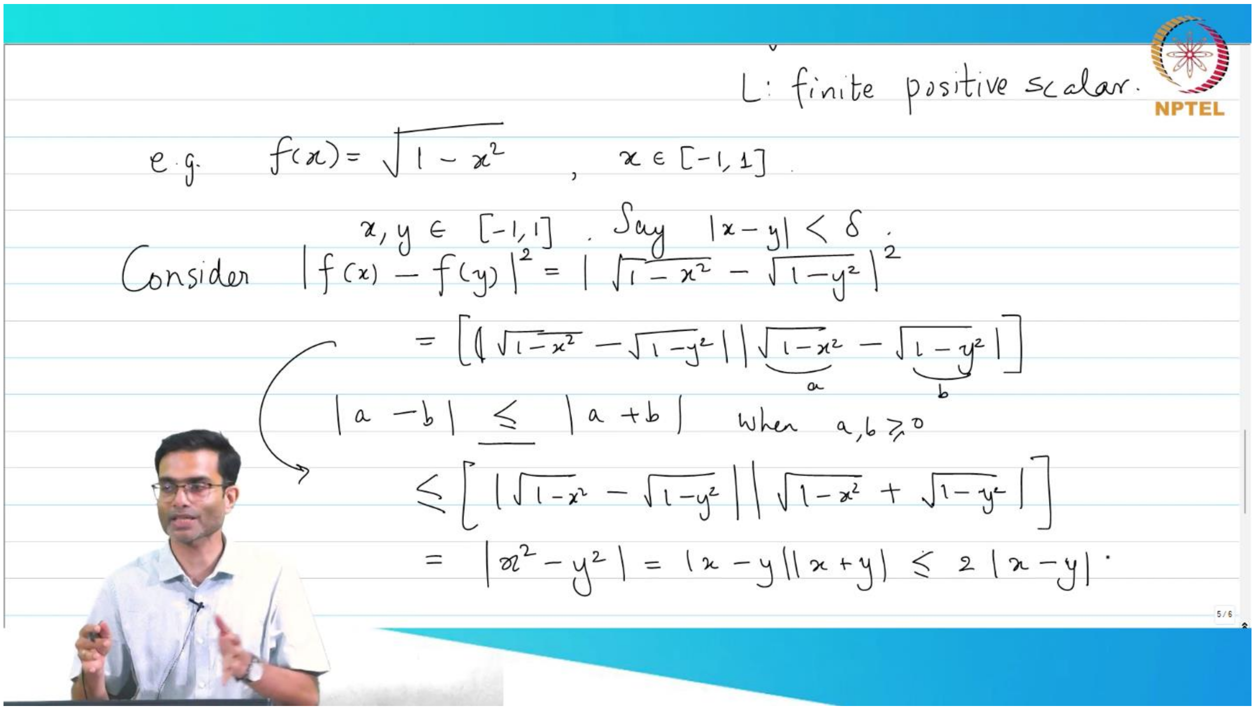
Finally, c is something which we will also use quite a bit in the course, which is called Lipschitz continuity. So, the function f, so I am going to take this distance $|f(x) - f(y)|$, okay? Is bounded by some constant L times the distance between x and y, okay? For all x and y belonging to the domain of f, okay?

So, notice the slight difference in flavor with definition a or definition b. In the definition a or definition b, the distance between the function values was just related to a constant δ and ϵ. But here, what are we saying? This distance is sort of proportional to the distance between the points x and y. Notice $\| x - y \|$ is appearing on the right-hand side here; it did not appear earlier, right? So, this is a difference, and this is called Lipschitz continuity.

So, why are we talking about these three different types of continuity? The thing is that if you are interested in going into analysis of optimization algorithms and trying to prove rates of convergence, these properties become important. If you are only interested in implementing an algorithm and getting your results and are happy to take someone else's report that this has super-linear convergence, this has quadratic convergence, you need not bother about this. L is just a scalar, finite scalar, yeah, finite positive number, okay, yeah.

It is not quite the same, right? You are saying that if I divide, you will get something like ϵ/δ over there, right? The trouble is that this ϵ and for example, δ can depend on x and y, right?

Then your constant of proportionality will become a function of x and y, but here we are saying that it is a positive quantity.



To clarify this, let us take a concrete example, and we will show you where a function is b, but not c, or some combination of that, that would clarify. Just looking at the definition is a little confusing, I agree. So, let us take an example. Let us take

$$f(x) = \sqrt{1 - x^2}$$

and to keep life in the real world, I should restrict x to where? -1 to 1. This guarantees that f maps from $[-1,1]$ to $\mathbb{R}$; otherwise, I would get complex numbers, okay?

And now we want to figure out what type of continuity this function satisfies. Now, $\sqrt{1 - x^2}$, intuitively, if you graph it for example, it looks like a continuous function, but which precise type of continuity does it satisfy is something that we want to see, okay? So, the usual procedure in all of this is take two points and take their difference and the difference of their function values and then see what happens, which way do the inequalities follow, right?

So, I am going to take x and y as two points that belong to $(-1,1)$, okay? I am going to choose, let us say that this is some point here $x = 0.4$ and let us say $y = 0.5$, okay? So, let us compute

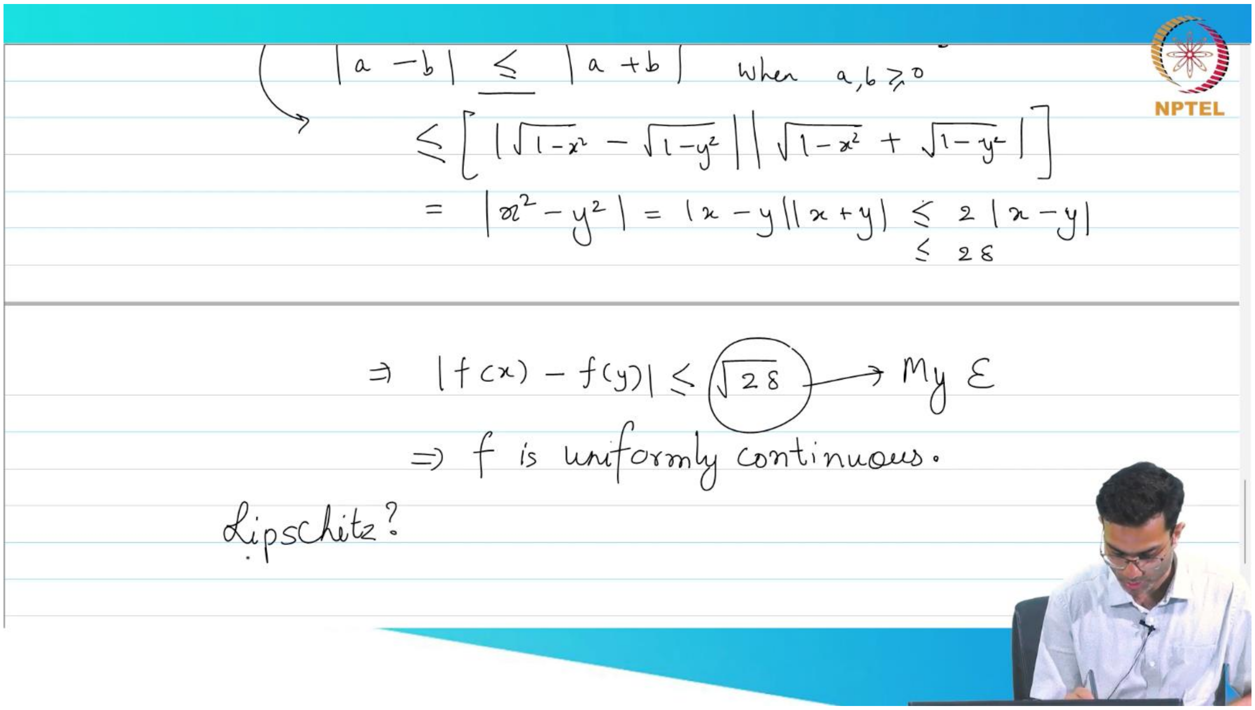
$$f(x) - f(y) = \sqrt{1 - (0.4)^2} - \sqrt{1 - (0.5)^2} = \sqrt{0.84} - \sqrt{0.75}$$

If you do that calculation, you will see that they are not too far apart, but still, we can compute it explicitly if we wish.

So, we can also see this from geometry, right?

So, as x approaches 1, the values also go to 0 over there, and likewise, as you go to -1, the distance between 1 and -1 is finite, and everything is fine over there. So, $\sqrt{1-x^2}$ is continuous everywhere; in fact, I can even talk about it being uniformly continuous. Now, if you took $1+x^2$ for instance, can someone tell me what kind of continuity does that possess? What do you think?

ww



Okay, this is quite obviously Lipschitz continuous, right? Because if you look at $|f(x)-f(y)| = |x^2-y^2|$ and if you just factor it out, you will get

$$|x-y||x+y|$$

So, clearly, if $\| x-y \| < \delta$, then this is just going to be a finite number, K, multiplied by $\| x-y \|$ for K as a constant, okay?

So, I could be Lipschitz continuous as well, and it is both continuous and uniformly continuous; there are many functions that fall into all three categories of continuity.

So, as I said, things are becoming a little more interesting. Let us take a step back, and I am going to introduce the notion of the differentiable function; it is much easier.

Let us say you have a function $f: \mathbb{R}^n \to \mathbb{R}$. It is differentiable at a point a if

$$\lim_{x \to a} \frac{f(x) - f(a) - \nabla f(a) \cdot (x-a)}{\| x-a \|} = 0$$

And intuitively, what this is saying is that as x approaches a, the difference in function values can be approximated by the linear term $\nabla f(a)(x-a)$, okay?

And this is your usual statement of differentiability where you have a smooth function, and this is essentially saying that the slope at that point approximates the function value. So, we will be using this as well.

Now, let me tell you what we want to prove from here on; so we will have these three types of continuity and differentiability. It is easy to see that continuity implies differentiability. However, uniform continuity is not necessarily differentiable, but Lipschitz continuity is equivalent to differentiability at that point.

And now we will go on to say that differentiability implies continuity.

We are now going to summarize what we have learned so far.

- If a function f is Lipschitz continuous, then f is continuous and differentiable.
- If a function f is uniformly continuous, then it is continuous.
- If a function f is differentiable at a point, then it is continuous at that point.

This will give you the tools to start optimizing these functions using numerical techniques.