INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

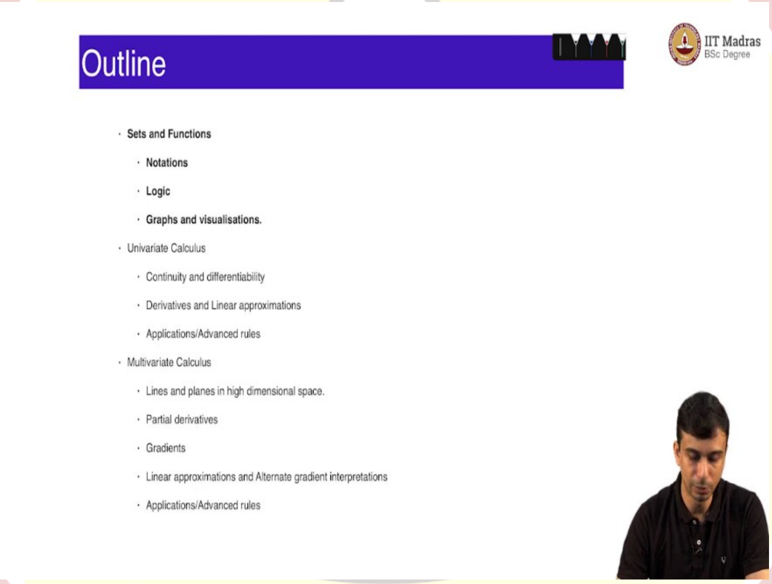
IIT Madras

ONLINE DEGREE

Machine Learning Foundations
Assistant Professor. Harish Guruprasad Ramaswamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras
Sets and Functions

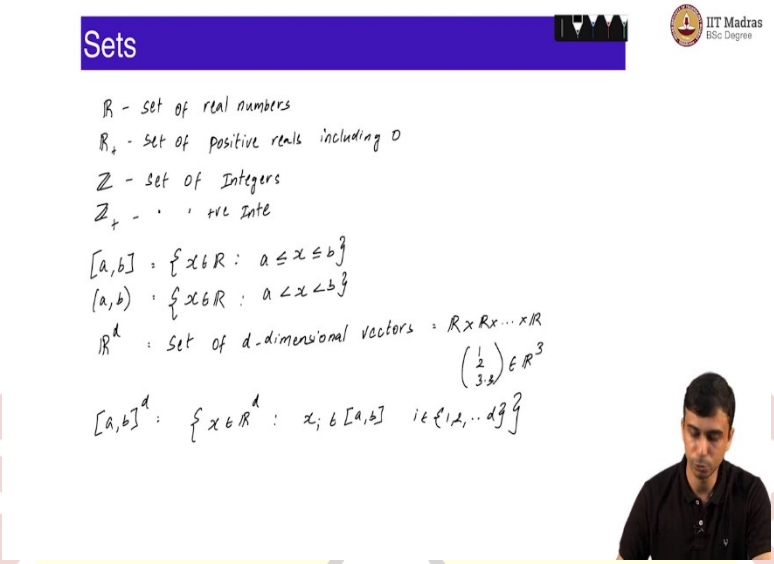
Hello everyone, welcome to the second week of the machine learning foundations course. In this week, we will look at the basic math tools that will be required for handling the rest of the course, I have titled it calculus because that is the majority of this course. And this will majorly overlap with the later part of the mathematics 2 course that you would have already done. But still, this has a different focus into some more geometric or a graphical focus rather than the course that you would have seen in math 2 for data science.

(Refer Slide Time: 0:47)



So here is a brief overview of this week's content. We will first take a look at some basic sets and functions that we will be using throughout the course and notations and the notations for them. And some basic logic statements that we will be making, and their math notation for that. And we will wrap the first part of this week with how to visualize functions and there are several ways of visualizing these functions, how to graphically visualize these functions, that will be the final part of the first lecture today.

(Refer Slide Time: 1:24)



Here are some basic sets that you will be using over and over throughout the course. The first set is R, which is simply the set of set of real numbers, which you would have seen already. And then you have R +, which is simply the set of real numbers, including 0 positive real numbers, including 0. Then we have Z, which is simply the set of integers. And Z + is a set of positive integers, including 0.

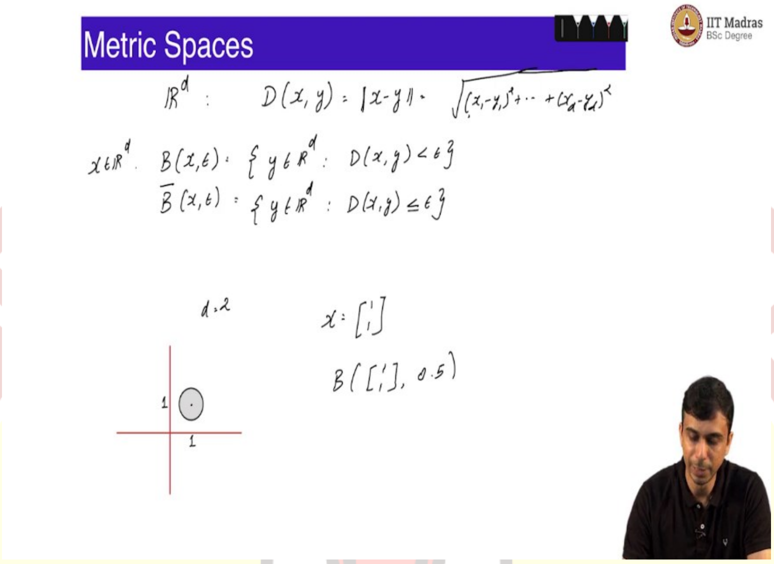
So, these are the four domain sets that will very often be using the course. And then there are a couple of other sets that we will use often, which is the closed interval a, b, this is the set of real numbers real numbers, such that they are between a and b, and the boundaries are inclusive, we include both a and be, so that is the closed interval. Then we have the open interval a , b, which is simply the set of all x in real a < x < b.

The only difference between these 2 sets is that in the second set, the boundary points a and b are not included. So, all of these sets can be exponentiated to power d, the Cartesian product d × will be taken for example, we can have Rd is simply the set of D dimensional vectors.

So, Rd is the same as writing R cross R cross d × an example element in for example, R3 would be 1 , 2 , 3.3 is an element in R3. You can have this power Cartesian product power to pretty much any set that you can think of for example, you can have a , b power d is simply the set of d dimensional vectors, but each element in the d dimensional vector is going to be between a and b.

So, we can write this compactly as the set of D dimensional vectors in Rd such that xi belong belongs to the closed interval a , b for all i belongs to 1 , 2 ,. So, these are some of the sets, the basic sets that we will notate throughout the course.

(Refer Slide Time: 4:31)



The next object that we will be interested in, in analysing is a metric space. So, what is a metric space? A metric space is simply a set with an extra structure associated with it. And the means that structure is the distance function. So, you can have a distance function D, which takes in 2 objects the set that is the distance function.

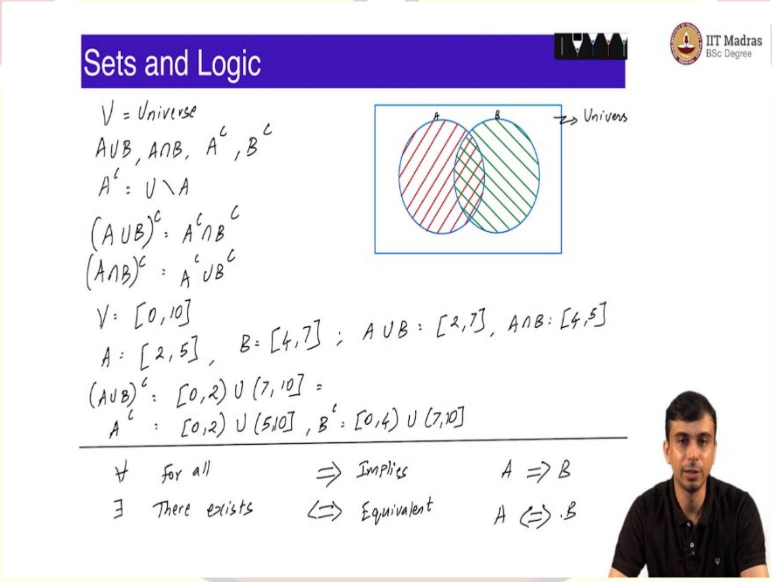
So, that the default metric space that you will be working throughout is Rd which is the D dimensional vectors with the distance function simply given by the Euclidean distance by that I mean D of X , Y is simply the norm of x - y, which is simply the square root of x1 - y1, the whole squared + … xd - yd, the whole square. So, this is the standard Euclidean norm. This will be the default metric space for most of our course.

This metric, the there is a crucial set that is associated with metric space, which is a ball, we can define an open ball of radius epsilon centered at x, that is given by b of x , epsilon is simply the set of y in Rd, such that the distance of y from x is strictly < epsilon. This is the open ball where we have some x is some element in Rd. And ball of x , epsilon is simply the ball of radius epsilon centered at x.

Some× we will also be used the closed ball we also using the closed ball, which is B bar of x , epsilon, which is the set of y in Rd such that d of x , y < or =epsilon, the only difference between the open ball and the closed ball is that the open ball does not contain the boundary, the closed ball does contain the boundary. What do I mean by that? Here is a simple example for the set.

So, imagine you are in D dimension, d =2 space, 2 dimensional space. And let us say for example, x is the vector 1 , 1. So here, this grey set is the ball of Centered at 1 , 1 and radius, let us say 0.5. This is the open ball, if you do not include the black boundary under it is actually =the close ball, if you include the boundary. If you do not include the boundary, it is the open ball, if you include the boundary, it is the closed ball. This is an example of a ball of radius or a given radius around a given center. The ball will be a very crucial part for doing a lot of analysis with metric spaces.

(Refer Slide Time: 7:28)



So now, we move to another crucial aspect of the math that we will do here, which is some sets and logic. So, you might have seen Venn diagrams, in your school itself, but we will still do a quick recap of that. So, you have sets living in a universe. So, that is that universe, let us denote it by the set V, V is the universe. And you have, let us say, sets A and B, which contain elements in the universe.

The basic operations that are associated with sets are union and intersection. So, you can have, we use the universe. You can have A union B, A intersection B, and A complement, and B complement. So, these are some of the basic sets associated with a given that offsets the A and B, A union B is simply the union of both the A and B.

For example, if this was a, and this was B, union B is simply the entire all elements in A and P, A intersection B is the elements that are in both A and B. A complement is simply the set of elements in the universe, but not in a and similarly B complement. So, you can in principle, write A compliment, as the universe set difference A. So, this is another notation that you might want to know this is the notion for set difference.

So, how do you denote the difference between 2 sets? This is the backslash symbol. That is A complement. Similarly, you can have B complement also, with the Venn diagram, you might have seen the De Morgan's laws, which are the basic laws of logic, essentially, what does it say? Well, it is a collection of 2 equations, which is A union B, the whole complement is simply =A compliment, intersection B complement.

And similarly, A intersection B, the whole complement is =A complement, union B complement. Graphically it is pretty easy to understand A union B the whole complement is simply the elements that are in the universe, but not neither A nor in B. So, that is essentially the unshaded region of this diagram, forms A union B the whole complement, that is clearly exactly =A complement intersection B complement.

So, that is the first law and similarly for the second law, A intersection B the whole complement is simply the set except for this intersection. That is A intersection B, the whole complement, and that is clearly exactly =A complement, union B complement, we will see this with a very simple example.

Let us say your universe V is the closed interval 0 , 10. This is the universe here. Let us say we have 2 sets A, A is =the closed interval, let us say 2 , 5 and let us say B is the closed interval given by 4 , 7. So, we have VA s, the closed interval 2 , 5, and B is the closed interval 4 , 7. Now let us just quickly verify the De Morgan's here.

What is A union B with this, we can say what is A union B union B is 2 , the interval 2 , 5 union the interval 4 , 7, which is clearly the interval 2 , 7. What is the A intersection B? It is clearly the interval, closed interval 4 , 5. So, what is A union B, the whole complement A union B the whole complement is simply the set 0 , the closed interval 0 10, difference, the set 2 7, and that is simply given by the union of the set 0 , 2, union 7 , 10.

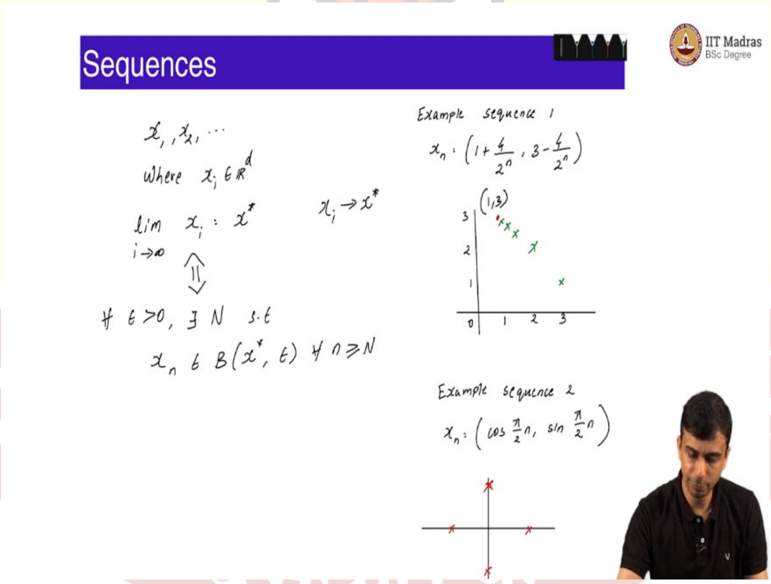
Note that I am using closed square brackets for 0 and 10. But parenthesis for 2 and 7 because 2 and 7 are not included in A union B, the whole complement. So, you can place A union B the whole complement is 0 , 2 union 7 , 10, with parenthesis for 2 and 7. This is clearly =A compliment, intersection B compliment, why is that A compliment is simply the we will just verify that quickly.

What is A compliment, well A compliment is the set 0 , 2 union 5 , 10 and B complement is simply the set 0 , 4 union 7 , the intersection of A complement and B complement is exactly =A union B the whole complement. So, this is a pretty simple way to simple verification, you can do a similar verification for A intersection B the whole complement also.

So, this is a DeMorgan's law and the basic set operations that you will need to be familiar and comfortable with. So, there is some logic statements that we will also be using throughout the course. And the four main types of logic modifiers that you will see are these first is you have inverted A, which is said to mean for all and you have an inverted E, which means there exists and then you have the arrow marks.

This, the double-sided arrow mark essentially means implies. And you might see a two-sided arrow mark which implies which saying equivalent to, so you can say A arrow B, A and B are two statements, you can say A arrow B is =saying that A implies B, you can have two sided arrows, which means that A is =B, A implies B and B implies A, then we have a two-sided arrow.

(Refer Slide Time: 14:02)



So, we will illustrate all these logic statements through another building block of analysis and calculus, which is sequences. What is a sequence a sequence is simply an ordered collection of elements. That is a sequence. Here is an example sequence, here is a collection of elements in each of these elements are going to lie in R2 xn is 1 + 4 by 2 power n , 3 - 4 by 2 power n.

First, you can see that each of these elements are two dimensional. So, these elements are in R2. What for example, let us say set n =1 is that n =1 x1 is simply 1 + 2, , 3 - 2, which is 3 , 1 which is this element here. You can substitute n =2, you would get 1 + 2 ,1 + 1, 3 - 1, which is 2 , 2. And you can keep on increasing n and let us say for large enough n, we can you have 1 + 4 by 2 power n is approximately going to be =1, and 3 - 4 by 2 power n is approximately going to be =3.

So, if you this is the first element is x1, the second element is x2, and you can keep on doing this as your n increases, you can see that the xn seems to approach this point, the right point here, which is simply 1 , 3. This is an example of a sequence. And another example of a sequence would be, let us say, this sequence where xn is cosine of pi by 2n , sine of pi by 2n.

This sequence also each element here lies in R2 and this is similar to the previous sequence, where which also like each element like. But the sequence you can see is going to be very different. Let us just try to do that. So, what if what is x1? x1 is simply cosine of pi by 2 , sine of pi by 2, which is 0 , 1. What about x2? x2 is cosine pi , sine pi, which is this, what about x3 is going to be this, x4 is going to be this. And x5 is again going to be exactly =x1 because of the periodicity of cosine and sine.

Now, this is also a valid sequence. But you can see that this sequence does not seem to converge towards any particular point. So, I will make the notion of convergence clear soon. But you can see the difference between this sequence and this sequence. So, the first sequence as the N essentially increases, you effectively, you have your xn be a constant, but in the second sequence, as n increases, it does not seem to stay still it keeps on moving.

So, this gives us the first important definition of a sequence, which is convergence, or limit can say that, let us say you have a sequence x1 x2 dot …, where xi are d dimensional vectors, we can say that limit i approaches infinity, xi is =x star if there exists an epsilon > 0, if for all epsilon > 0, there exists an integer n such that your x, small n belongs to the ball centered at x star , epsilon for all n > or =capital.

Those the equivalent definition here so we can have a 2-sided arrow. When someone says the limit of the sequence is x star, what they mean technically is the statement, we will try to parse the statement quickly. What does this say this says that for whatever small epsilon that you choose, there exists some number integer n, such that, after that integer, you are xn is going to stay within a ball of radius epsilon centered at x star.

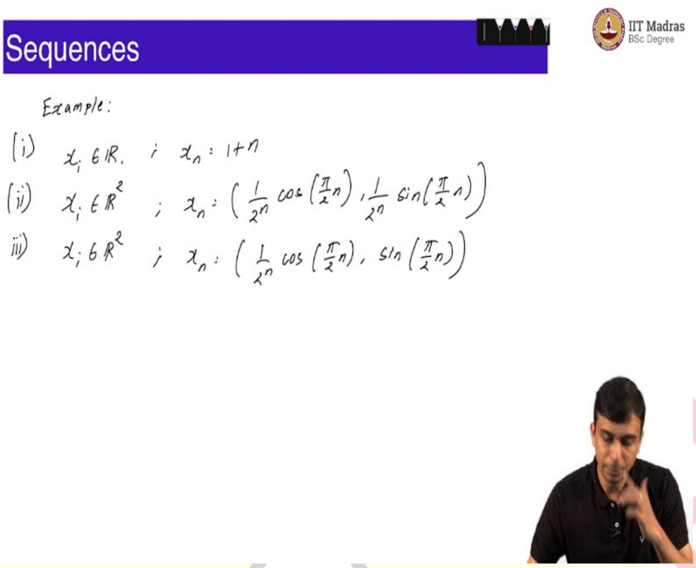
Let us try to check that for the first sequence that we had, whatever epsilon equals, which means that how small a ball that you can choose around x star, you can see that after some large enough n, your xn's are going to stay inside that ball, you can choose however small radius that you want, you can choose a radius of 0.1, 0.01 0.001 can use whatever radius around x star which is in this case, 1 , 3, you can consider a circle or a ball of radius epsilon around 1 , 3.

And eventually, your sequence is going to go inside that ball and stay inside and it is not going to go out the polygon for, this is going to be true for whatever radius that you choose as long as it > 0. So, that this means that this particular sequence here converges to 1 , 3, or in other words, limit i approaching infinity x i is =1 , 3. So, that is the, this is an example of a convergent sequence.

You cannot say the same for the second sequence. For example, I can draw a ball of radius let us say 0.1 around any point that we can think of, let us say I draw a ball of radius 0.1 around this point. Yes, occasionally you do go your sequence does go inside this ball, but it also goes out it never stays completely inside the ball after a point. So, this particular sequence the second sequence does not converge to any point at all. So, this is an example of a non-convergent sequence.

The idea of a limit, and this also explains the logic. So, this is for all epsilon, which means that however small radius that you choose, there exists a point beyond which your sequence stays inside that ball of radius epsilon centered at x star. This is equivalent to saying that the limit i approaching infinity, xi is =x star or you might also write this as simply xi approaches x star, this is another alternate notation for writing this where your xi tends to star x star or xi approaches x star limit i is approaching infinity, x is =x dot all these are equivalent statements for the same mathematical fact.

(Refer Slide Time: 20:53)

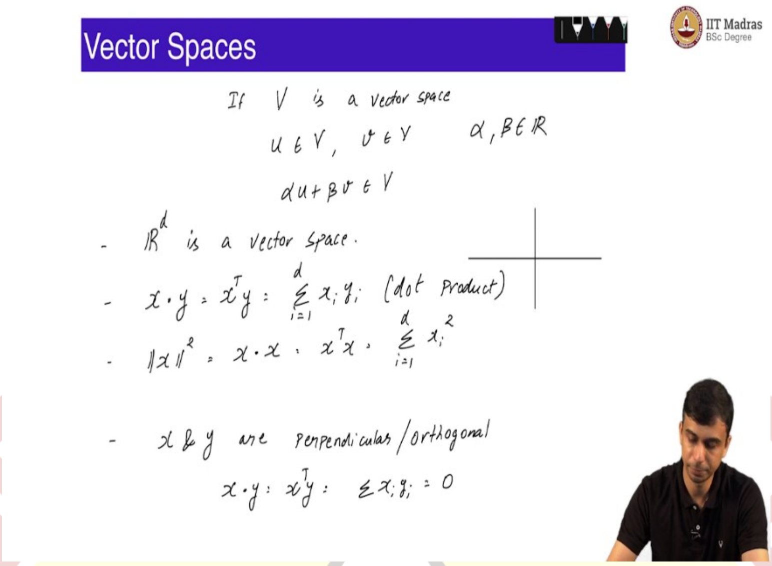


So, this here are a couple of exercise problems for you to think about. Here is one. So, for each of these examples, sequences, think, what could converge if, does the sequence converge or not? If it does converge, what is the convergent point? So, let us say we are first sequences, xi is just real value, not even R2 just real value. What is xn? xn is simply =let us say 1 + n. So, does this sequence have a convergent point? Does it converge to anything? That is the first question.

The second question is, let us say your xi are 2 dimensional vectors xi in R2. And let us say your xn is given by 1 by 2 power n, cosine of pi by 2 n , 1 by 2 power n, sine of pi by 2. The sequence looks similar to the non-convergent sequence we saw earlier, but just that now you have a1 by 2 power n and on top of it. Does this sequence converge to anything?

Then finally, you have a third sequence xi R2 where your xn is simply given by 1 by 2 power n, cosine of pi by 2 n. But the second coordinate is different. It is just sine of pi by 2 n. Just think about these 3 sequences. And do these 3 sequences have a limit? Do they converge anything? If so, what? If not, can you show that this particular these 3 sequences once these sequences converge, do not converge to anything. This is a pretty simple exercise. I hope you can do this.

(Refer Slide Time: 22:51)



So, now we will move on to the next fundamental building block of this course, which is a vector space. A vector space is also similar to a metric space in the sense in the sense that there is also a set with an extra structure associated with it. Vector space, you can simply call this a collection of vectors, but not any arbitrary collection of vectors, this collection of vectors must satisfy certain properties, there are several properties but the most important property is that is as follows.

If v is a vector space, then this particular property must hold there are other properties but this is the most crucial property you have, let us say an element u belongs to v. And let us say u have another element v also belongs to v let us say α , $\beta$are some arbitrary real value scalars. If u, if v is a vector space, capital V is a vector space, then any linear combination of its elements is also in the in this in the sector, that is, you have α u + β V, also belongs to the vector space. This is the most crucial requirement of a vector space. There are other requirements also, but this is the most important one.

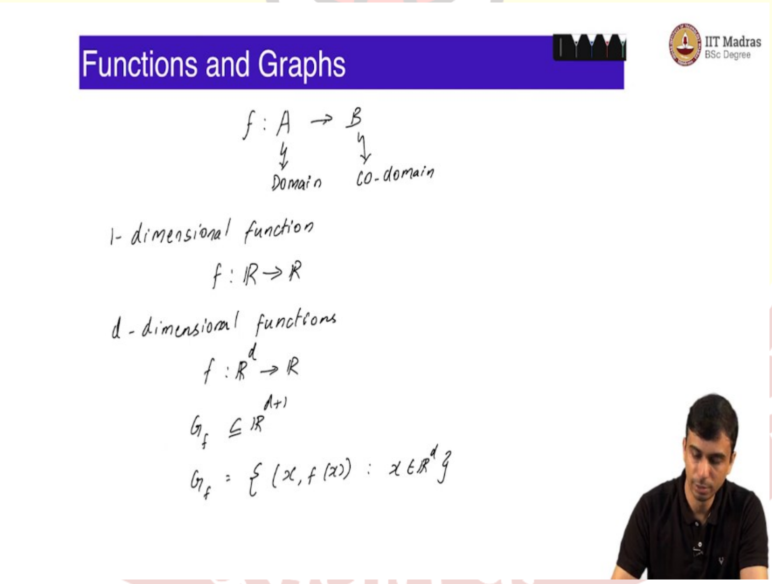
And for our purposes, the most important vector space is simply going to be Rd Rd is a vector space. Why is that just kind of obvious to see this, let us say you have 2 3 dimensional vectors. If you add them, the output is also clearly a 3-dimensional vector. Let us say you have 2 3-dimensional vectors u and v 5 ×u + 7 × V is also a 3-dimensional vector. So, Rd is clearly a vector space. According to this definition, at least it also satisfies all the other requirements of a vector space.

Some× a vector space is also associated with what is called a dot product or an inner product. You have, for example, let us say a vector, let us say x and y belongs to the vector space V, then x dot y is =x transpose y is simply sum over i =1 to d xi yi. This is the dot product of 2 vectors in the vector space and this is the dot product.

The norm of vector square norm square is simply given by x dot x or x transpose x, which is simply sum over i =1 to d x square. The vector space can also have these two properties, attributes associated with it, which is the dot product or the inner product and the norm, for our purposes we will mostly be working with Rd are subsets of it. So, it does R vector space will always have a dot product and a norm.

One notion that will be very often useful for us with vector spaces is the notion of perpendicularity or orthogonality. Let us say you have two elements x and y in a vector space, you will say x and y are perpendicular or orthogonal to each other, are orthogonal to each other, if x dot y is =x transpose y is =sigma xi yi =0. If you have the dot product of two vectors is =0, then we will call these two vectors as orthogonal or perpendicular to each other.

(Refer Slide Time: 27:04)

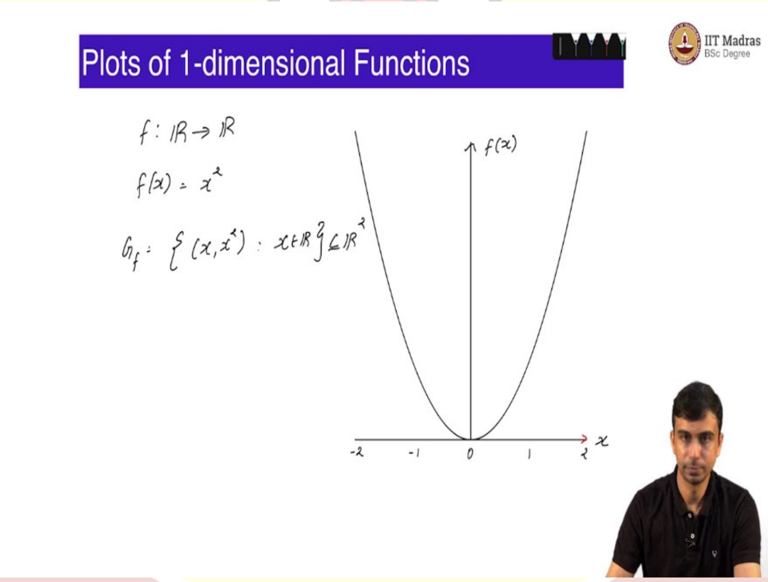


Moving on to the other important building block of our course, which is functions. Functions are simply mappings from one set to another set. So, a function f let us say is a mapping from A to B, A and B are sets is called the domain of the function f, and B is called the codomain of the function f. When we say real valued functions, we mean the coding codomain of that function is R. We will be working mostly with the real value functions.

Say one dimensional function means that you have a function from reals to reals that is, the domain is R1 and co domain is also R1. We will often also be working with D dimensional functions or multivariate functions where the function f is a mapping from Rd to R. So, that is a function a function is simply a mapping from the domain to its codomain. Sometimes for visualization purposes, it is also nice to have a graph of a function.

What is a graph of a function, a graph of a function is a subset. So, if you have a function f from Rd to R, the graph of this function, let us call it Gf, is a subset of Rd + 1. Gf is simply the set of x , f of x such that x is in Rd. The graph of a function is you vary your input over the domain, and you concatenate your input, which is an Rds, this is a D dimensional vector. This is a scalar. So, if you concatenate a d dimensional vector and a scalar, you get a d + 1 dimensional vector. So, that is why this is a subset of the graph of a function f is a subset of Rd + 1.

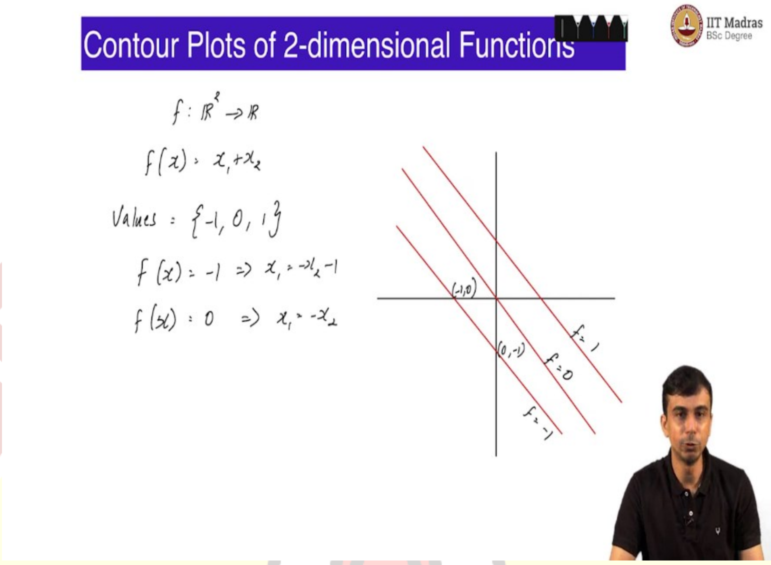
(Refer Slide Time: 29:33)



So, here are simple illustrations of visualizing simple functions first. Now, one-dimensional function, let us say one dimensional function f from R to R. The easiest way to visualize that is a simple plot where you just have an x axis, where you try different inputs and on the y axis, you simply plot f of x.

So, you for example, in this case, this function is simply f of x is =x squared. So, very simple function. And you can easily visualize this by a plot, it is possible to do that for one dimensional functions. And what is the graph of this function graph of this function is a subset of R2 now, because f is a function from R to R. The graph of this function is a subset of R2, and Gf is simply =x , x squared such that x belongs to a subset of R2. But really, it is just a, you can imagine, the graph on a paper corresponds to that all you have ink on the paper for the scope corresponds to a subset of R2 and that is the graph of this function.

(Refer Slide Time: 31:01)



Things become a little bit harder with two dimensional functions. That is, if you have a two-dimensional function that is f is from R2 to R, it is no longer possible to easily plot it because you will need one more dimension. And paper is just 2-dimension papers, our screens are just 2 dimension, you cannot really plot this such a function. And 2 dimensional functions are really very useful for going beyond 1D. And they are the simplest illustrations, which are non-trivial in most situations.

And so, it is nice to have a way to visualize two dimensional functions. And the standard way to visualize a two-dimensional function is to use contour plots. So, what do we mean by that, let us say we have a simple function f from R2 to R. Let us say f of x is =x1 + x2. So, x is a two-dimensional vector, which means that there has two components x1 and x2 F of x is, say x1 + x2.

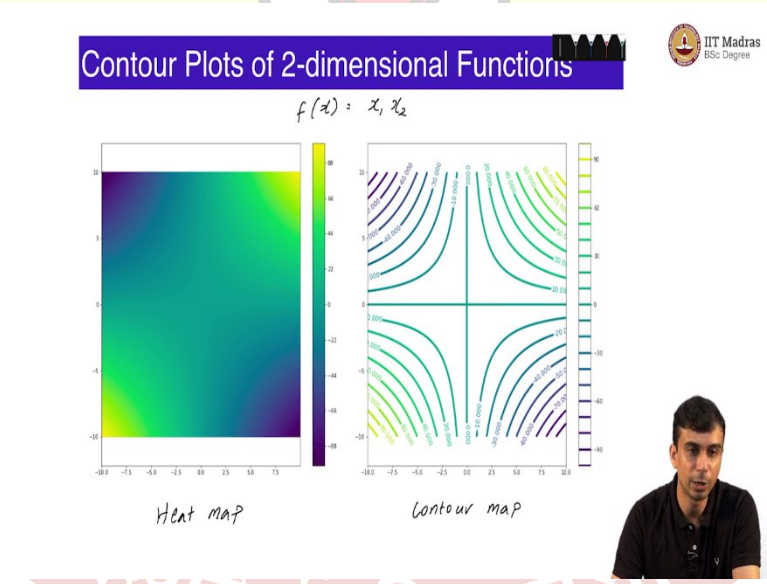
How do you visualize this function, the standard would utilize this function is to think of a set of values, the output of this function can take, for example, let us say the values that you are going to give are, let us say - 1 , 0 , 1, it can be more values also, but for now, let us just say - 1 , 0 , 1. What do we mean by that, we take the function f, and we set f of x to be =- 1, we essentially can consider all solutions of this x and plot that as a curve.

So, when you said f of x is =- 1, which means that x1 + x2 is =- 1, which x1 is =- x2 - 1. So, which means that x1, let us see, if you plot on x1 , x2 which satisfy this constraint, you would simply get this. In particular, it would pass through the points given by 0 , - 1 - 1 coma 0. Let us label this particular red line by the contour F is =- 1.

Similarly, you can do that for other, the other values in the value set, which is f of x is =0 implies, x1 is =x2, x1 is =- x2. And that would correspond to this set. We will label this also by the property and contour which is f is =0. And similarly, you can have f =1 would correspond to this not being particularly neat inclined these figures, but this should give you an idea of what we mean by contours.

So, these for example, you could have more values here, you could have - 1, 0, 1, 2, 3, 4, 5, you would have more contacts. So, each of these lines are called contours. And this might remind you of some things you would have done in your geographies in your school. For example, you might have plotted isotherms are isobars, which are lines of constant temperature or lines of constant pressure. Those are exactly contour plots. Contour plots are the standard way to visualize 2D functions.

(Refer Slide Time: 35:02)



Here is a contour plot of a slightly more complicated function where f of x is =x1 into x2. So, you can see that if x1 and x2 is set to 0, it means that x1 is =0 or x2 equals 0. That is why you have this particular contour labelled by 0, you can for example, set f of x is =10 and you would get this particular 4 contracts. So, you can have we can have 1 , 10 is a valid way of getting 10 - 1 or - 1 is also valid.

So, these both of these put together form the contour for f equals 10. Similarly, you can get f =- 10 would correspond to these two contours. This is a contour plot. And some× it is nice to

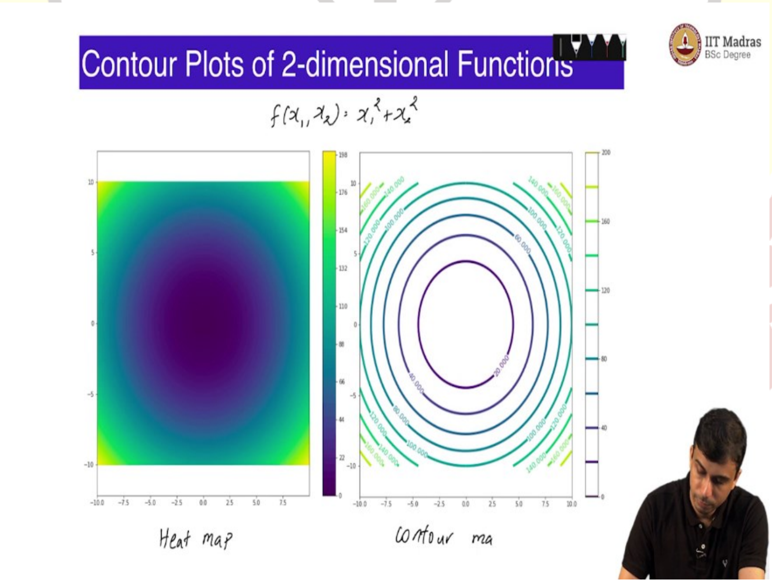
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

actually add color to a contour plot in technically speaking, you do not need color for a contour plot, you could just draw lines and label each contour by its value. But you could have colors also in this case, what I have done is you can see that yellow corresponds to high values, and purple corresponds to low values, and each contour has the same color.

Technically speaking, you could increase the number of contours to really large numbers instead of here, I have done I have drawn contours for 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and then - 10 - 20, - 30, - 40, and so on, but there is nothing stopping you from trying more controls for let us say 0, 1, 2, 3, 4, 5 or 0.1, 1.1, 1.2. If you drew enough contours, you would fill the entire region. If you do enough coloured contours, you would fill the entire region and that would correspond to this particular figure on the left. So, this figure is called a heat map. And the figure on the right is a contour map.

So, both have the heat map essentially gives you more information than the contour map, but some× it might be difficult to see the geometric shape of the contours from the heat map for example, here, it is not so easy to see that the shape of the contours are these curves, contour map gives that information, but the heat map technically contains more information about the function because it is infinite in many contours or there in a heat map.

(Refer Slide Time: 37:34)



Here is one other example of another function, f of x1 , x2 is simply x1 squared + x2 squared is either contours or circles. And the heat map also looks circular, vaguely circular in shape. This is the heat map and this is a contour. So, these are two standard ways to visualize our 2D functions. With that, I think we can wrap up the basic setup and basic notation and symbols and visualization techniques that you will need for the rest of this week. And for the rest of the course also

Machine Learning Foundations

Chapter 2: Calculus

Harish Guruprasad Ramaswamy
IIT Madras

Outline

- **Sets and Functions**
  - **Notations**
  - **Logic**
  - **Graphs and visualisations.**
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Sets

$\mathbb{R}$ - set of real numbers

$\mathbb{R}_+$ - set of positive reals including 0

$\mathbb{Z}$ - set of Integers

$\mathbb{Z}_+$ - " " +ve Inte

$[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$

$(a,b) = \{x \in \mathbb{R} : a < x < b\}$

$\mathbb{R}^d$: Set of d-dimensional vectors $= \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$

$$\begin{pmatrix} 1 \\ 2 \\ 3.3 \end{pmatrix} \in \mathbb{R}^3$$

$[a,b]^d : \{x \in \mathbb{R}^d : x_i \in [a,b] \quad i \in \{1,2,..d\}\}$

Metric Spaces

$\mathbb{R}^d$: $D(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_d - y_d)^2}$

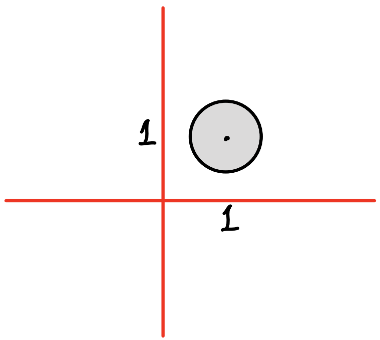
$x \in \mathbb{R}^d$. $B(x, \epsilon) = \{ y \in \mathbb{R}^d : D(x, y) < \epsilon \}$

$\overline{B}(x, \epsilon) = \{ y \in \mathbb{R}^d : D(x, y) \leq \epsilon \}$

$d = 2$

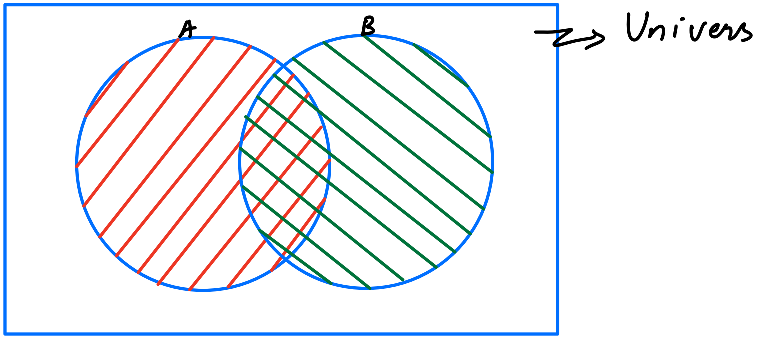
$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$B\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, 0.5 \right)$



Sets and Logic

U = Universe



$A \cup B$, $A \cap B$, A^c, B^c

$A^c = U \setminus A$

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$

$U = [0, 10]$

$A = [2, 5]$, $B = [4, 7]$; $A \cup B = [2, 7]$, $A \cap B = [4, 5]$

$(A \cup B)^c = [0, 2) \cup (7, 10] = A^c \cap B^c$

as $A^c = [0, 2) \cup (5, 10]$, $B^c = [0, 4) \cup (7, 10]$

$\forall$	For all	$\Rightarrow$	Implies	$A \Rightarrow B$
$\exists$	There exists	$\Leftrightarrow$	Equivalent	$A \Leftrightarrow B$

Sequences

$x_1, x_2, \ldots$

where $x_i \in \mathbb{R}^d$

$\lim_{i \to \infty} x_i = x^*$ $\qquad$ $x_i \to x^*$

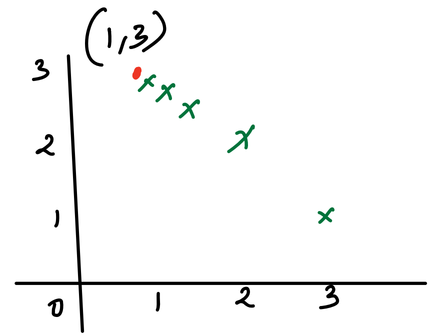
$\Updownarrow$

$\forall \epsilon > 0, \exists N$ s.t

$x_n \in B(x^*, \epsilon) \ \forall n \geq N$

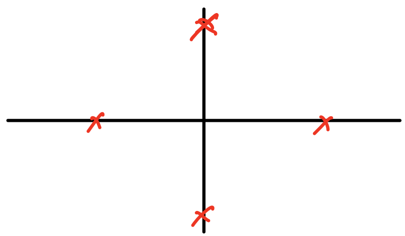
Example sequence 1

$x_n = \left(1 + \frac{4}{2^n}, 3 - \frac{4}{2^n}\right)$



Example sequence 2

$x_n = \left(\cos \frac{\pi}{2} n, \sin \frac{\pi}{2} n\right)$



Sequences

Example:

(i) $x_i \in \mathbb{R}$. ; $x_n = 1+n$

(ii) $x_i \in \mathbb{R}^2$; $x_n = \left(\frac{1}{2^n} \cos\left(\frac{\pi}{2} n\right), \frac{1}{2^n} \sin\left(\frac{\pi}{2} n\right) \right)$

(iii) $x_i \in \mathbb{R}^2$; $x_n = \left(\frac{1}{2^n} \cos\left(\frac{\pi}{2} n\right), \sin\left(\frac{\pi}{2} n\right) \right)$

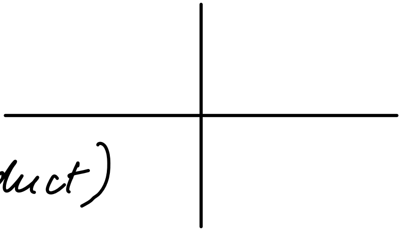
Vector Spaces

If V is a vector space

$$u \in V, \quad v \in V \qquad \alpha, \beta \in \mathbb{R}$$

$$\alpha u + \beta v \in V$$

- $\mathbb{R}^d$ is a vector space.



- $x \cdot y = x^T y = \sum_{i=1}^{d} x_i y_i$ (dot product)
- $\|x\|^2 = x \cdot x = x^T x = \sum_{i=1}^{d} x_i^2$

- x & y are perpendicular/orthogonal

$$x \cdot y = x^T y = \sum x_i y_i = 0$$

Functions and Graphs

$$f : \underbrace{A}_{\text{Domain}} \rightarrow \underbrace{B}_{\text{co-domain}}$$

1- dimensional function

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

d - dimensional functions

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$G_f \subseteq \mathbb{R}^{d+1}$$

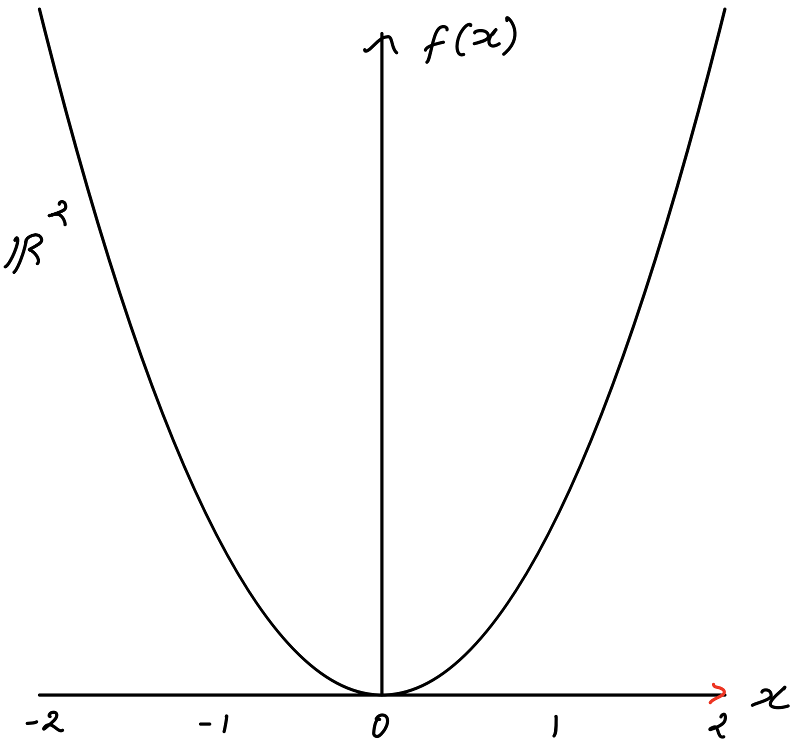
$$G_f = \{ (x, f(x)) : x \in \mathbb{R}^d \}$$

Plots of 1-dimensional Functions

$f : \mathbb{R} \to \mathbb{R}$

$f(x) = x^2$

$G_f = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R}^2$



Contour Plots of 2-dimensional Functions

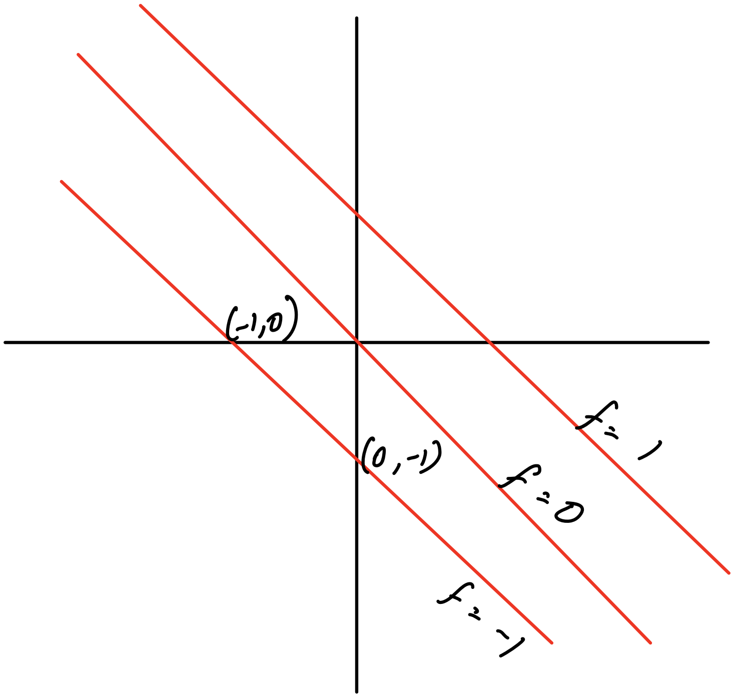
$f : \mathbb{R}^2 \to \mathbb{R}$

$f(x) = x_1 + x_2$

Values $= \{-1, 0, 1\}$

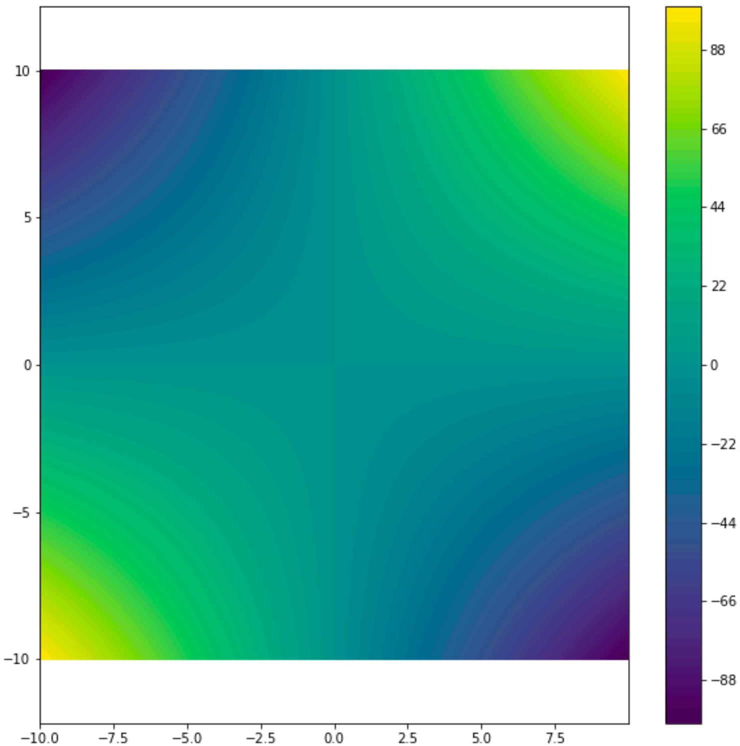
$f(x) = -1 \Rightarrow x_1 = -x_2 - 1$

$f(x) = 0 \Rightarrow x_1 = -x_2$

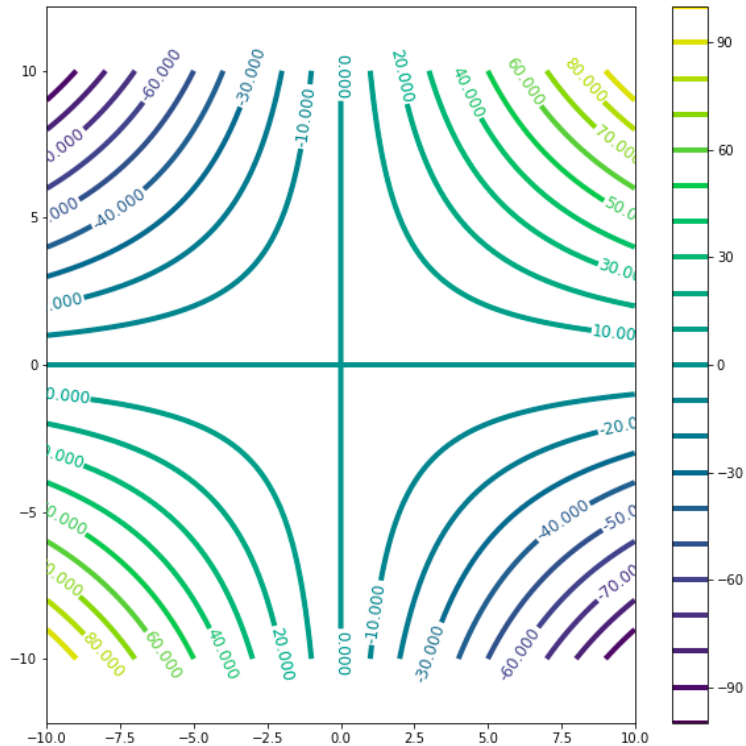


Contour Plots of 2-dimensional Functions

$f(x) = x_1 x_2$



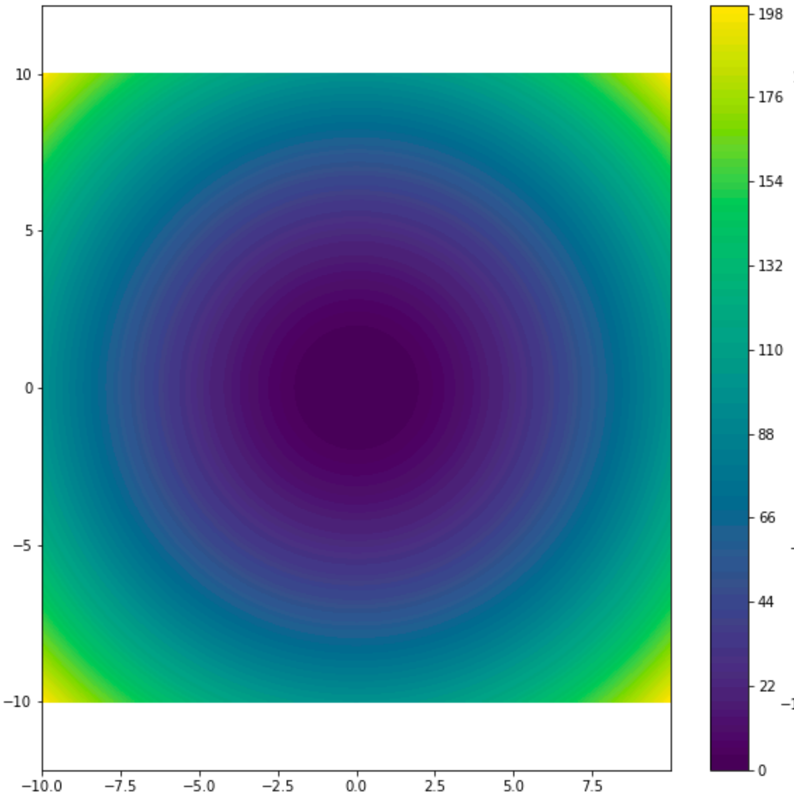
Heat map



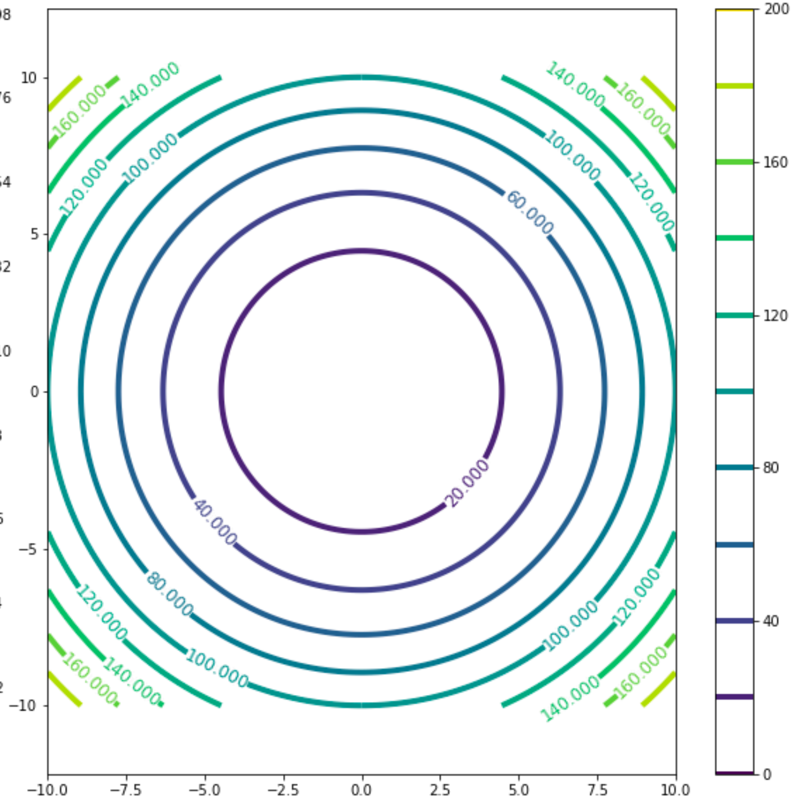
Contour map

Contour Plots of 2-dimensional Functions

$$f(x_1, x_2) = x_1^2 + x_2^2$$



Heat map



contour map.

Outline

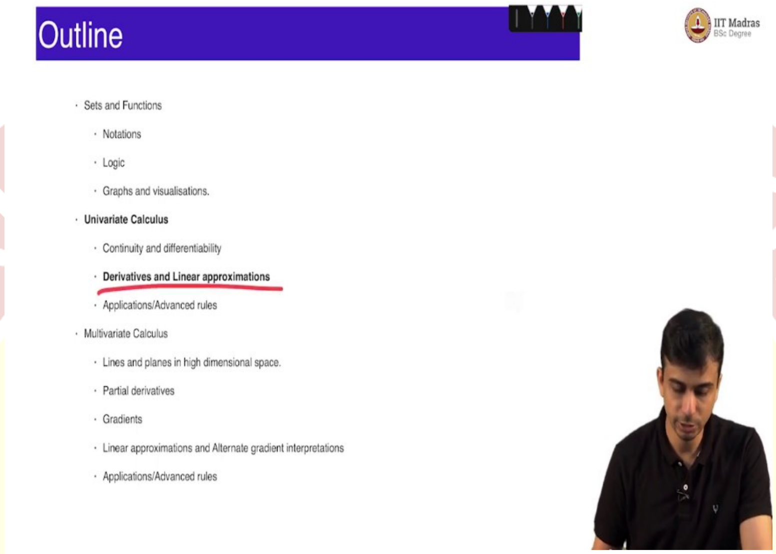
- Sets and Functions
    - Notations
    - Logic
    - Graphs and visualisations.
- **Univariate Calculus**
    - **Continuity and differentiability**
    - Derivatives and Linear approximations
    - Applications/Advanced rules
- Multivariate Calculus
    - Lines and planes in high dimensional space.
    - Partial derivatives
    - Gradients
    - Linear approximations and Alternate gradient interpretations
    - Applications/Advanced rules

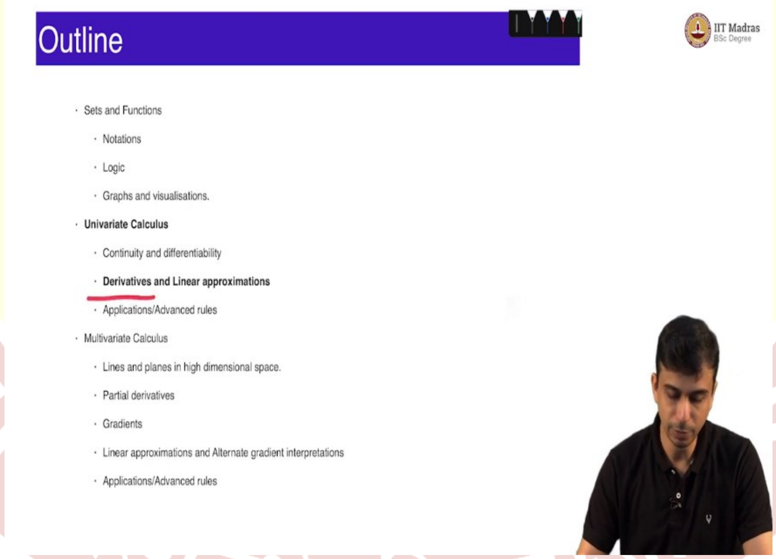
IIT Madras

ONLINE DEGREE

Machine Learning Foundation
Assistant Professor Harish Guruprasad Ramaswamy
Department of Computer Sciences and Engineering
Indian Institute of Technology, Madras
Univariate Calculus: Derivative and Linear Approximations

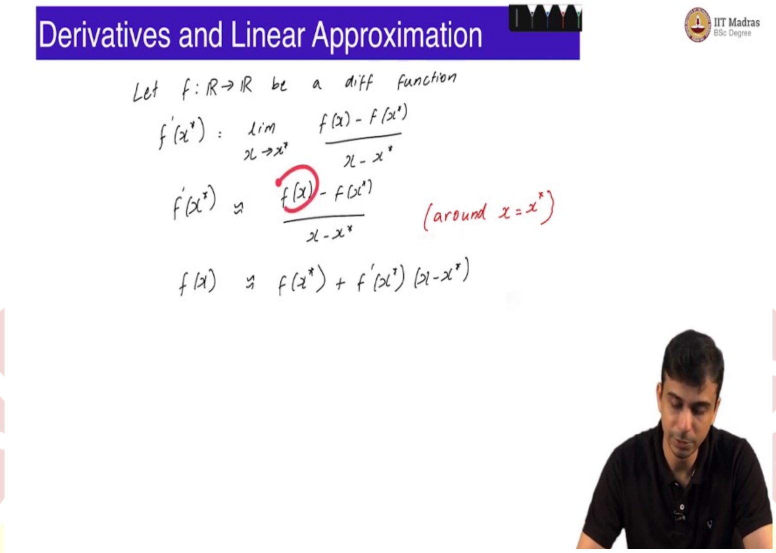
(Refer Slide Time: 00:13)

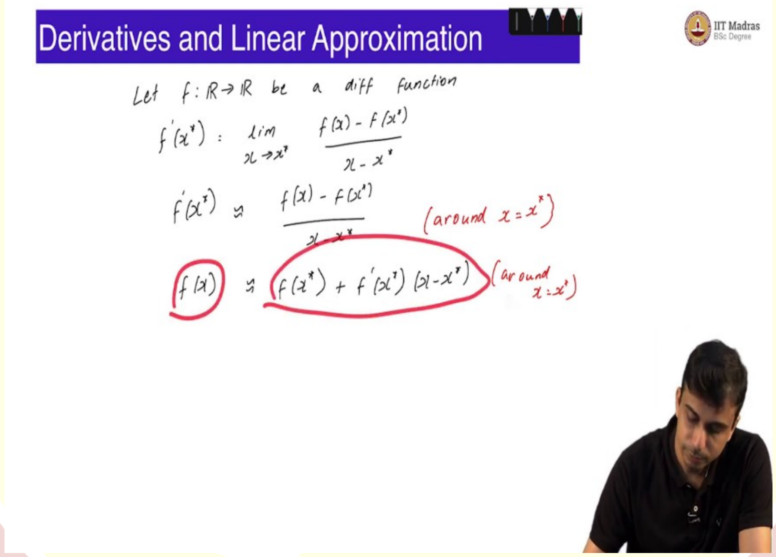




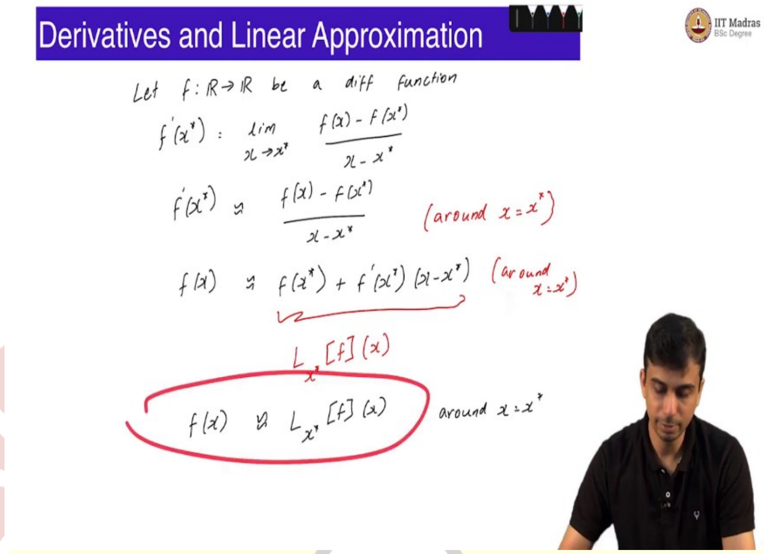
Hello, everyone, and welcome to another lecture of machine learning foundations. We will continue with the, the univariate calculus ideas that we had begun, last class. In this class, we will mainly focus on the derivative, and its link to what are called linear approximations.

(Refer Slide Time: 00:32)





सिद्धिर्भवति कर्मजा



So, let us write down, let us consider a differentiable function from R to R. So, let f from R to R be a differentiable function. What is the definition of the derivative? Well, f dash of x * = limit x approaching x * f of x - f of x * by x - x *. So, we do have a limit term here. So, we cannot do direct multiplication by x - x * and you cannot move this here.

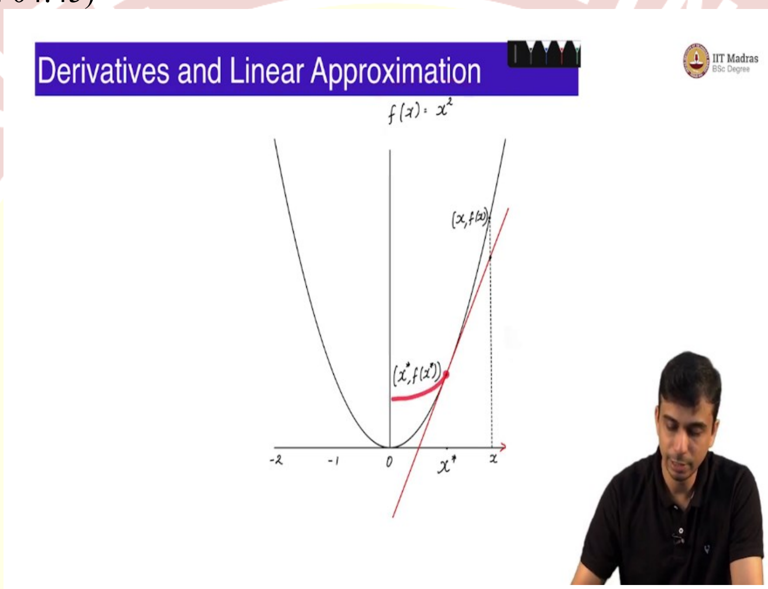
But let us assume that. So, let us assume that x is approximately= x *. That is, we are in a regime where x is approximately equal x *, in which case the limit x equal x approaching x * is satisfied. So, we can say that f of x * is approximately= f of x - f of x * by x - x *, in the neighborhood around x equals x *. So, this is true only when x is approximately= x *.

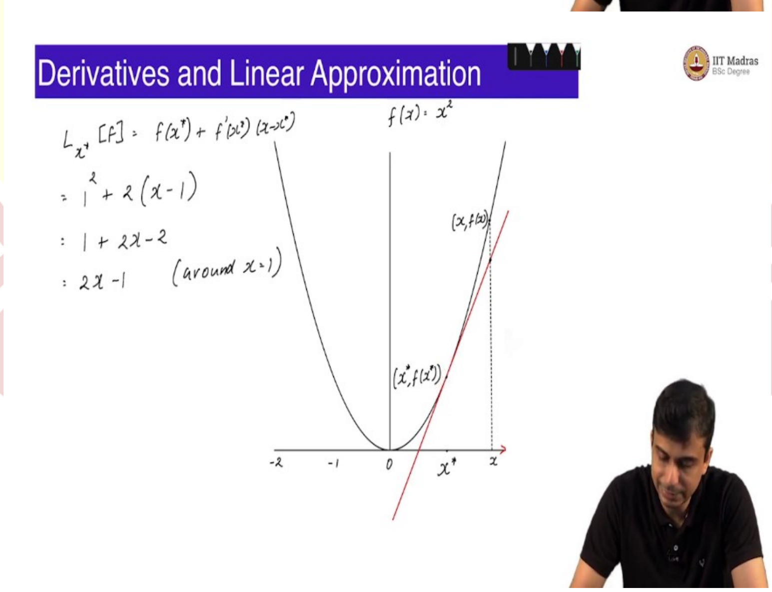
Only then, we can remove this limit statement without affecting logic. Now, that we have this, we can move things around and write an alternative expression, which is f of x is approximately= f of x * + f dash of x * into x - x *. So, f dash of x * is approximately= f of x - f of x * by x - x * around x= x *, which means that f of x is approximately= f of x * f dash of x * into x - x * + f of x *.

This is also true around x = x *. When x is close to x * this is true. And this is called a linear approximation. So, you can view this, the left-hand side is a function of x. The right-hand side also can be viewed as a function of x, you can view x * and f dash x * as constants, which means that f of x * is a constant, f dash of x * is also constant. This entire term also can be viewed as a function of x.

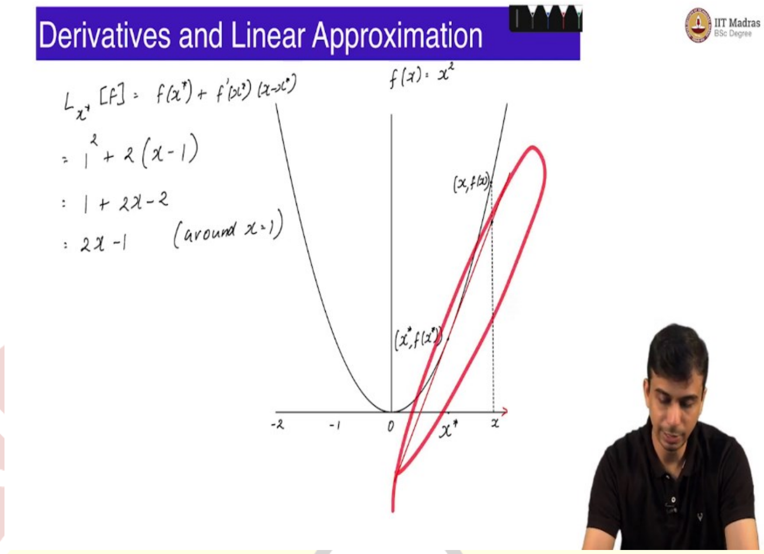
And we are going to denote that by L x * of f. This entire thing is a function of f. And if you want to evaluate it at a point x, so, it is L x * of f of x. So, what the approximately= statement says is f of x is approximately= L x * of f evaluated at x around x = x *. This is the key idea, the key expression in essentially, first order calculus. So, this is f of x can be approximated by f of x * + f dash f x * index - x * around x equals x * is the, is the key approximation idea that we will be using for a major part of machine learning.

(Refer Slide Time: 04:45)





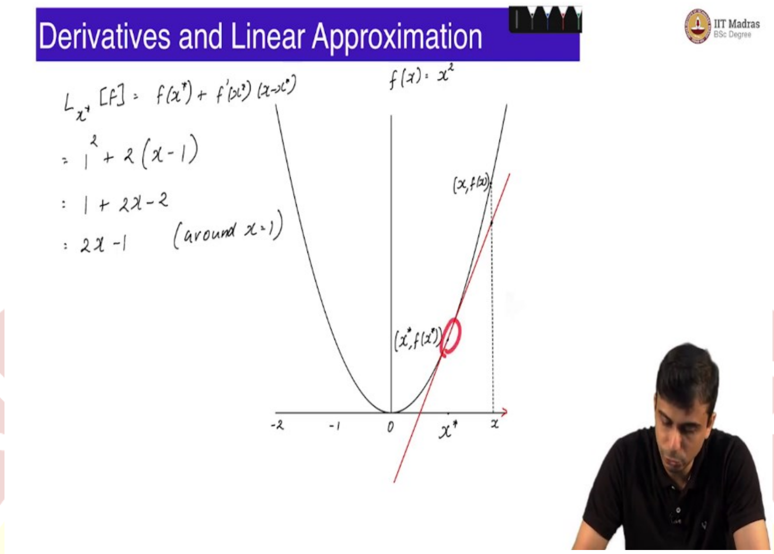
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

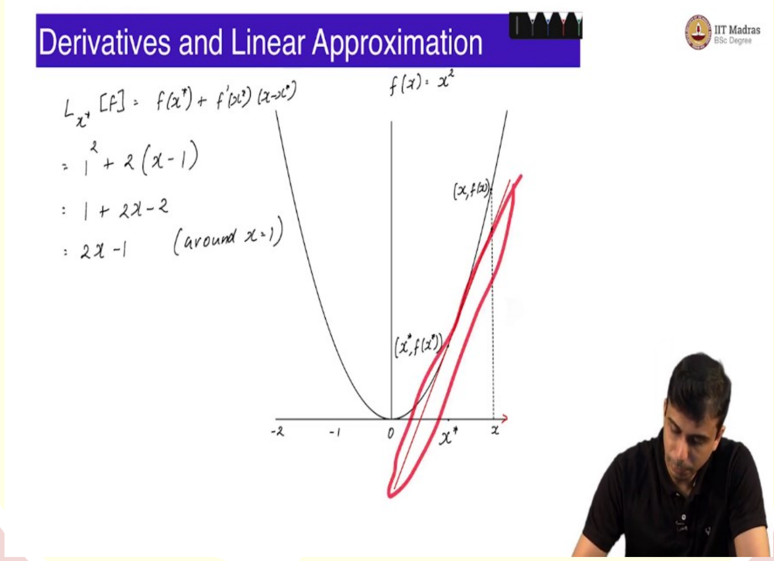


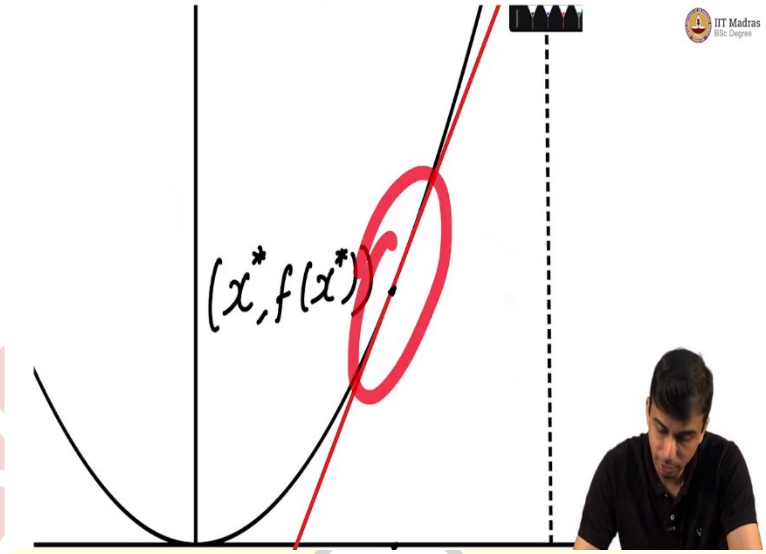
So, what does this mean for us? So, let us give some graphical illustration of this. Let us consider the simple function f of x= x squared, which we know to be differentiable. And let us pick some x * is equal r= 1. So, let us pick x * = 1, and consider to the point x * , f of x *. What is f dash of x *? F dash of x * is simply= 2, that simply follows from calculus.

So, f of x is x squared f dash of x is 2x, f dash of x * is 2. So, now we can write down what the linear approximation actually is. What is a linear approximation? Well, L x * of f is simply= f of x * + f dash of x * into x - x *, which happens to be 2 squared + 2 into x - 2, which is 4 + sorry, it is 1 squared + 2 into x - 1. We said x * = 1, so, 1 + 1 squared + 2 into x - 1, which is 1 + 2x - 2, which = 2x - 1. So, what we are saying is the linear approximation of f around x, around 1, around x= 1, this is valid around x= 1. This red line corresponds to the graph of 2x - 1.

(Refer Slide Time: 06:41)

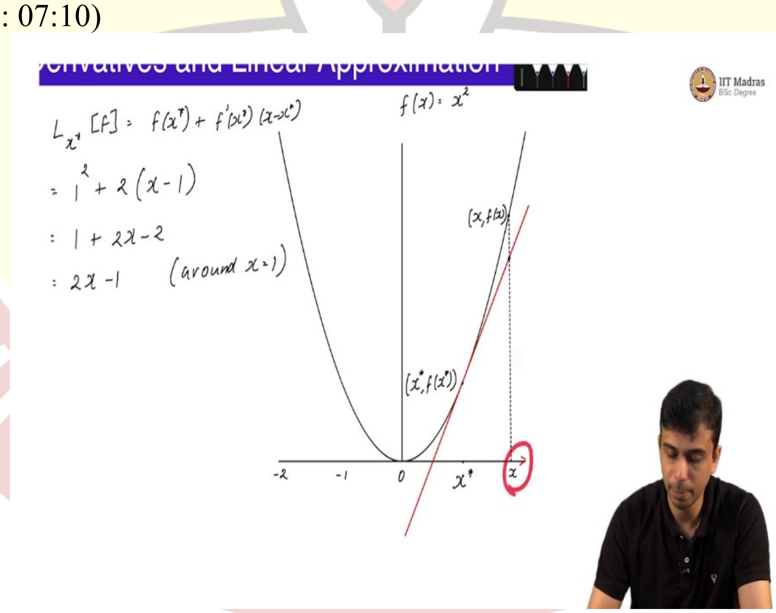


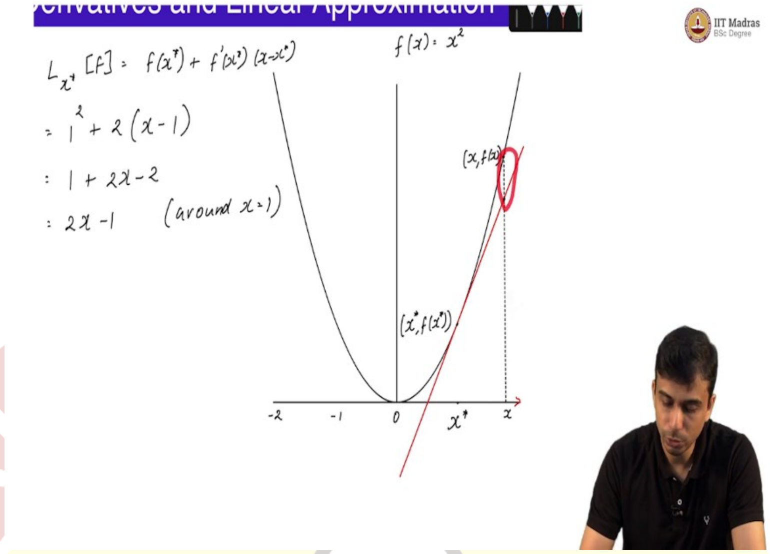


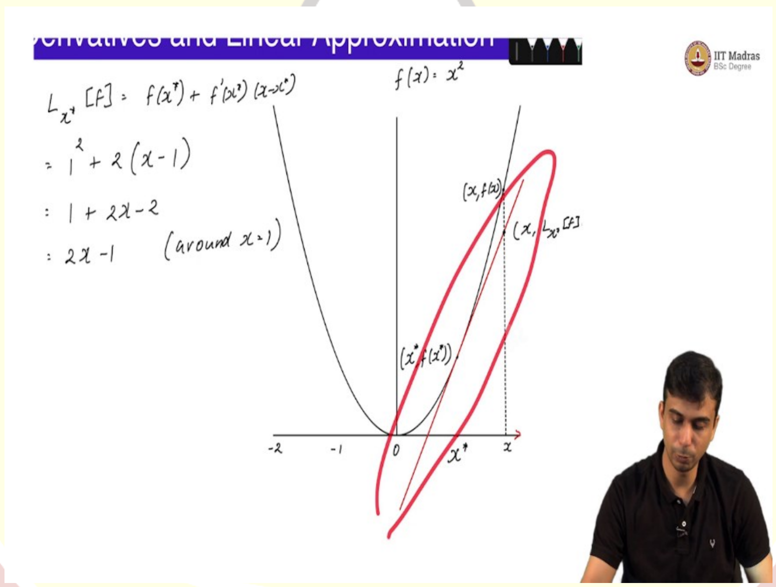


And you can see that this particular graph is an approximation to the black curve, which corresponds to x squared around x * = 1. So, you can see that x * = 1, it exactly touches it, it is exact. The farther your x moves from x *, the worse your approximation is, around x = x *, your red line and black line coincide quite well.

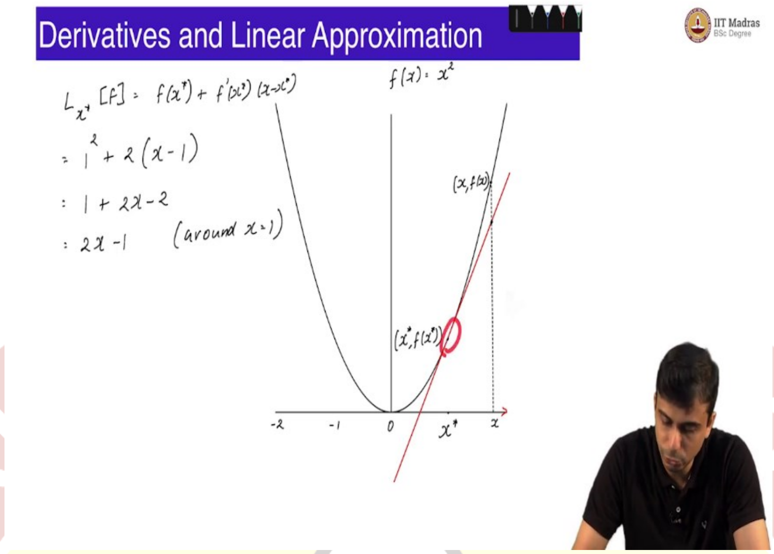
(Refer Slide Time: 07:10)





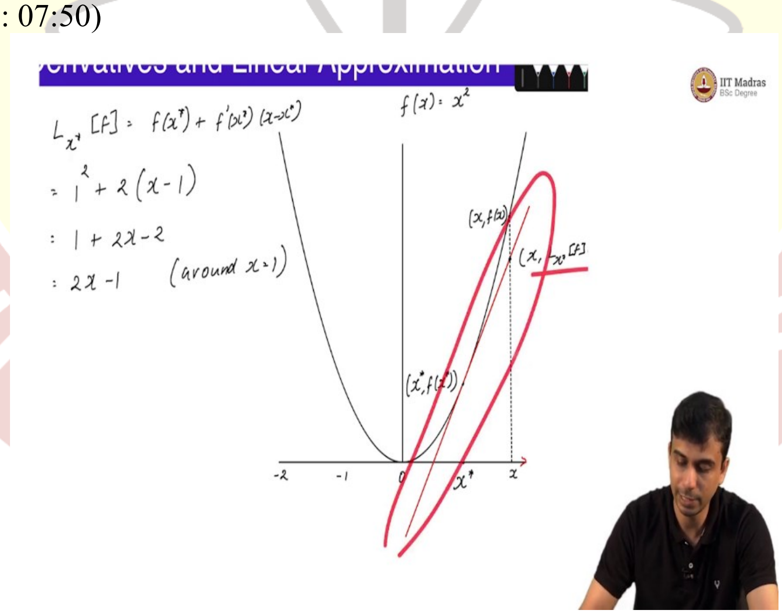


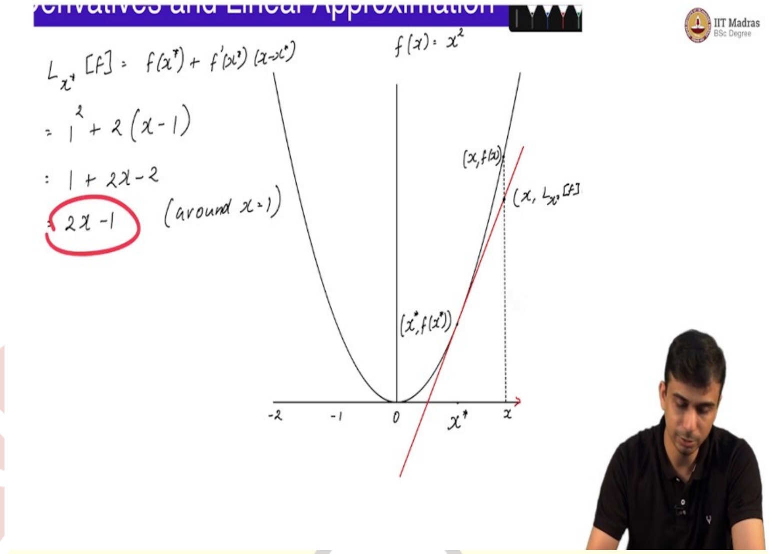
सिद्धिर्भवति कर्मजा

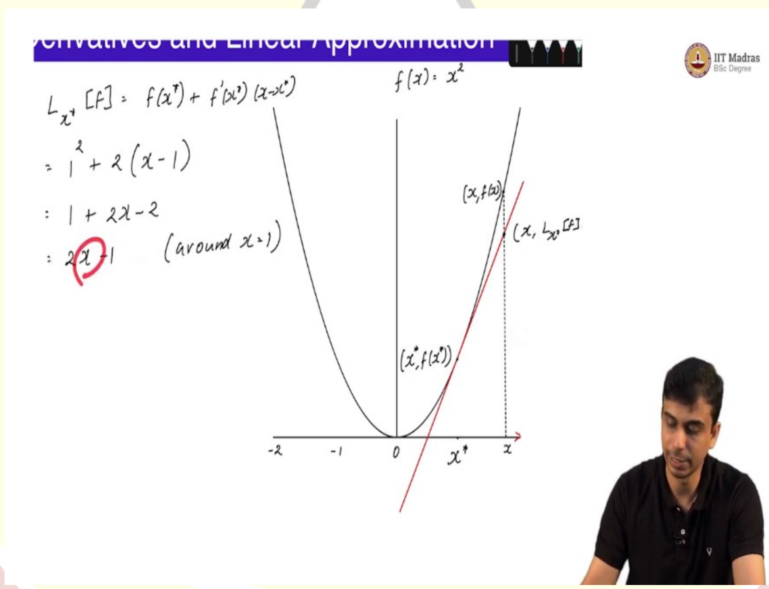


That is the reason why it is a linear, it is a approximation. Assume, as your x moves farther and farther away from x *, it becomes a (())(07:17) approximation. For example, at this point, let us say x = 2, we can see that there is actually a gap between f of x and L x * of f evaluated at x. So, this black, this point here corresponds to x , L x * of f evaluated at x. So, that is the reason why I am saying that L x * of f is an approximation of f around x= x *.

(Refer Slide Time: 07:50)

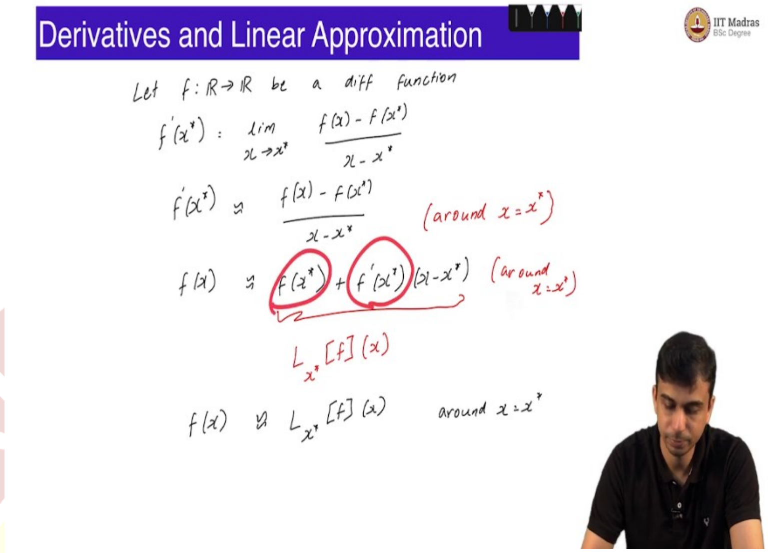


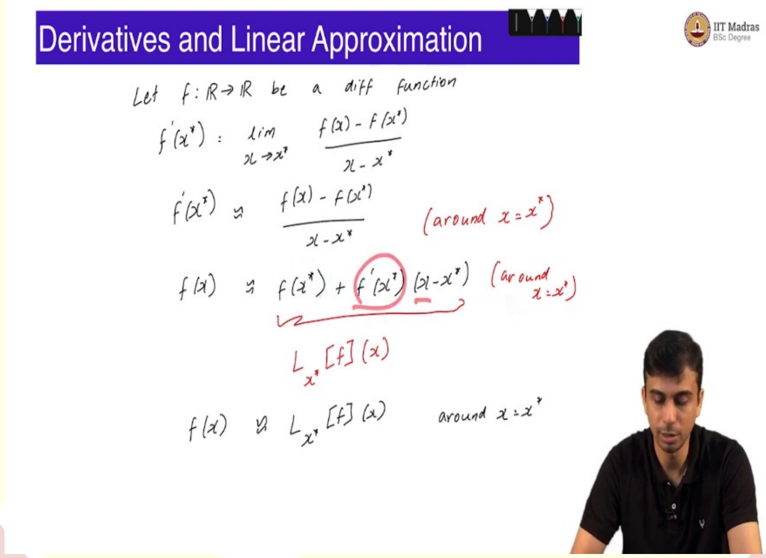


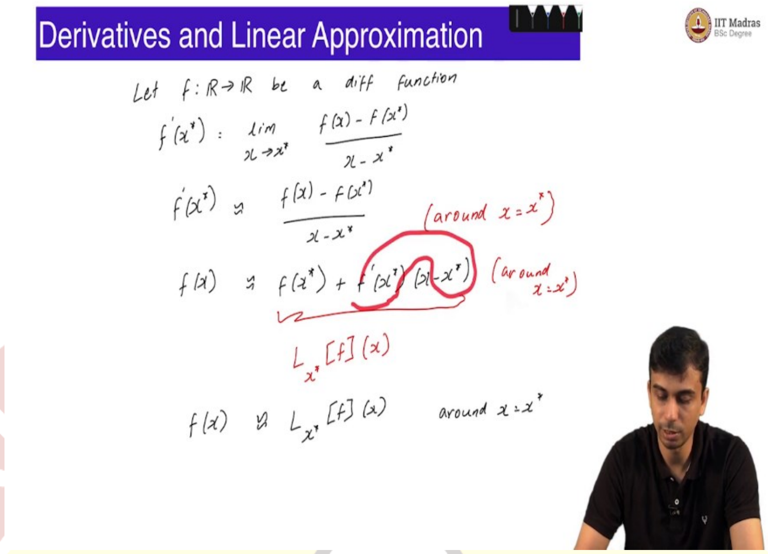


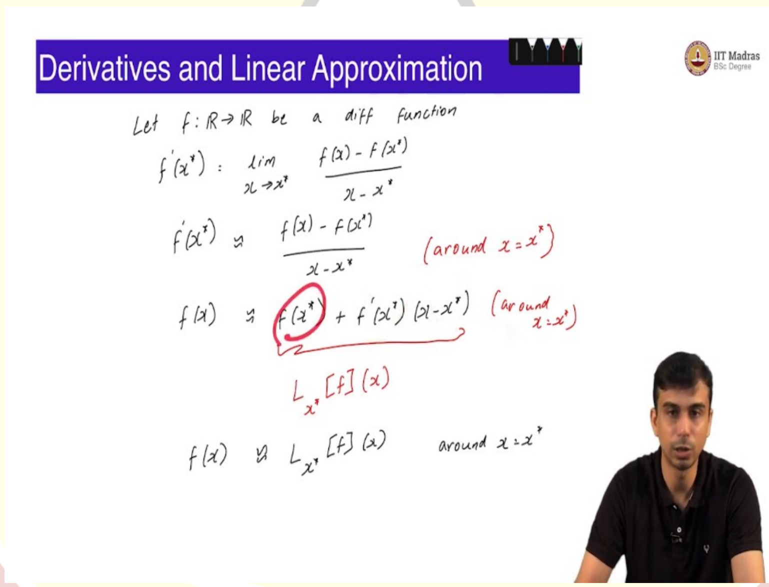
So, that is, that is what we mean by saying that L x * of f is a linear approximation. And why is it a linear approximation? You can see that L x * of f, the graph of Lx * of f is a straight line. Why is that? Because this 2x - 1 is a linear function of x, the coefficient is it is, it is a equation of a straight line.

(Refer Slide Time: 08:17)





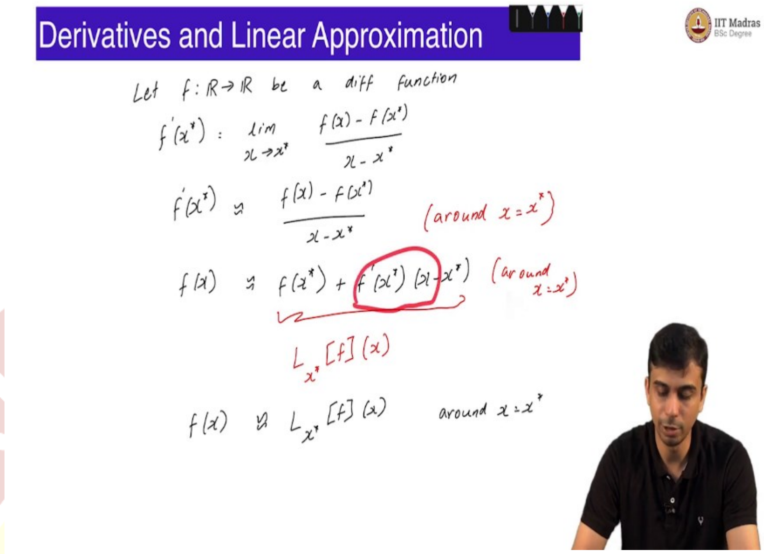


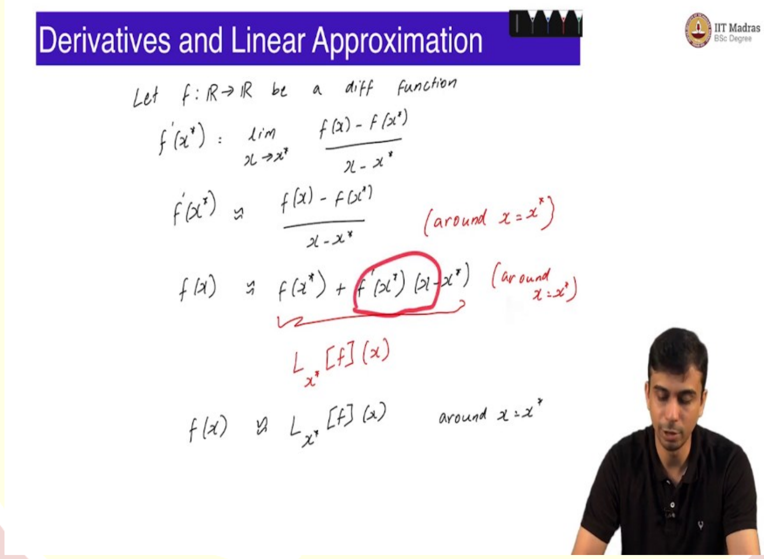


In general, also, it is true because you can see that L x * of f viewed as a function of x is simply the coefficient is f dash x * which is a constant. We have, this as a constant and constant times x, and f dash x * into - x * is also a constant.

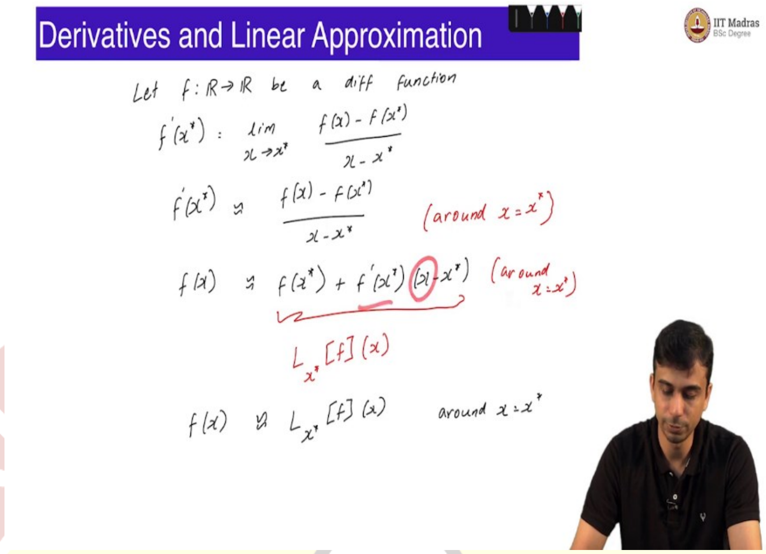
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

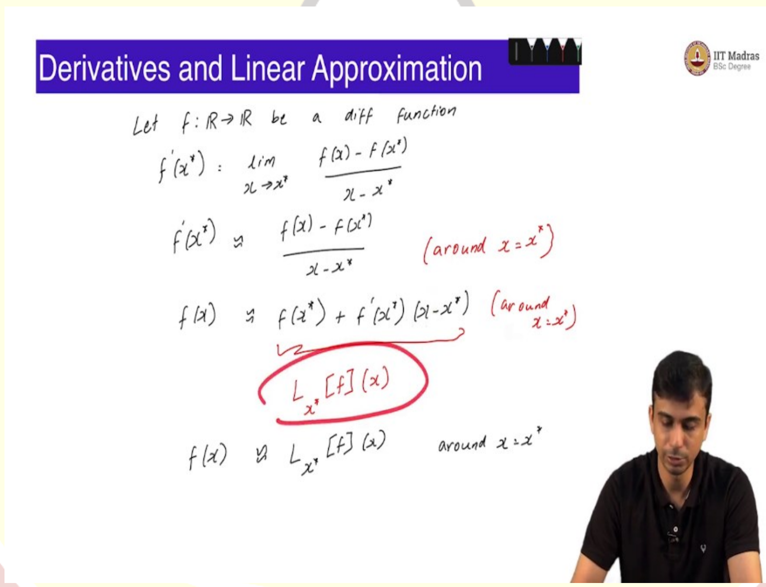
(Refer Slide Time: 08:34)





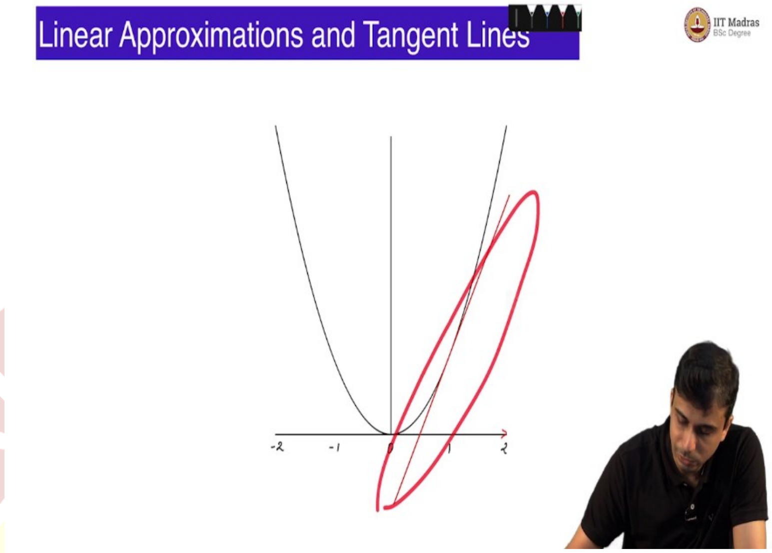
INDIAN INSTITUTE OF TECHNOLOGY MADRAS • सिद्धिर्भवति कर्मजा •

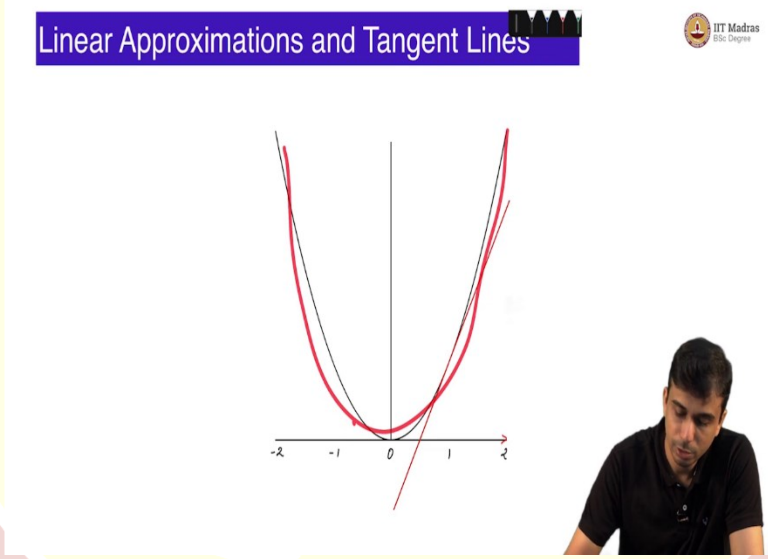


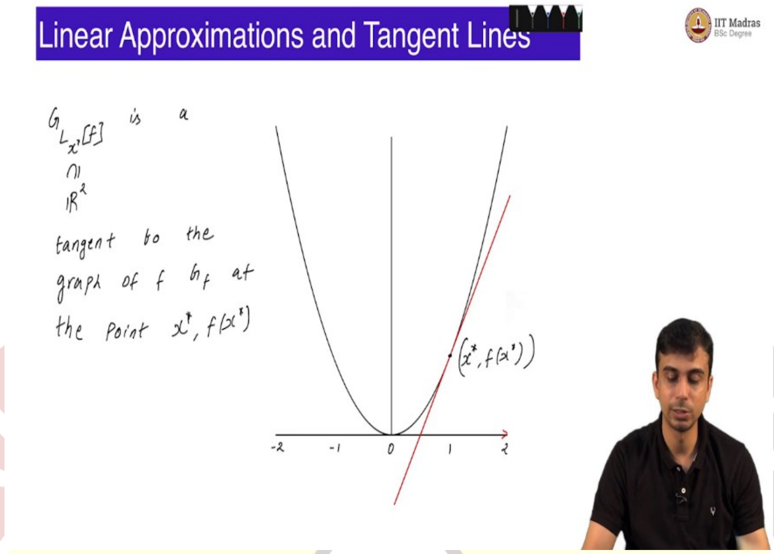


So, f of x * + f dash of x *, sorry, f of x * - f dash x * expression is a constant. And you have f dash x * into x, which is a linear coefficient of x multiplied by x. So, that is the reason why this is called a linear function. And this is an approximation of f around x= x *.

(Refer Slide Time: 08:57)





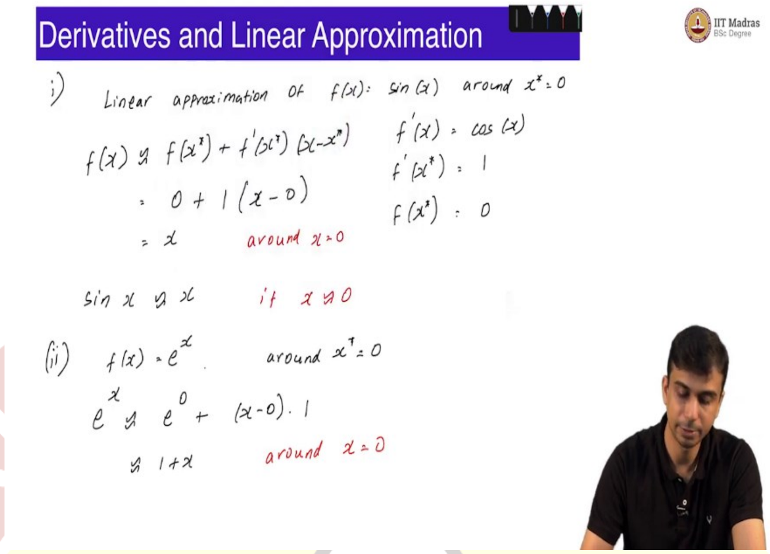


So, that is, that is about linear approximations. And you can, you would also seen another interpretation of linear approximation via what is called tangent lines. So, tangent is a, a geometric object, while a linear approximation is, a functional object. What is a tangent? Well, a tangent is, I will geometrically define vaguely dominantly define, one set is a tangent to another set if it touches the other set.

So, if to, we can say a line is a tangent to a set, if that line touches the first set. I am not going to define touching in a formal sense because that is way too much work, but it is. In this case, you can see that the red line which corresponds to the graph of the linear approximation of f around x * is a tangent to the graph of f. Or in other words, I am going to write it in a formal way, which is your graph of L x * of f, L x * of f is also a function.

So, the graph of that function is also going to be a subset of R2, in the case of functions from R to R, this is a subset of R2, is a tangent to the graph of f Gf at the point x * , f of x *. So, this point is the tangent, is the point at which these two sets touch each other. That is x * , f of x *. This is the definition of a tangent line, a linear approximation to form tangent lines to the graph, the graph of linear approximations forms tangent lines to the graph of the function.

(Refer Slide Time: 11:18)

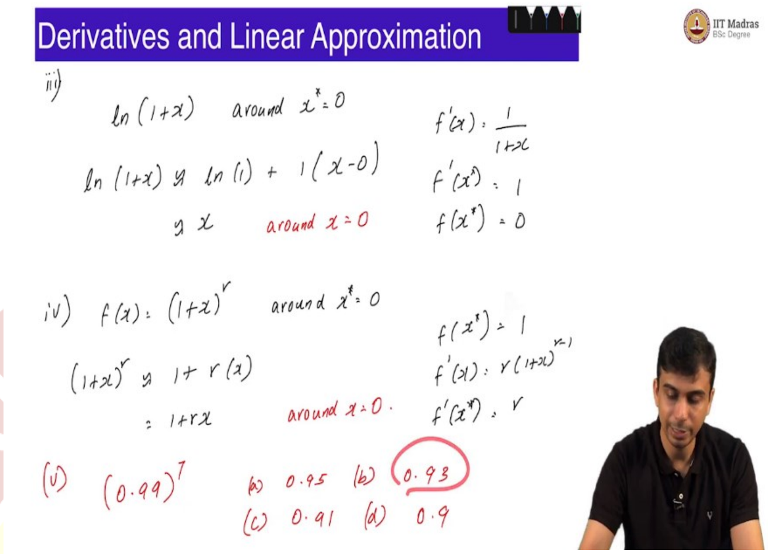


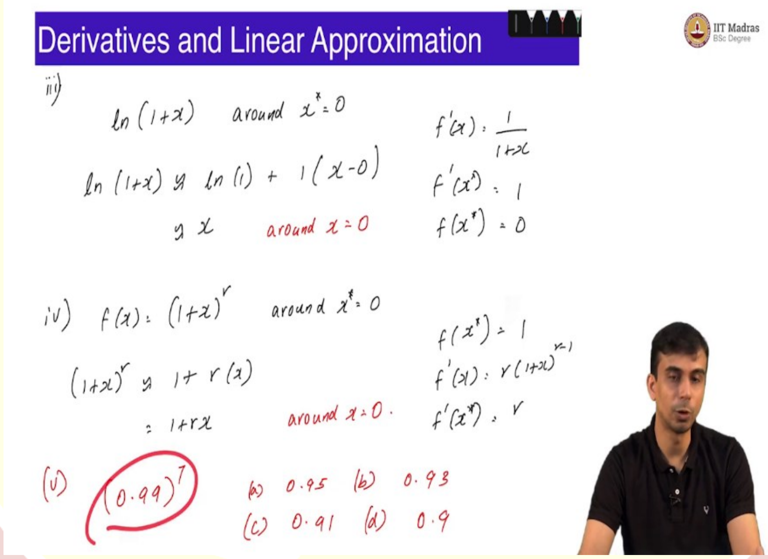
Here. now we can, we have the basics of linear approximations and tangent lines, we can see some examples. The first examples perhaps the most famous one, linear approximation, give the linear approximation of f of x = sin x around x * which = 0. So, to do that, you need f dash of x. What is f dash f of x? F dash of x is simply cos x, f dash of x * just (())(12:08) 0, which is 1. And f of x *, which is sin 0, which is 0.

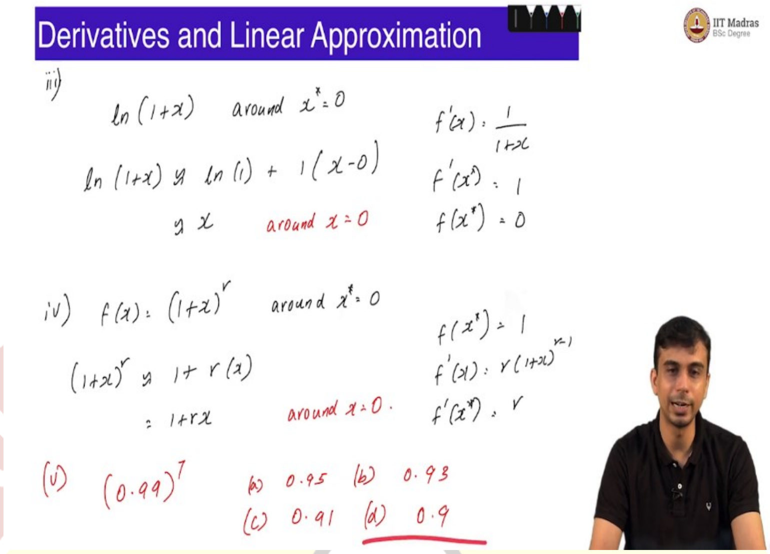
With these three facts put together, we can say f of x is approximately= f of x * + f dash of x * times x - x *. What is that? Which is 0 + 1 into x - 0, which is and when is this valid? This is valid around, around x = 0. When x is approximately, x is close to 0. This is a value approximation, or in other words, it says that sin x is approximately= x, if x is approximately= 0.

This is the famous sin theta by limit, sin theta by theta = 1 thing that you would have derived in your school. Let us do one more classic (())(13:18). Let us say f of x = e power x. I want the linear approximation of this function around x * = say, 0. In this case, what would be, what would be e power x? That is approximately= e power 0, + x - 0 into f dash of 0, which = 1. So, e power x is approximately= 1 + x around x = 0. So, I am going to highlight the fact that e power x is approximately 1 + x is true only around x= 0. And I am going to highlight that, similarly this also deserves a highlight.

(Refer Slide Time: 14:18)







So, a few more classic examples. Let us do what is log of 1 + x around x * = 0. F of x= log 1 + x, what is f dash of x? F of x= 1 by 1 + x, f dash of x * = 1. And f of x * = 0. So, we can put all that together and say log of 1 + x is approximately= log of 1, which is 0 + f dash of x *, which is 1 into x - 0. So, log of 1 + x is approximately= x around x = 0.

So, we will do one more example. Let us say f of x = 1 + x power r, r is some integer, let us say 2, or 3 or 7 it does not matter. So, 1 + x power r around, you want approximate 1 + x power r, around x = 0. So, what is f of, what is f of x *? F of x * = 1, f dash of x in general is simply= r into 1 + x, or r - 1. And f dash of x * in particular, is going to be= r.

We can put all of that together and say that 1 + x power r is approximately= 1 + r into x - 0, (())(16:21) just 1 + r x. And this is also valid around x= 0. So, this should give you a idea of all, pretty much a lot of calculus makes approximations very easy to write down. So, this is e power x is approximately 1 + x around x = 0 or log 1 + x is approximately= x around x= 0, and so on.

You can generally apply this to any function that you have, if you want an approximation around, which is valid around a particular point x *, you can just use f of x * + f dash x * into x - x *. And the advantage of the linear approximation is that it is a much more simpler object than the original function. And this is going to be the key idea that is going to drive most of optimization.

Let us apply this particular idea for a simple question. Now, here is a, an exercise question. Which of the four options now uses (())(17:35)? I want 0.99 power 7, which of the four are closest to 0.99

power 7. The four options are, a let us say 0.95, b is 0.93, c is 0.91 and d is 0.9. So, you have four options here, and the exercise is to find which of these four options is closest to 0.99 power 7.

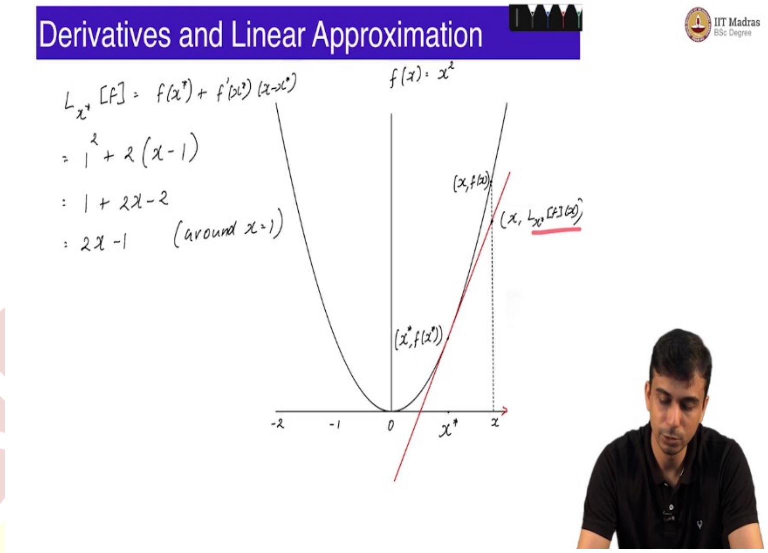
Neither of these are exact, but the question is, which of these is closest. So, 0.99 power 7 should immediately think of 1 + x power 7, where x = - 0.01. And - 0.01 is close enough to 0 that you can apply a linear approximation, which means that the 0.99 power 7 is very easily approximated by 1 + 7 into - 0.01, which is 0.93. So, 0.93 is a (())(18:42) approximation.

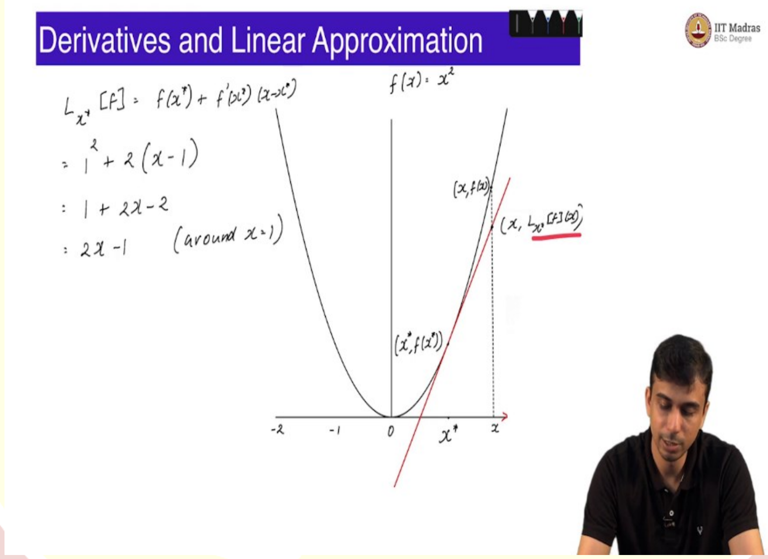
So, you can immediately get an answer of 0.93, for a complex thing like 0.99 power 7, based on the approximations. And intuitively, it makes sense. So, 0.99 power 7 essentially corresponds to a 1 percentage decay of let us say, your money in a bank every year. What (())(19:01) after seven years? How much would you have, how much loss would you have made or how much money would remain in the bank?

A first approximation would say, well, there is compound interest happening, but it is seven years. So, small enough number of years and one percentage is a small percentage that I can say that 0.99, that is, after 7 years, I would have lost seven percentage, which means that I will be left with 0.93 of my original money.

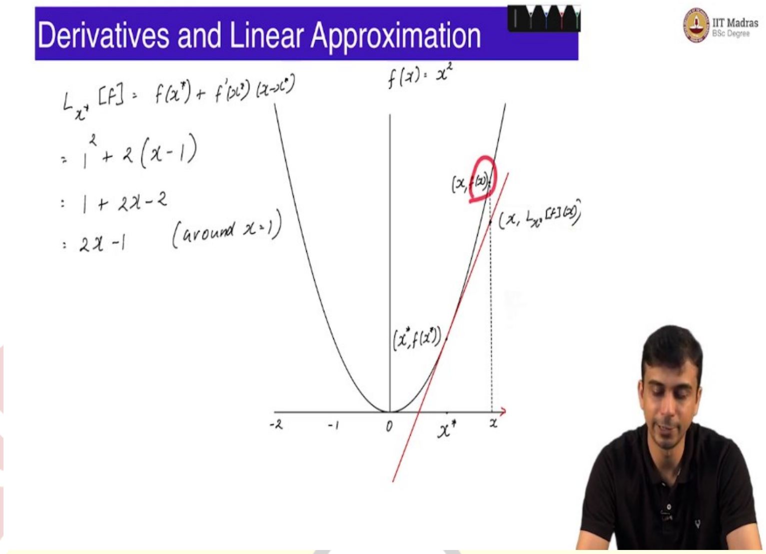
Linear approximation essentially makes this a more formal way. I mean, what approximation are you, by approximating compound interest by simple interest, what you are doing is essentially linear approximations. And this is a valid approximation as long as the rate of interest is small enough, that is, in this case, 1 - 1 percentage is a small enough rate that you can do this in a valid way. And later now, later, we will see that this is not always valid. And in fact, for large enough let us say instead of 1 percentage loss, we had 10 percentage loss.

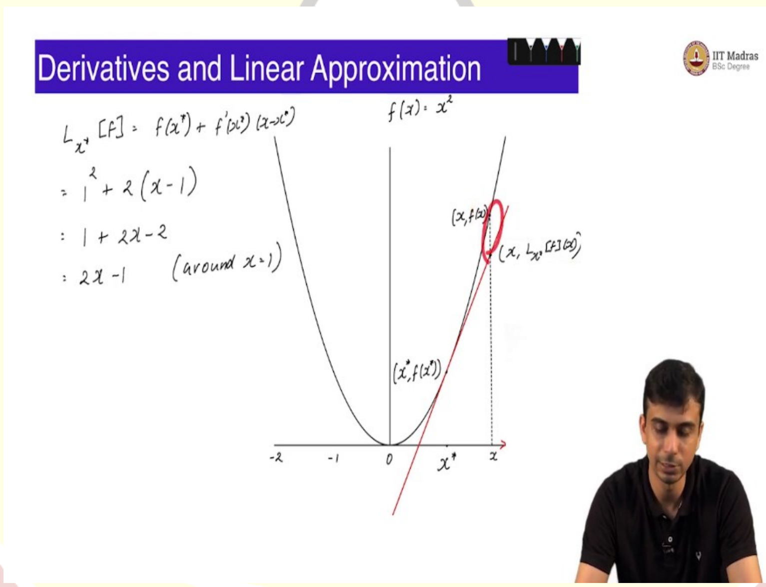
(Refer Slide Time: 20:08)





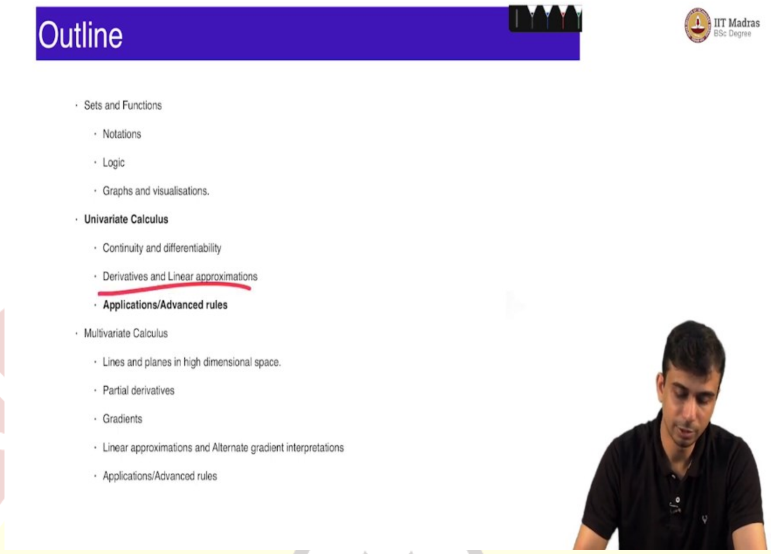
INDIAN INSTITUTE OF TECHNOLOGY MADRAS · सिद्धिर्भवति कर्मजा ·





This would no longer be true, because of the same reason why, L x * of f is only an approximation of f around x equals x *. As you move farther and farther away from x *, your Lx * of f at evaluated at x is not going to be exactly= f of x, it is going to be different from f of x.

(Refer Slide Time: 20:29)



So, with that, we have come to the end of derivatives and linear approximations. We will have a look at some slightly more advanced rules and applications of the derivative and linear approximations now.

Outline

- Sets and Functions
    - Notations
    - Logic
    - Graphs and visualisations.
- **Univariate Calculus**
    - Continuity and differentiability
    - **Derivatives and Linear approximations**
    - Applications/Advanced rules
- Multivariate Calculus
    - Lines and planes in high dimensional space.
    - Partial derivatives
    - Gradients
    - Linear approximations and Alternate gradient interpretations
    - Applications/Advanced rules

Derivatives and Linear Approximation

Let $f : \mathbb{R} \to \mathbb{R}$ be a diff function

$$f'(x^*) = \lim_{x \to x^*} \frac{f(x) - f(x^*)}{x - x^*}$$

$$f'(x^*) \simeq \frac{f(x) - f(x^*)}{x - x^*}$$ (around $x = x^*$)

$$f(x) \simeq \underbrace{f(x^*) + f'(x^*)(x - x^*)}_{L_{x^*}[f](x)}$$ (around $x = x^*$)

$$f(x) \simeq L_{x^*}[f](x)$$ around $x = x^*$

Derivatives and Linear Approximation

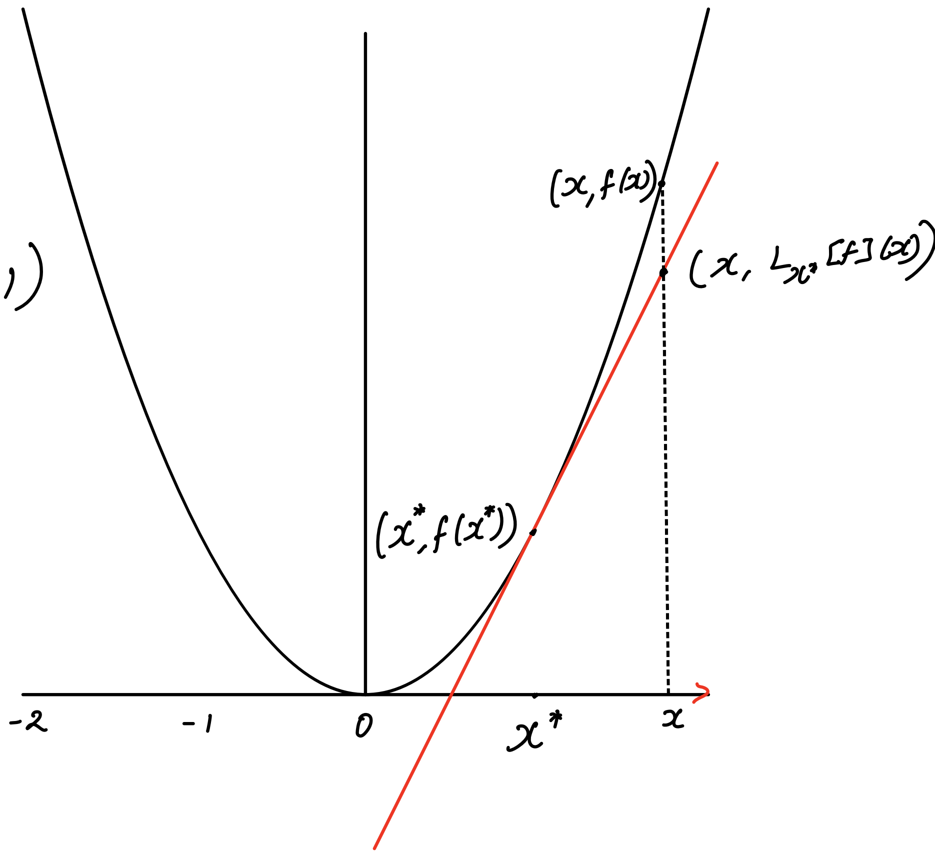
$$L_{x^*}[f] = f(x^*) + f'(x^*)(x - x^*)$$

$$= 1^2 + 2(x-1)$$

$$= 1 + 2x - 2$$

$$= 2x - 1 \qquad (\text{around } x = 1)$$

$f(x) = x^2$



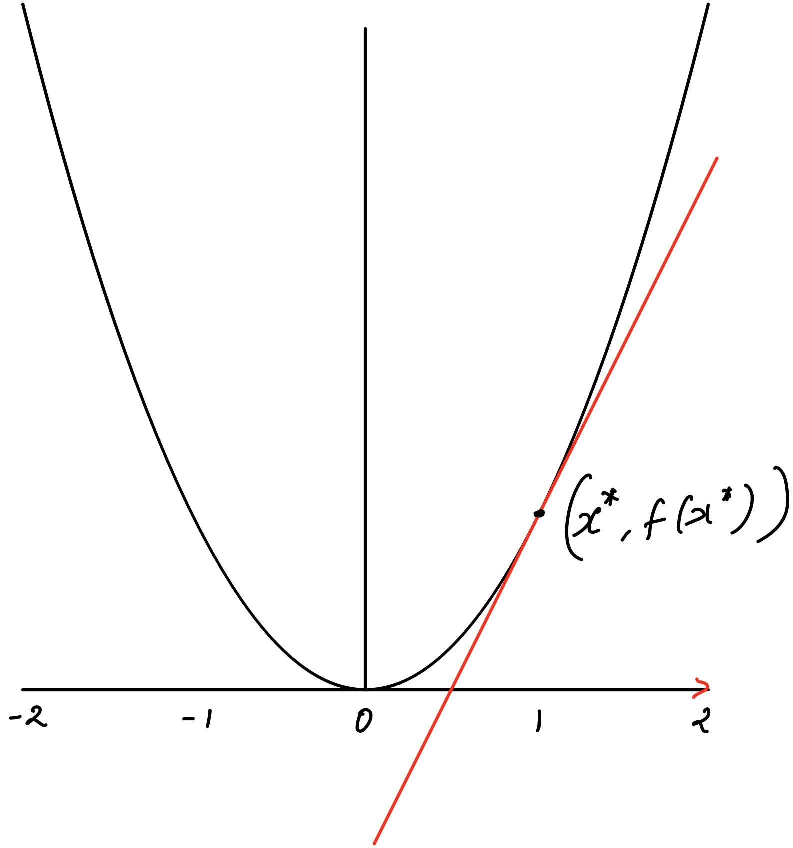
Linear Approximations and Tangent Lines

$G_{L_{x^*}[f]}$ is a

$\cap$

$\mathbb{R}^2$

tangent to the graph of f G_f at the point $x^*, f(x^*)$



Derivatives and Linear Approximation

i) Linear approximation of $f(x) = \sin(x)$ around $x^* = 0$

$f'(x) = \cos(x)$

$f'(x^*) = 1$

$f(x^*) = 0$

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*)$$

$$= 0 + 1(x - 0)$$

$$= x \quad \text{around } x = 0$$

$$\sin x \approx x \quad \text{if } x \approx 0$$

(ii) $f(x) = e^x$. around $x^* = 0$

$$e^x \approx e^0 + (x - 0) \cdot 1$$

$$\approx 1 + x \quad \text{around } x = 0$$

Derivatives and Linear Approximation

iii)

$\ln(1+x)$ around $x^* = 0$

$f'(x) = \frac{1}{1+x}$

$f'(x^*) = 1$

$f(x^*) = 0$

$\ln(1+x) \approx \ln(1) + 1(x-0)$

$\approx x$ around $x = 0$

iv) $f(x) = (1+x)^r$ around $x^* = 0$

$f(x^*) = 1$

$f'(x) = r(1+x)^{r-1}$

$f'(x^*) = r$

$(1+x)^r \approx 1 + r(x)$

$= 1 + rx$ around $x = 0$.

(v) $(0.99)^7$

(a) 0.95 (b) 0.93

(c) 0.91 (d) 0.9

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- **Univariate Calculus**
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - **Applications/Advanced rules**
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- **Univariate Calculus**
  - **Continuity and differentiability**
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Continuity of Functions

$$f : \mathbb{R} \to \mathbb{R}$$

is continuous at $x^* \in \mathbb{R}$ if for all sequences $x_1, x_2 \cdots$ converging to x^* we have that $f(x_i)$ converges to $f(x^*)$

$$\lim_{i \to \infty} x_i = x^* \Rightarrow \lim_{i \to \infty} f(x_i) = f(x^*)$$

$$\lim_{x \to x^*} f(x) = f(x^*)$$

E.g 1 : $f(x) = x^2$, $x^* = 2$

x_i : $3, 2.5, 2.25, \cdots \to 2$

$f(x_i)$: $9, 6.25, 2.25^2, \cdots \to 4$

Continuity of Functions

e.g 2 : $f(x) = \text{sign}(x) = \begin{cases} -1 & \text{if } x < 0 \\ +1 & \text{if } x = 0 \\ +1 & \text{if } x > 0 \end{cases}$

$x^* = 0$

$x_i : 1, \frac{1}{2}, \frac{1}{4}, \cdots \rightarrow 0$

$f(x_i) : 1, 1, 1, \cdots \rightarrow 1$

$x_i : -1, -\frac{1}{2}, -\frac{1}{4}, \cdots \rightarrow 0$

$f(x_i) : -1, -1, -1, \rightarrow -1$

Continuity of Functions

E.g.3 : $f(x) = \begin{cases} 2x+1 & \text{if } x > 1 \\ 3 & \text{if } x \leq 1 \end{cases}$

E.g.3 $f(x) = \frac{1}{x}$,

x_i : 1, 1/2, 1/4, ...

$f(x_i)$: 1, 2, 4, 8

E.g.3 $f(x) = \cos(1/x)$

x_i : 1, 1/2, 1/4, ...

$f(x_i)$: cos(1), cos(2), cos(4),

Differentiability of Functions

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x^* \in \mathbb{R}$

if $\lim_{x \to x^*} \dfrac{f(x) - f(x^*)}{x - x^*}$ exists.

$f'(x^*)$

f is NOT continuous at x^* $\Rightarrow$ f is NOT differentiable at x^*.

E.g. 1 $f(x) = |x|$

$x_i : 1, \frac{1}{2}, \frac{1}{4}, \ldots \to 0$

$\dfrac{f(x_i) - f(0)}{x_i} : 1, 1, 1 \to 1$

$x_i : -1, -\frac{1}{2}, -\frac{1}{4} \ldots$

$: -1, -1, -1$

Differentiability of Functions

E.g : 2

$$f(x) = \begin{cases} 4x+2 & \text{if } x \geq 2 \\ 2x+8 & \text{if } x < 2 \end{cases}$$

E.g : 3

$$f(x) = \begin{cases} 4x+2 & \text{if } x \geq 2 \\ 2x+6 & \text{if } x < 2 \end{cases}$$

$$\lim_{x \to 2^+} \frac{f(x) - f(2)}{x-2} = 4$$

$$\lim_{x \to 2^-} \frac{f(x) - f(2)}{x-2} = 2$$

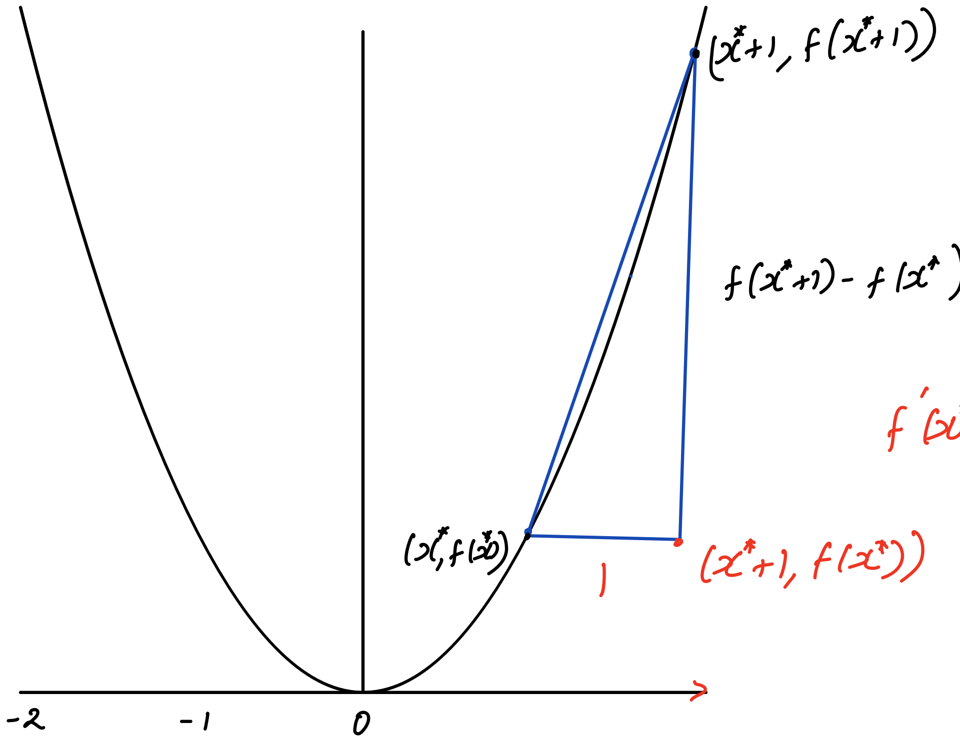
Differentiability of Functions

E.g. 4:

$$f(x) = \begin{cases} 4x+2 & \text{if } x \geqslant 2 \\ x^2+6 & \text{if } x < 2 \end{cases}$$

$$\lim_{x \to 2^+} = \lim_{x \to 2^-} = \frac{f(x) - f(2)}{x-2} = 4$$

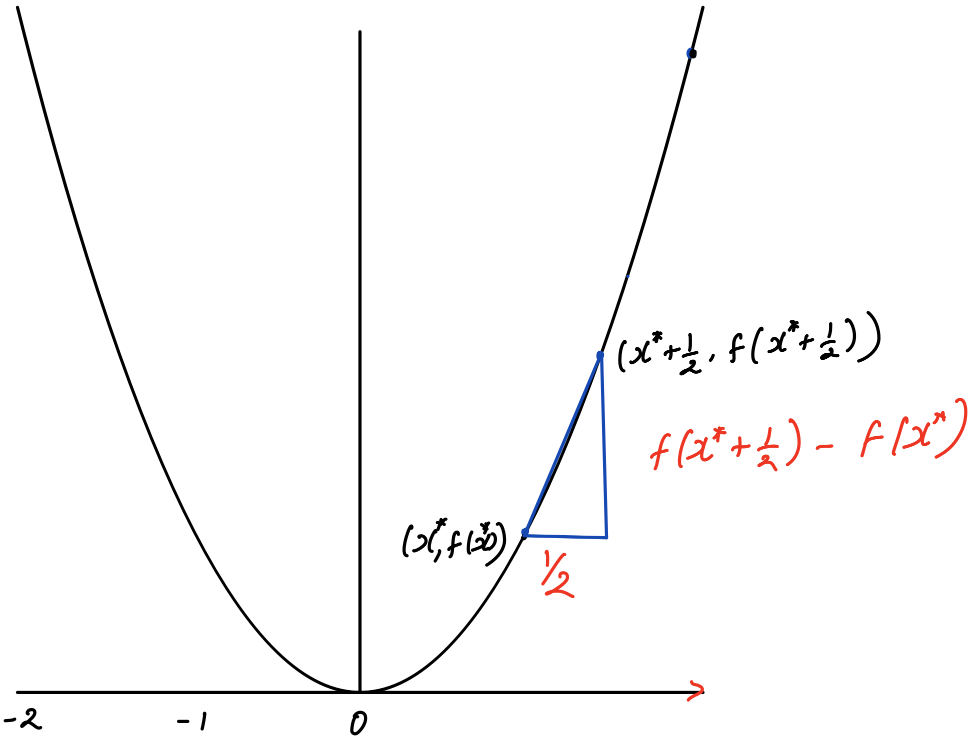
Differentiability of Functions



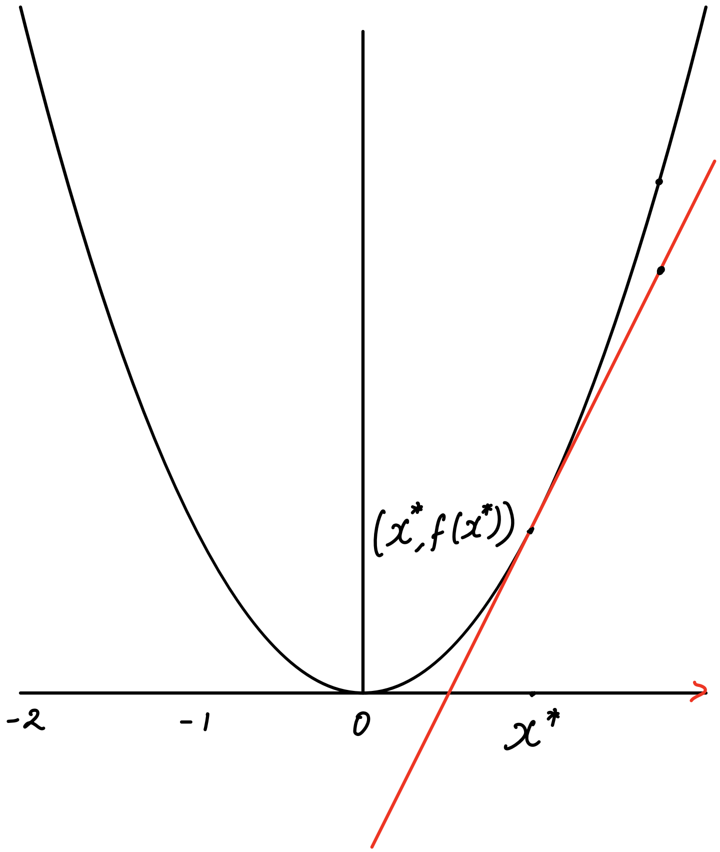
$$f'(x^*) = \lim_{x \to x^*} \frac{f(x) - f(x^*)}{x - x^*}$$

$$f'(x^*) = \lim_{h \to 0} \frac{f(x^*+h) - f(x^*)}{h}$$

Differentiability of Functions



Differentiability of Functions



Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- **Univariate Calculus**
  - Continuity and differentiability
  - **Derivatives and Linear approximations**
  - Applications/Advanced rules
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- **Univariate Calculus**
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - **Applications/Advanced rules**
- Multivariate Calculus
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Higher Order Approximations

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*) \quad \text{(Linear apx)}$$

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2} f''(x^*)(x - x^*)^2 \quad \text{(quadratic)}$$

e.g.: 1

$$f(x) = x^2$$

$$f'(x) = 2x$$

$$f''(x) = 2$$

$$x^2 \approx (x^*)^2 + 2x^*(x - x^*) + \frac{1}{2} \cdot 2 \cdot (x - x^*)^2$$

$$= x^2$$

Higher Order Approximations

Approx e^x around $x^* = 0$

$$e^x \approx e^0 + e^0 (x - x^*) + e^0 \cdot \frac{1}{2} \cdot (x - x^*)$$

$$= 1 + x + \frac{x^2}{2}$$

Ex: Which is closest to $(1.1)^7$

(a) 1.7 (b) 1.9 (c) 2.1 (d) 2.3

$f(x) = (1+x)^7$, $f'(x) = 7(1+x)^6$, $f''(x) = 42(1+x)^5$

$f'(0) = 7$ $f''(0) = 42$

$$f(0.1) \approx 1 + 7(0.1) + \frac{1}{2} \cdot 42 (0.01)$$

$$= 1.91$$

Higher Order Approximations

Product Rule

$$f(x) = g(x) \cdot h(x)$$

$$f'(x) = ? \qquad x^* = 0$$

$$f(x) \approx (g(0) + x\, g'(0))(h(0) + x\, h'(0))$$

$$= g(0)h(0) + x\left[g'(0)h(0) + h'(0)g(0)\right]$$

$$+ x^2 g'(0)h'(0)$$

$$L_x[f] = f(0) + x f'(0)$$

$$f'(0) = g'(0)h(0) + h'(0)g(0)$$

Chain Rule

$$f(x) = g(h(x))$$

$$\approx g\left(h(0) + h'(0)x\right)$$

$$\approx g(h(0)) + g'(h(0))\left[h(0) + h'(0)x - h(0)\right]$$

$$= g(h(0)) + g'(h(0))\,h'(0)\,x$$

$$f(x) \approx f(0) + f'(0)\cdot x$$

$$f'(0) = g'(h(0))\,h'(0)$$

Chain Rule

(i) $\dfrac{e^{3x}}{\sqrt{1+x}}$ Give LA around $x=0$

$$\frac{e^{3x}}{\sqrt{1+x}} \approx (1+3x)\left(1-\frac{x}{2}\right)$$

$$\approx 1+\frac{5}{2}x$$

(around $x=0$)

(ii) Give Lin. Apx.

$e^{\sqrt{1+x}}$ around $x=1$

$$e^{\sqrt{1+x}} \approx e^{\sqrt{2}} + \frac{e^{\sqrt{2}}}{2\sqrt{2}}(x-1)$$

(around $x=1$)

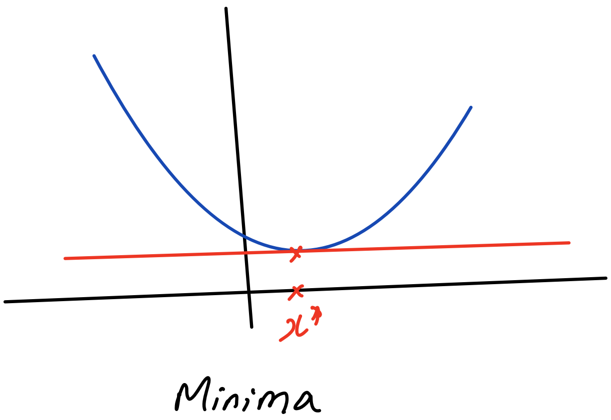
Maxima, minima and saddle points

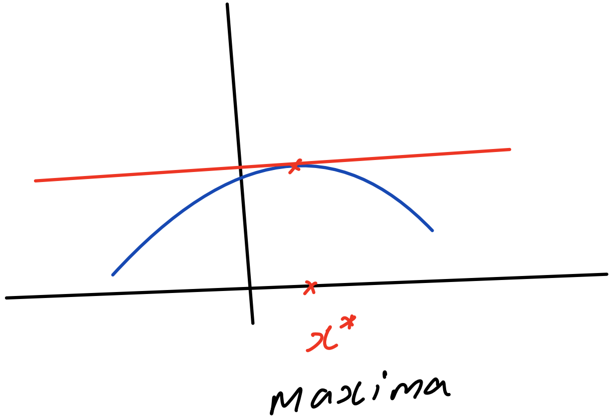
$$L_{x^*}[f] = f(x^*) + f'(x^*)(x - x^*)$$

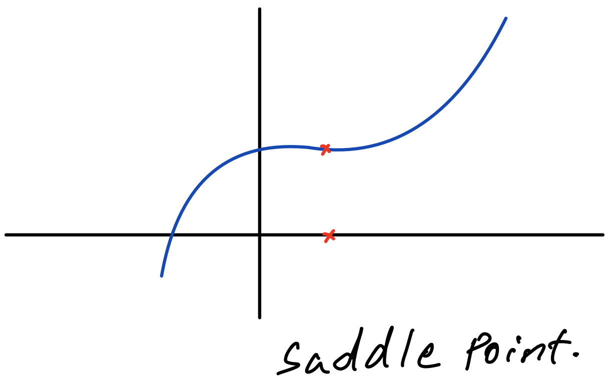
$f'(x^*) = 0 \iff x^*$ is a critical point of f

$$L_{x^*}[f] = ?$$

Maxima, minima and saddle points







Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - **Lines and planes in high dimensional space.**
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

$f : \mathbb{R} \to \mathbb{R}$

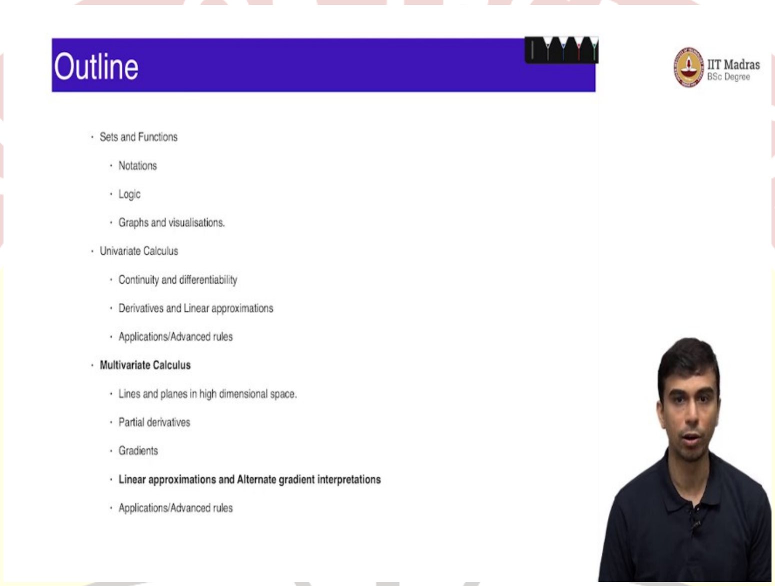
$f : \mathbb{R}^d \to \mathbb{R}$

IIT Madras

ONLINE DEGREE

Machine Learning Foundations
Professor. Harish Guruprasad Ramaswamy
Department of Computer Science & Engineering
Indian Institute of Technology, Madras
Multivariate Calculus: Linear approximation and applications

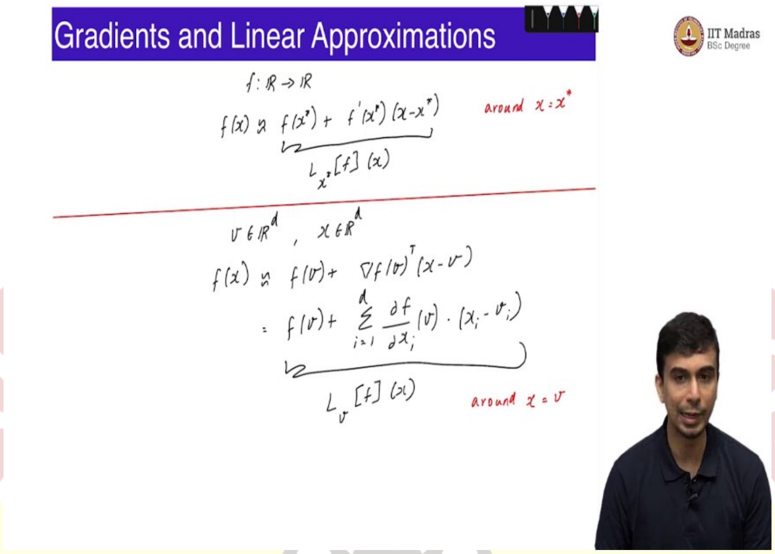
(Refer Slide Time: 0:13)



Hello, everyone, and welcome to another lecture on machine learning foundations. In the last class, we introduced and define partial derivatives and gradients. And that is the first and standard interpretation of a gradient, which is it is a package containing the components, all the components of the partial derivatives. You have the partial derivatives with the respect of all variables and put them all together into one vector that is the gradient that is the computational interpretation of the gradient.

In today's lecture, we will take a look at the other interpretations of the gradient which are more geometric and more intuitive in nature than the computational approach.

(Refer Slide Time: 0:51)



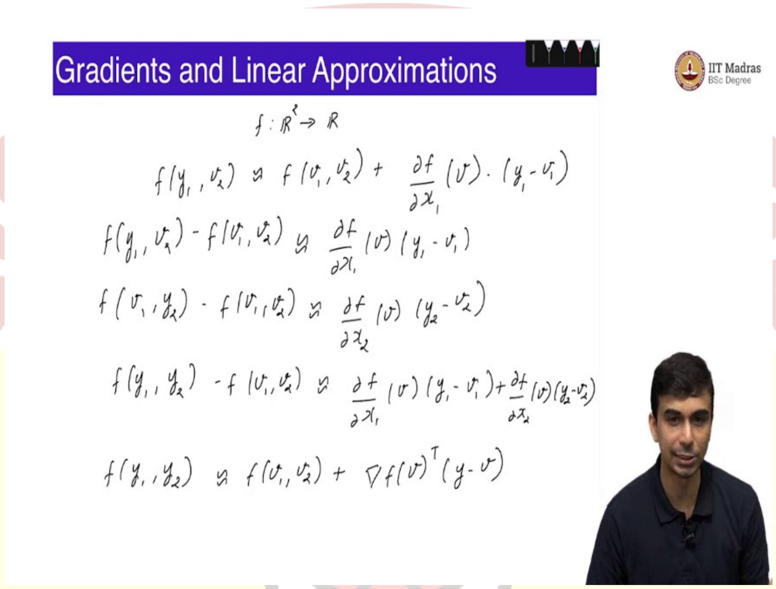
The first of which is the linear approximation interpretation. We already have seen linear approximations for one dimensional functions. If you have a function f from R to R, and you want to approximate this function f around x close to x star this is what you would do, f of x is approximately equal to f of x star + f dash of x star into x - x star. This is what you have been doing for linear approximations. And we denoted this term as the linear approximation of f around x star evaluated at x. And this is a valid approximation when x is close to x star. This is for one dimensional functions.

How can we do the same thing for functions from Rd to R? In other words, the it is, x, now imagine v is a vector, a d dimensional vector and x is in its neighborhood, x is a d dimensional vector and x is the close to v. Now, you want to be able to approximate x, the answer, I will give the answer first and then give various interpretations of it, which is f of x is approximately equal to f of v + grad f of v transpose x - v. You can see the similarities to the one dimensional case. So, f of x, you first have the f of v here, which corresponds to f of x star here, and you have grad f of v transpose x - v, which corresponds to f dash of x star into x - x star.

We will break this down even further as f of v + sum over i equal to 1 to d doh f by doh xi into xi - vi. I have just broken down the gradient into its components, doh f by doh xi evaluated at the point v multiplied by xi - vi. And we will denote this also in a very similar way. We will call this as the linear approximation of f around the point v evaluated at x.

And when is this approximation valid, this approximation is valid around x, approximately around x equal to v. When x is close to v, this is a valid approximation of f. So, we will give one interpretation or one possible interpretation of this lean approximation.

(Refer Slide Time: 3:55)



The way we do that is by breaking down into multiple one dimensional functions. We will make life easier for ourselves and just consider a function, two dimensional function. This easily extends to multi dimensional functions as well, but we will do that for just two dimensional function. Let us say an f from R2 to R, f of x1 , v2 is approximately equal to f of v1 , v2.

We are viewing the second variable as a constant, the first variable as the variable. So, we are approximately around v1. So, f of x1 , v2 is approximately equal to f of v1 , v2 + doh f by doh x1 evaluated at v into x1 - v1. Here is a small notation issue. We will instead of calling it x1, we will call it y1. What is, this is just a standard one dimension linear approximation, that is we are approximating the function f as viewing as just as a function of its first argument, that is f of y1 , v2 is approximately equal f of v1 , v2. And we are using the partial derivative of f with respect to the first argument at v × y1 - v1.

We will simplify, we will change the notation a little bit more, f of y1 , v2 - f of v1 , v2 is approximately equal to doh f by doh x1 at v into y1 - v1. So, you can do the same thing for the second argument also, that is f of v1 , y2 - f of v1 , v2 this is approximately equal to doh f by doh x2, which is the second argument, at v into y2 - v2.

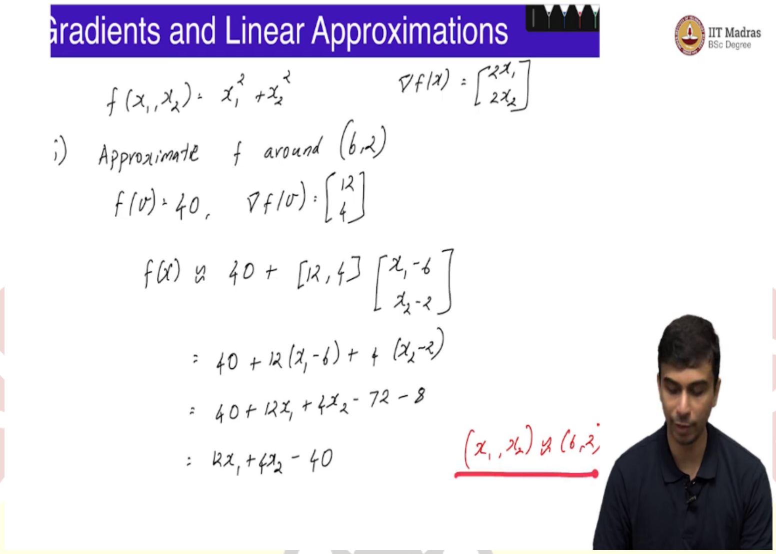
So, let us interpret these two approximations. What this says is, as I move from v1 , v2 to y1 , v2, that is as I move along the first axis, v1 changes to y1, v2 remains the same. This is the change in the value of the function. The value of the function changes by doh f by doh x1 evaluated at v into y1 - v1.

Similarly, here, this expression says that as I move from v1 , v2 to v1 , y2 the change is given by doh f by doh x2 evaluated at v into y2 - v2. But what we are now interested in is what if both of them move simultaneously? What if, what can I say about f of y1 , y2 - f of v1 , v2. The simplest answer is to, well, if one variable changes, this is the change, if another variable changes is the change. If both variables change, the change is simply given by the addition of these two changes, which is doh f by doh x1 at v into y1 - v1 + doh f by doh x2 at v into y2 - v2.

Or in other words, now we will write down the full linear approximation, which is f of y1 , y1 is approximately equal to f of v1 , v2 + doh f by doh x1 evaluated at v × y1 - v1 + doh f doh x2 evaluated at v × y2 - v2. Here, we are going to make our life easier. And I am just going to call it the gradient of f at v transpose y - v. This is exactly our linear approximation expression.

Remember, that is the gradient of f of x is approximately f of v + grad f of v transpose x - v. Or in other words, f of y is approximately equal to f of v + grad f of v transpose y - v. Y and x are just dummy variables, they do not really matter.

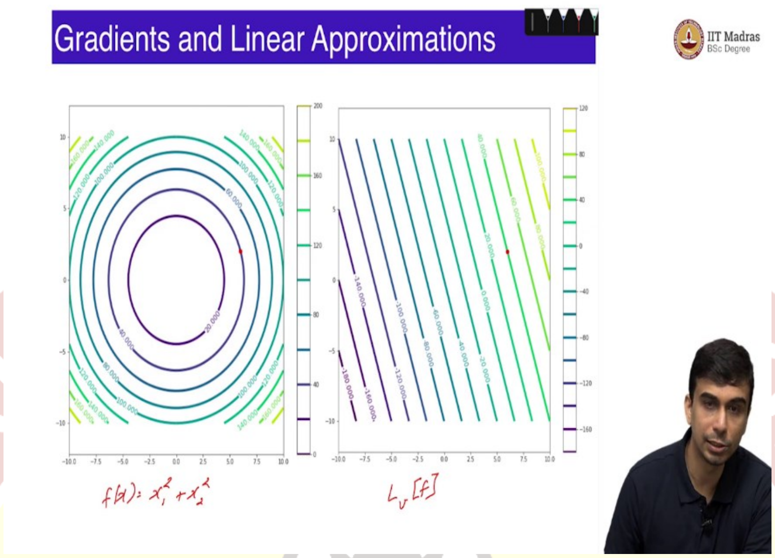
(Refer Slide Time: 8:36)

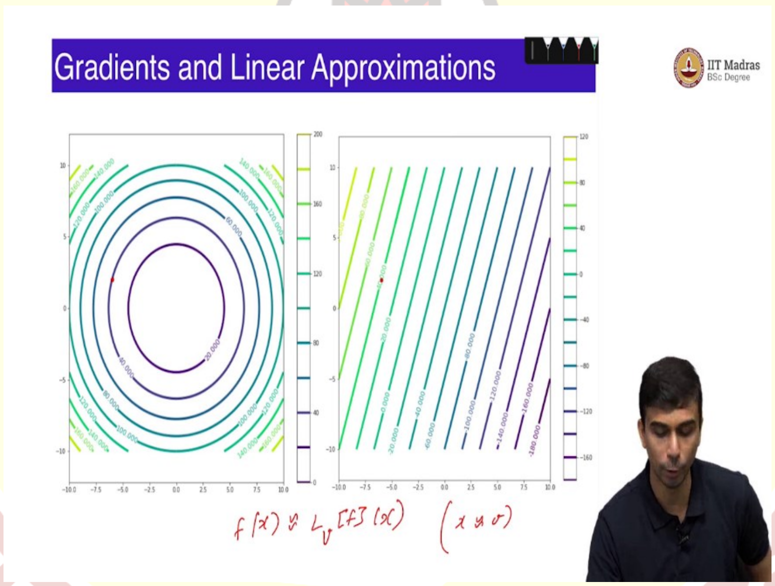


So, here are a couple of examples to run. So, we will take a simple example first. Let us consider the function f from R2 to R, which is f of x1 , x2 is equal to x1 square + x2 square. Now, here is the question, approximate f around, I want the linear approximation of f around the point let us say 6 , 2. So, v is equal to 6 , 2 in our case. What is f of v first? F of v is equal to 6 square + 2 square which is 40. What is grad f of v? Grad f of v is 2v1, 2v2 which is 12 , 4. This is because grad f of x is 2x1, 2x2, so grad f evaluated at v in this case 2v1, 2v2 which is 12 , 4.

With this we can actually give the answer which is f of x is approximately equal to 40 + grad f of v, which is transposed 12 , 4 × x - v, which is x1 - 6, x2 - 2. We can simplify this further and say 40 + 12 into x1 - 6 + 4 into x2 - 2, which would be 40 + 12x1 + 4x2 - 72 - 8, which would be 12x1 + 4x2 - 40. This is the linear approximation of f, which is x1 square + x2 square around the point 6 , 2. So, this is particularly valid around what point, well, x1 , x2 is approximately 6 , 2. When your vector x is a neighborhood of 6 , 2, this is a valid approximation.

(Refer Slide Time: 11:04)





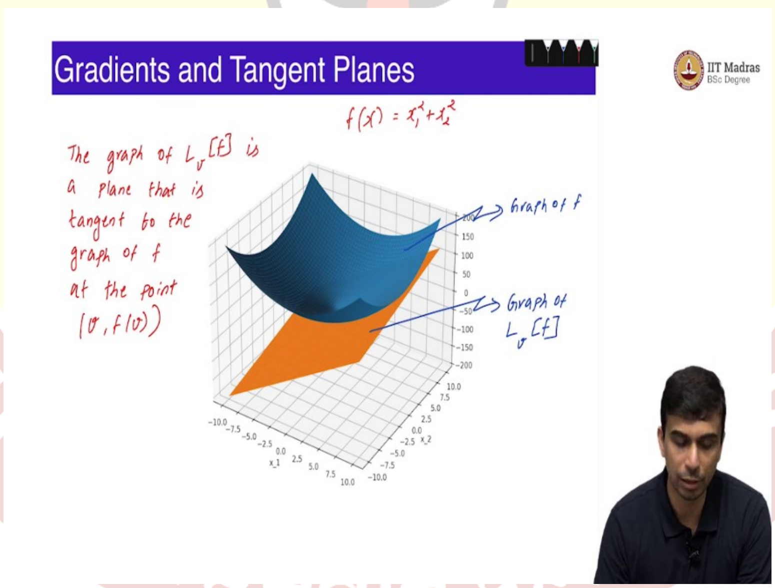
Let us visualize this further. Here is a contour plot of the function f, which is f of x is equal to x1 square + x2 square. Here is the point 6 , 2. And on the right you have the contour plot of the function, Lv of f, which we just derived here, which is 12x1 + 4x2 - 40. And because this is a linear function, you can see that the contours are all straight lines and they are equally spaced. This is a linear function.

A couple of things to note immediately, which is f of x evaluated at 6 , 2, the value is equal to 40. Consider this circle has value 40. And you can see that the linear approximation of f, are the point 6 , 2 also has the exact value 40. So, the linear approximation is exactly equal to the function value at the point of approximation. At other points, you do not know, but it is exactly equal at the point of approximation. And there are other properties also. But that is the first property that you should keep in mind.

And the linear approximation does change with the point that you choose. If you want approximate the function f around the point 6 , 2, this is the linear approximation. But if you want to approximate the same function around a point, let us say - 6 , 2, the linear approximation changes. You still have that property that is at the point, you have 40 here and 40 here. The f of v is exactly equal to Lv of f evaluated at v.

At a general point x, this is not true. At the general point x, you only have approximation, f of x, the one to change v to x, this is not exactly equal. This is only approximately equal. And that is also true only when x is approximately equal to v, only when around x equal to v do we have this expression. So, that is the linear approximation viewpoint. And the approximation is completely defined by the gradient.

(Refer Slide Time: 13:20)



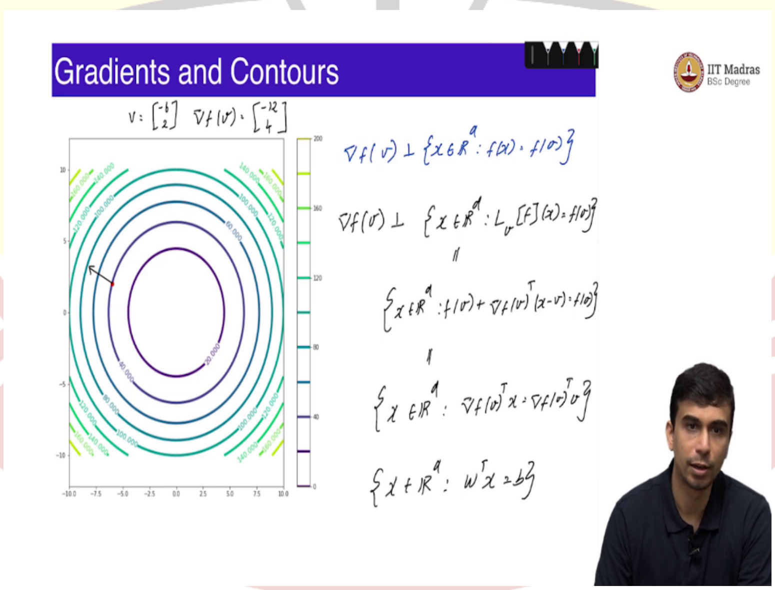
And the same way that we have a linear approximation viewpoint for one dimensional functions and tangent line viewpoint, we can have a very similar viewpoint here, which is you can use tangent planes as graphs.

Let us, this is the tangent plane interpretation of the gradient. That is, if you have the graph of the linear approximation of f around the point v, that is the graph of Lv of f, which is going to be a plane. Recall, graph of a function f from Rd to R is a subset of Rd + 1. Here you have a three dimensional plot of the function f of x is equal to x1 square + x2 square. So, the graph of Lv of f is a plane, because Lv of f is a linear function and linear functions, the graph of linear functions correspond to planes.

A graph of Lv of f is a plane that is tangent to the graph of f at the point v , f of v. So, v is a d dimensional vector and if v of v is a scalar, so this is together is a d + 1 dimensional vectors. So, for example, here you have the blue curve corresponds to the blue surface corresponds to the graph of f. And the orange plane corresponds to the graph of the linear approximation of f around the point v.

You can see that this graph of the linear approximation is a plane which is given and it touches the surface, the blue surface, at a point. And that point is exactly equal to v , f of v. So, this is the other interpretation. That is we have the linear approximation interpretation of gradient, then we have the tangent plane interpretation. The tangent plane is also given by the linear approximation, which in turn is given by the gradient.

(Refer Slide Time: 15:48)



And we have the next interpretation, which uses the contour sets. Here is the statement for the gradient, which as you can see that the gradient of f evaluated at, let us say, at some point v is going to be perpendicular to the contour set given by or the level set given by the set of x in Rd, such that f of x is equal to f of v.

For example, let us say you take the point - 6 , 2 and the gradient of f at that point is simply - 12 , 4. So, what I have done here is I have taken the point - 6 , 2 and I have plotted the vector - 12 , 4 as a vector here. The lengths do not matter, only the direction matters. You can see that the direction, this black arrow is almost perpendicular. It is exactly perpendicular to this level set.

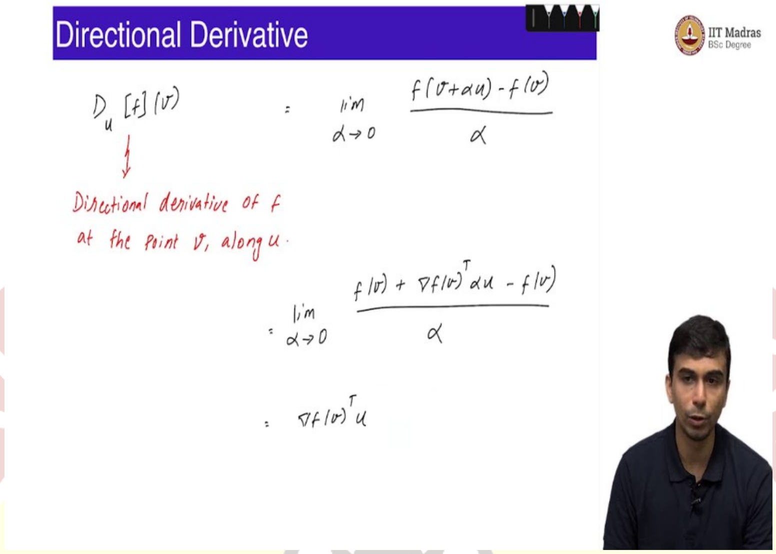
What is this level set? This level set is exactly equal to the set of all x in Rd, such that f of x equal to f of v, which is f of x equal to 40. This is the level set corresponding to f of x equal to f of v, which is f of x equal to 40. And you have the gradient of the function f at v is - 12 , 4, which is perpendicular at that point. And I am not going to prove this completely. But what I am actually going to prove is this statement, which is quite obvious, which is that the gradient of f at v is actually perpendicular to the set x in Rd such that the linear approximation of f evaluated at x is equal to f of v.

So, these two statements are not the same, but, not same as of now, but we will, they are actually the same under certain extra regularity assumptions, but this statement over is very easy to prove. Why is that, because this on the set on the right side is actually a plane. Why is that you have the approximation on f at x is a linear function of x. And this is exactly, planes have a perpendicular vector inbuilt to them. What is that? Well, we will evaluate this set. This set is simply the set of x in Rd such that f of v + grad f of v transpose x - v is equal to f of v.

We will do some moving, term moving around, we get this is the set of x in Rd such that grad f of v transpose x is equal to grad f of v transpose v. And this for the equation of a plane is set of x in Rd such that w transpose x is equal to b. If you have such a plane, you will already know that the vector w is perpendicular to that plane. And you can see that this is exactly of that form grad f of v corresponds to w.

So, grad f of v is actually perpendicular to this plane, which we have a proved now. So, that is the other interpretation of the gradient that is the gradient of a function at a point v is going to be perpendicular to the contour passing through that point.

(Refer Slide Time: 19:27)



Next we move to another useful application or interpretation of the gradient which has to do with directional derivative. What is the directional derivative? We have already seen an example of directional derivative which is, well, we will first say an English redefinition of the directional derivative. We will define directional derivative of the function f at the point v along the direction u. So, you have the function f from Rd to R, u and v are vectors in D dimensional space, u and v are elements in Rd.

So, this is the notation for, I am going to read this as the directional derivative of the function f. I will write this, directional derivative of the function f at the point v along the direction u. This is the, this is how you read the statement. What is the definition of this? Well, the definition is going to be, first create some space here. This is limit α approaching 0, f of v + α u - f of v by α. Let us try to interpret this expression.

What this saying is, we want to move from v to along the direction u. You are at the point v and you want to move along the direction u and you are trying to quantify the rate of change of the function as you move along the direction u. That is the direction and derivative of f along u. How will we compute this thing?

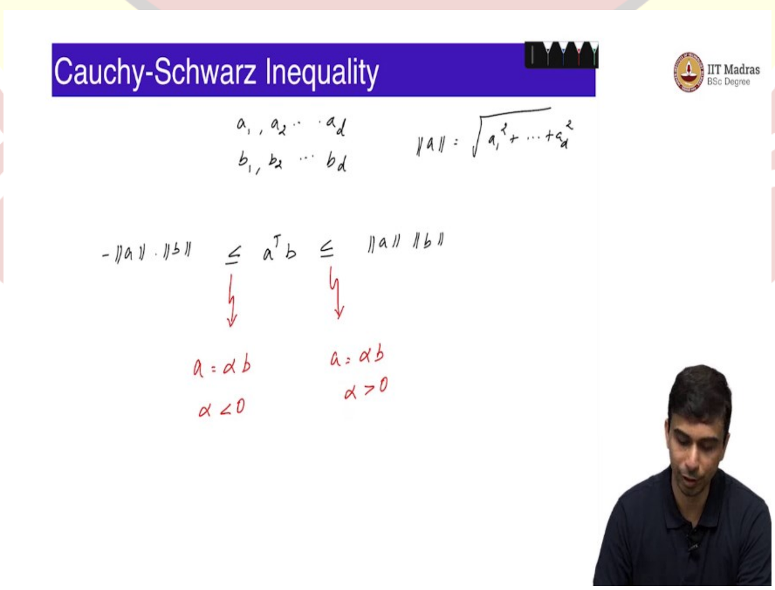
Here, you can use our linear approximations, because α is going to approach 0, you can approximate f of v + α u using linear approximations. This is going to be a limit α approaches 0, f of v + α u is approximately equal to f of v + grad f of v transpose α u - f of v by α. You can see that f of v cancels and α also cancels. So, you are left with grad f of v transpose u.

So, this is nice and particularly interesting because it says that to compute the directional derivative of the function along any direction, all that you need to do is to take the . product of the direction with the gradient. And this is a, in particular, this is nice, because you want to, you are at, let us say, in machine learning, in practice, you often want to minimize a function, which means that you are most approaches take an iterative thing where you are at the current iterate, let us say v, and you want to decide which direction to move.

And you want, you are going to move to new it new iterate from v to let us say v dash, but you want to decide which direction to go. A very natural notion is you are at, you can choose to move at the direction, which decreases the function the most or increases the function the most if you want to maximize it. So, the notion that you are looking for is to what direction to move so that you can increase or decrease the function the most. And the directional derivative is a very important tool for doing that.

We will actually attack that next. And before we do that, however, we will take a slight detour to attack to, actually give a very basic inequality known as the Cauchy-Schwarz inequality, which is useful for controlling the inner product of two vectors grad f of v transpose u.

(Refer Slide Time: 23:28)

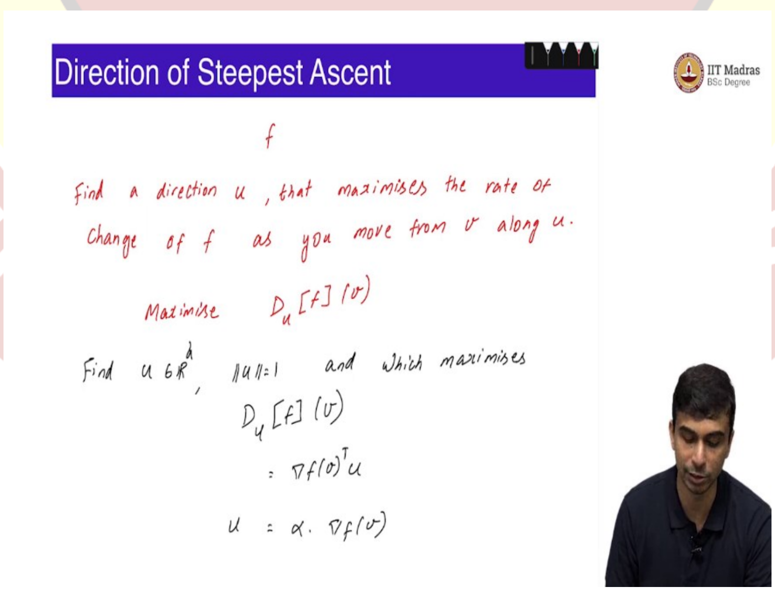


What is that? Well, you have, let us say you have two vectors a and b, whose, they are d dimensional vectors, their components are a1, a2, ., ., . ad, and similarly, b1, b2, ., ., ., bd. You have two d dimensional vectors a and b, and you want to be able to control their inner product. You want to be able to control a transpose b. What we can show is a transpose b is upper bounded and lower bounded by these two terms, which is upper bounded by norm a × norm b and lower bounded by - norm a into norm b.

This norm is the standard Euclidean norm that we have already defined, which is norm a is simply square root of a1 square + a2 square + ., ., . ad square. So, a transpose b is upper bounded and lower bounded between these two terms. And you also happen to know that when does this equality happen. You want to know when this equality happens. And there is a very clear characterization for that.

This equality happens when a is equal to α × b, where α is a scalar, which is negative. If you have a negative scalar α and a is equal to α × b, then you have a transpose b is exactly equal to - norm a, norm b. And when does this equality happen? Well, this equality also happens when a is equal to α b, but α is a positive scalar. So, in other words, if you want to maximize the term a transpose b, you will have, you will choose b is equal to, a is equal to α b and α to be greater than 0.

(Refer Slide Time: 25:27)



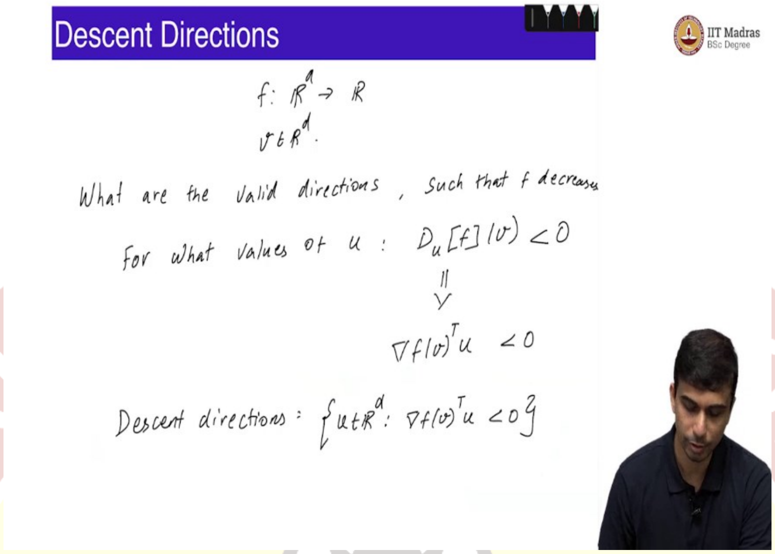
And this will all lead us to the next question here, which is, let me clean this. Now you want to find the direction of steepest ascent. Let us say now you are trying to maximize a function f, let us say you have a function f, and you want to maximize it. You are at a point current detractor is v. And you want to choose a direction, find a direction u that maximizes the directional derivative. You want to maximize the rate of change of f as you move from v along u.

In other words, you want to maximize the directional Du of f evaluated at v. Find u which maximizes this term. That is your goal. And we will do a little bit more formally now. What is the direction? A direction cannot have arbitrary length, so we will fix the direction to be a unit norm.

So, find u in Rd such that norm u is equal to 1 and which maximizes the directional derivative of f evaluated at v. But we already know this is exactly equal to grad f of v transpose u. So, v is fixed. The current point is fixed. The only thing that is not fixed is the direction along with u move. So, you want to find a u which maximizes grad f of v transpose u. And that is already given by the Cauchy-Schwarz inequality, which means that you should pick u to be a scalar multiple, a positive scalar multiple of grad f of v.

And you already know that u is a unit norm vector, so α is simply going to be 1 by norm of grad f of v. So, this gives the other interpretation of the gradient which is the gradient is simply the direction of the steepest ascent. If you are at a point v and you want to move along the direction which increases the function the fastest, you would choose the gradient. So, that is the other interpretation, which is the direction of steepest ascent.

(Refer Slide Time: 28:22)

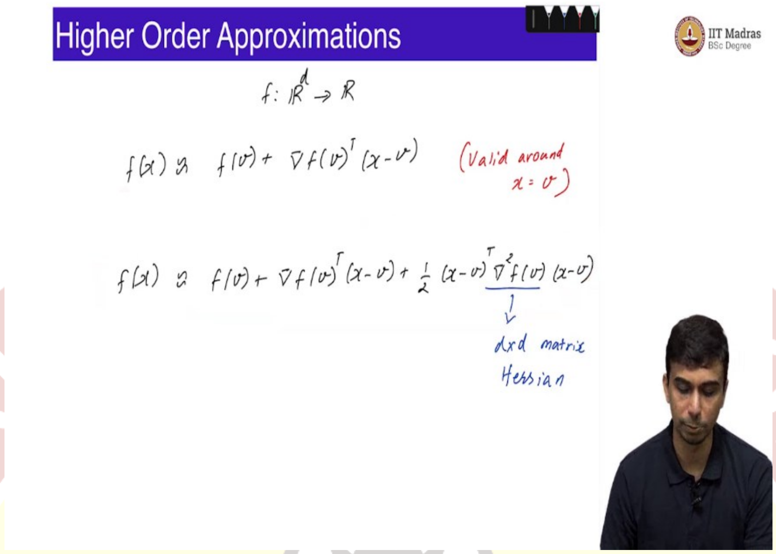


We can finally wrap this up by saying you are at a point let us say v. You have a function f from Rd to R. You are at a point v in Rd. Now, let us say you want to minimize this function. What all directions can you go, what are the valid directions, such that f decreases, or in other words, for what values of u then you say the directional derivative of f along u evaluated at v is less than 0 or in other words, this is grad f of v transpose u less than 0.

So, the set of descent directions is simply the set of u in Rd such that grad f of v transpose u is less than 0. So, this is another tool that you will need when you do optimization that is the set of directions along which you can move so that the function value decreases from the current value. So, that is your descent direction. These are all descent directions.

With that, we will wrap up the alternate grid interpretations. And there are some simple, I mean, there are some slightly advanced rules which I will briefly skim over. We will not go into great detail on these, because we will not be using most of these, but we will do that anyway.

(Refer Slide Time: 30:25)



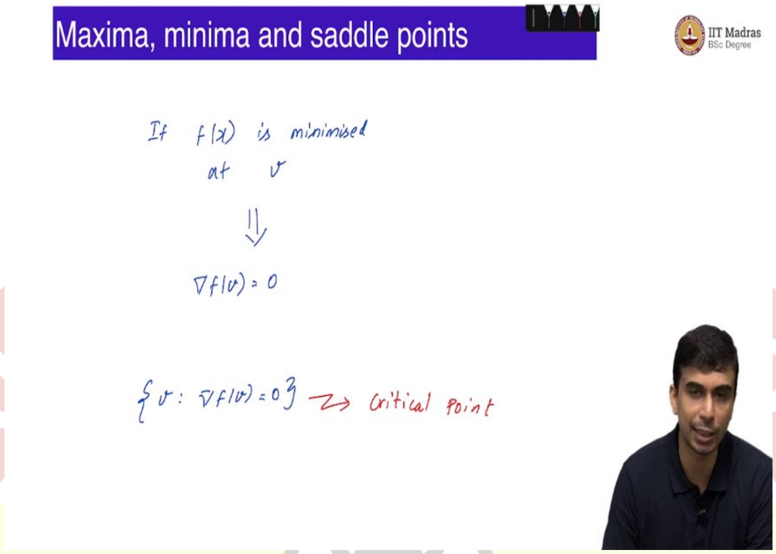
The same way that you did for functions f from R to R, you had linear approximations and you also had higher order quadratic approximations. You can do a very similar thing for functions f from Rd to R also as follows. You have a function f from Rd to R, the linear approximation f of, is simply given by f of x is approximately f of v + grad f of v transpose x - v. This is valid around x is equal to v. This is the linear approximation.

You can do a higher order approximation the same way that you did for one dimensional functions or this is little bit more complicated, which is you can also say f of x is approximately equal to f of v +, the first term looks the same, grad f of v transpose x - V. The second term on the other hand is half x - v transpose, well, the hessian of f evaluated at v into x - v. I will not go into great detail into expanding the second term, but it is analogous to the quadratic term that we had for one dimensional functions.

That is, you have f of x, this is exactly equal to the linear approximation and the extra term, you have this grad square f of v which is called the hessian. This term is a d cross d matrix called the hessian and x - v is a one cross v vector, x - v transpose is one cross v vector and x - v is a d cross one vector together this is a scalar.

I will not go into great detail on this, but you might find more advanced optimization algorithms use a quadratic approximation than the linear approximation, but most of machine learning you will be sticking to linear approximations which is this.

(Refer Slide Time: 32:55)



And the same way that you had maxima, minima and saddle points for one dimensional functions based on the derivative, you have a very similar situation here with gradients and maxima, minima and saddle points that is, if f of x is minimized at some point at, let us say, at some point v, then you have that grad f of v is the zero vector.

But it is not true the other way around, grad f of v equal to 0 does not mean that f of x is minimized at v. But however, if your function is minimized at v then grad f of v is equal to 0, but it is also true for maxima. So, any point where grad f of v equal to 0, the set of v such that grad f of v equal to 0 is called critical points. These points are called critical points. Note however that gradient is a vector. So, this 0 here corresponds to the vector zero not a scalar 0.

For special classes of functions, critical points can be shown to be exactly equal to minima, but in general critical points are not minima, I mean, not guaranteed to be minima or maxima. But in either case critical points are interesting points to analysis. If you want to find the maxima or minima of a function, you can just restrict your attention to points which satisfy this condition. This is called the first order necessary condition for optimality.

And you only focus your attention on critical points and check whether they can be candidate minima or maxima. This is another application of the gradient where you set the gradient to 0 and find interesting points which are critical points and check whether they satisfy your condition. With this, we will wrap up.

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - **Linear approximations and Alternate gradient interpretations**
  - Applications/Advanced rules

Gradients and Linear Approximations

$$f : \mathbb{R} \to \mathbb{R}$$

$$f(x) \approx \underbrace{f(x^*) + f'(x^*)(x - x^*)}_{L_{x^*}[f](x)}$$

around $x = x^*$

$$v \in \mathbb{R}^d, \quad x \in \mathbb{R}^d$$

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$

$$= \underbrace{f(v) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(v) \cdot (x_i - v_i)}_{L_v[f](x)}$$

around $x = v$

Gradients and Linear Approximations

$$f : \mathbb{R}^2 \to \mathbb{R}$$

$$f(y_1, v_2) \approx f(v_1, v_2) + \frac{\partial f}{\partial x_1}(v) \cdot (y_1 - v_1)$$

$$f(y_1, v_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1)$$

$$f(v_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) - f(v_1, v_2) \approx \frac{\partial f}{\partial x_1}(v)(y_1 - v_1) + \frac{\partial f}{\partial x_2}(v)(y_2 - v_2)$$

$$f(y_1, y_2) \approx f(v_1, v_2) + \nabla f(v)^T (y - v)$$

Gradients and Linear Approximations

$$f(x_1, x_2) = x_1^2 + x_2^2 \qquad \nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

(i) Approximate f around $(6,2)$

$$f(v) = 40, \qquad \nabla f(v) = \begin{bmatrix} 12 \\ 4 \end{bmatrix}$$

$$f(x) \approx 40 + [12, 4] \begin{bmatrix} x_1 - 6 \\ x_2 - 2 \end{bmatrix}$$

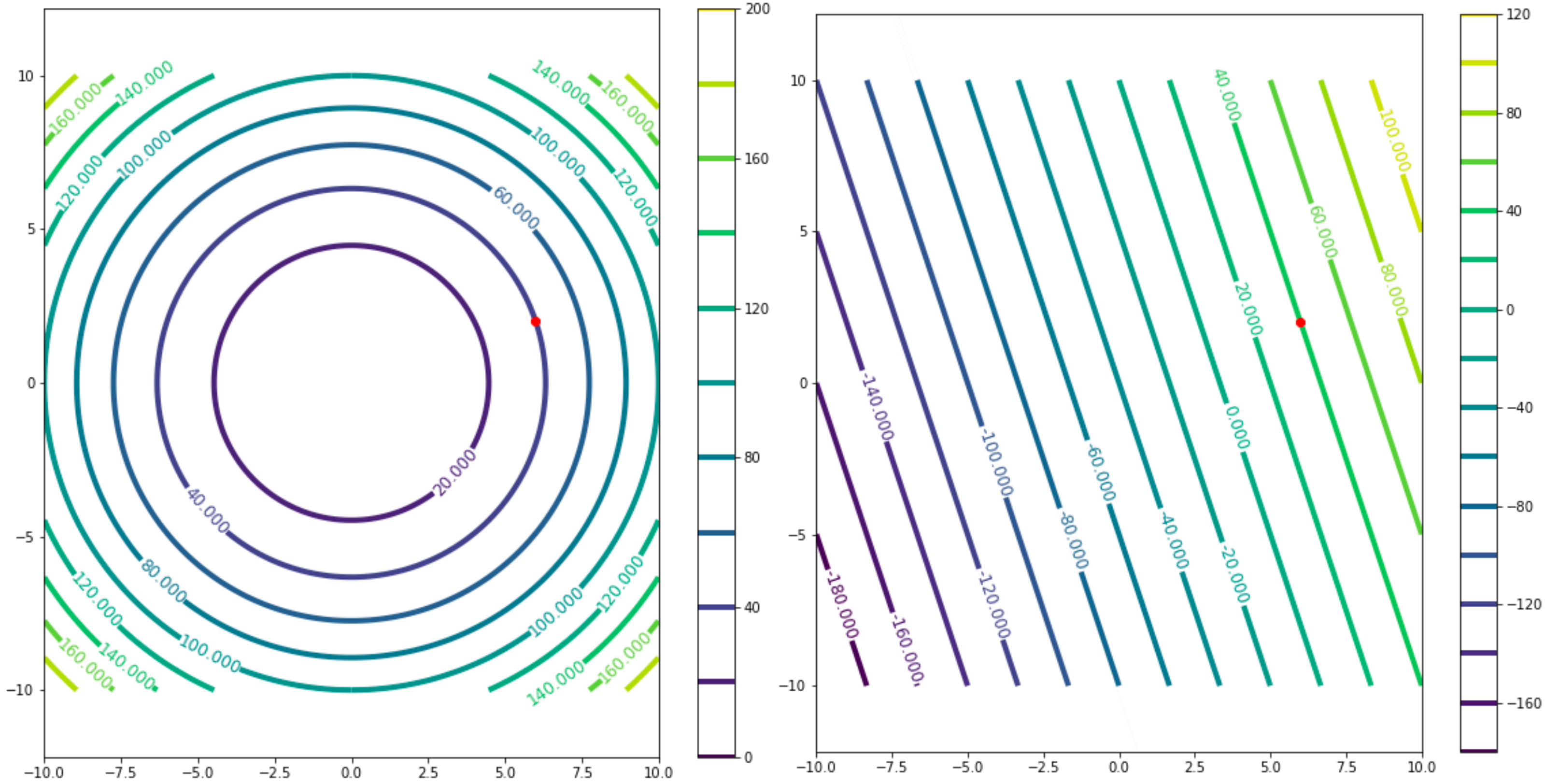
$$= 40 + 12(x_1 - 6) + 4(x_2 - 2)$$

$$= 40 + 12x_1 + 4x_2 - 72 - 8$$

$$= 12x_1 + 4x_2 - 40$$

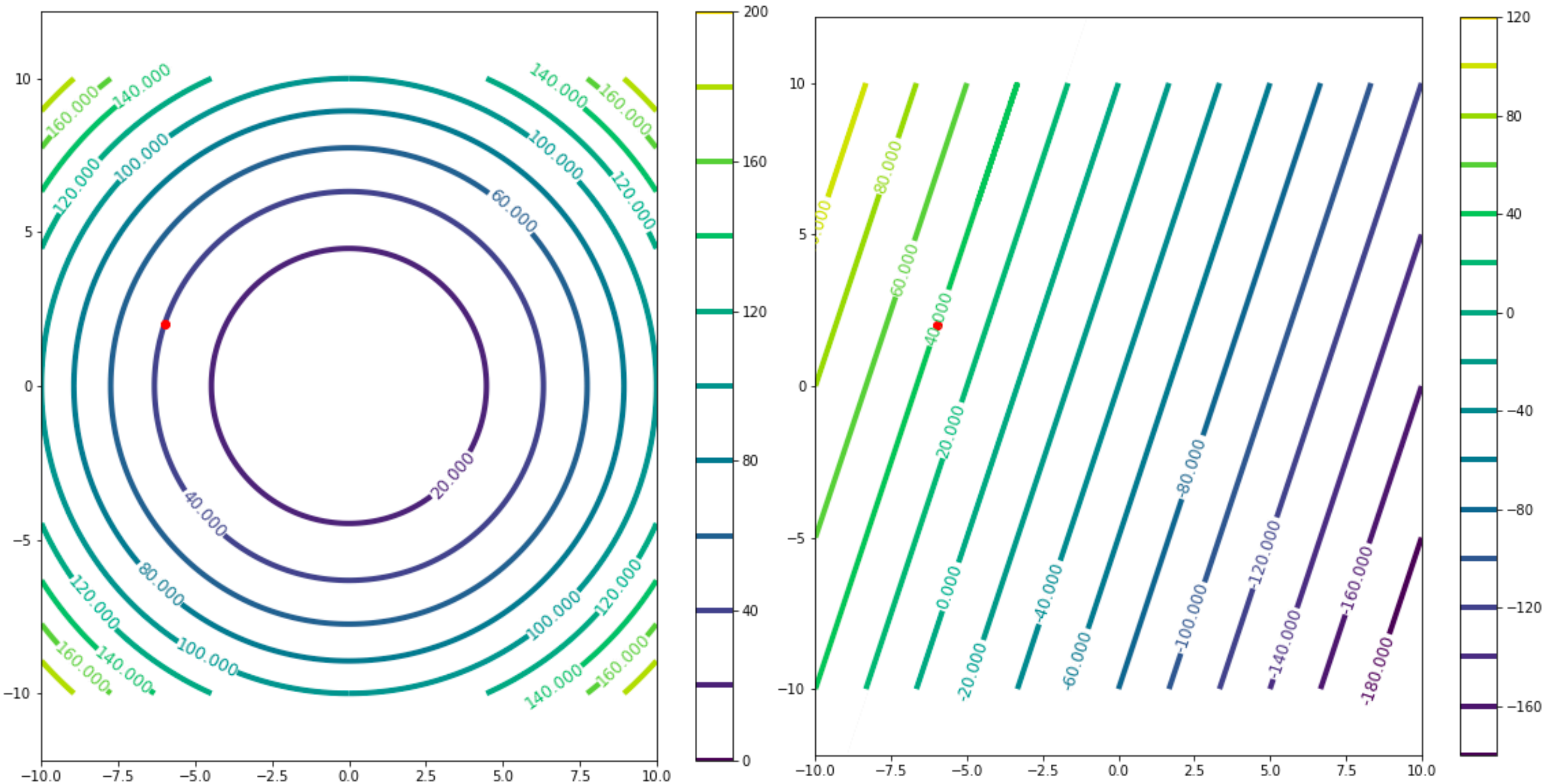
$$(x_1, x_2) \approx (6, 2)$$

Gradients and Linear Approximations



$f(x) = x_1^2 + x_2^2$ $L_v[f]$

Gradients and Linear Approximations

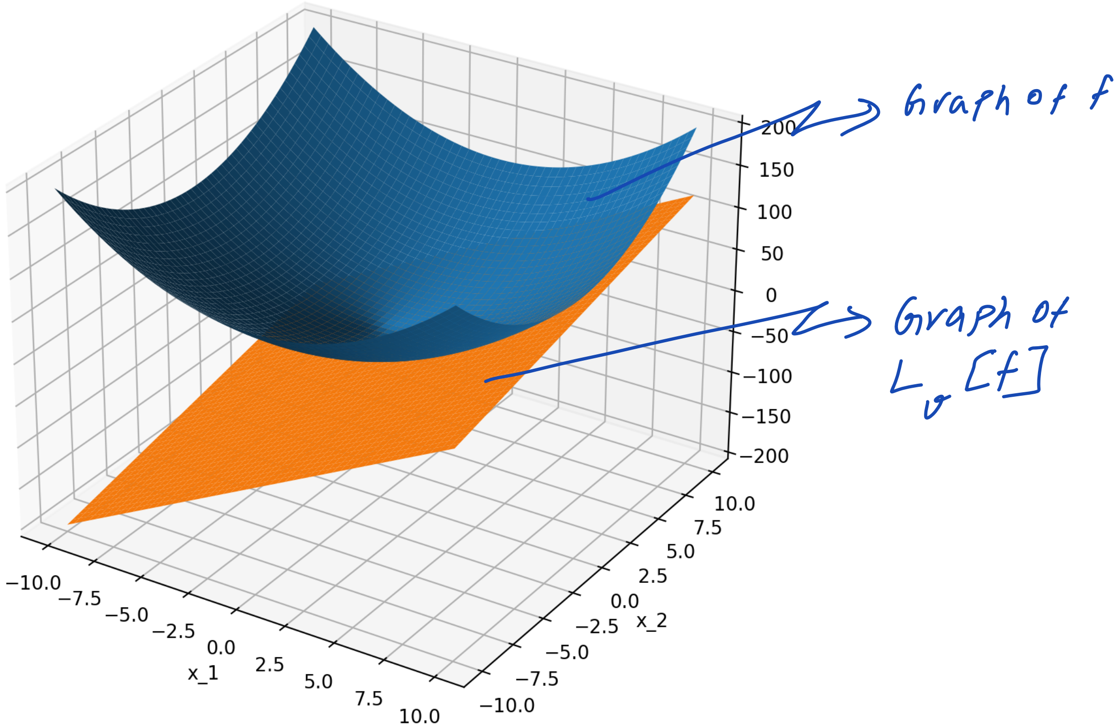


$f(x) \approx L_v[f](x) \quad (x \approx v)$

Gradients and Tangent Planes

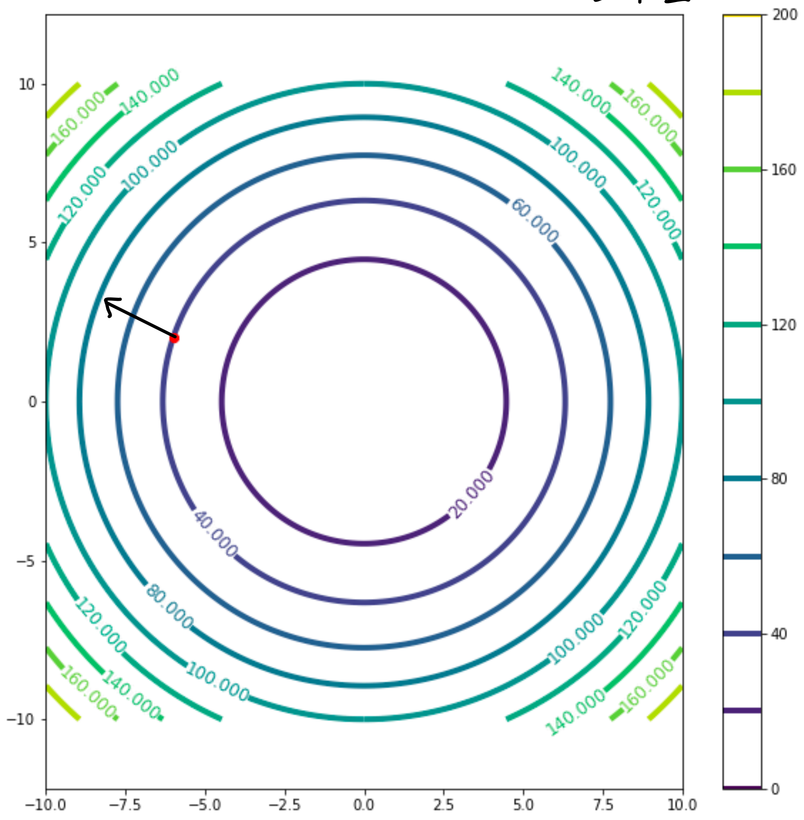
$f(x) = x_1^2 + x_2^2$

The graph of $L_v[f]$ is a plane that is tangent to the graph of f at the point $(v, f(v))$



Gradients and Contours

$v = \begin{bmatrix} -6 \\ 2 \end{bmatrix} \quad \nabla f(v) = \begin{bmatrix} -12 \\ 4 \end{bmatrix}$



$\nabla f(v) \perp \{x \in \mathbb{R}^d : f(x) = f(v)\}$

$\nabla f(v) \perp \{x \in \mathbb{R}^d : L_v[f](x) = f(v)\}$

$\parallel$

$\{x \in \mathbb{R}^d : f(v) + \nabla f(v)^T (x - v) = f(v)\}$

$\parallel$

$\{x \in \mathbb{R}^d : \nabla f(v)^T x = \nabla f(v)^T v\}$

$\{x \in \mathbb{R}^d : w^T x = b\}$

Directional Derivative

$$D_u[f](v) = \lim_{\alpha \to 0} \frac{f(v+\alpha u) - f(v)}{\alpha}$$

Directional derivative of f at the point v, along u.

$$= \lim_{\alpha \to 0} \frac{f(v) + \nabla f(v)^T \alpha u - f(v)}{\alpha}$$

$$= \nabla f(v)^T u$$

Cauchy-Schwarz Inequality

$a_1, a_2 \cdots a_d$

$b_1, b_2 \cdots b_d$

$\|a\| = \sqrt{a_1^2 + \cdots + a_d^2}$

$-\|a\| \cdot \|b\| \leq a^T b \leq \|a\| \|b\|$

$a = \alpha b$

$\alpha < 0$

$a = \alpha b$

$\alpha > 0$

Direction of Steepest Ascent

f

Find a direction u, that maximises the rate of change of f as you move from v along u.

Maximise $D_u[f](v)$

Find $u \in \mathbb{R}^d$, $\|u\| = 1$ and which maximises

$$D_u[f](v)$$

$$= \nabla f(v)^T u$$

$$u = \alpha \cdot \nabla f(v)$$

Descent Directions

$f: \mathbb{R}^d \to \mathbb{R}$

$v \in \mathbb{R}^d.$

What are the valid directions, such that f decreases

For what values of u : $D_u[f](v) < 0$

$$\Updownarrow$$

$\nabla f(v)^T u < 0$

Descent directions : $\{u \in \mathbb{R}^d : \nabla f(v)^T u < 0\}$

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - Linear approximations and Alternate gradient interpretations
  - **Applications/Advanced rules**

Higher Order Approximations

$$f : \mathbb{R}^d \to \mathbb{R}$$

$$f(x) \approx f(v) + \nabla f(v)^T (x - v)$$

(Valid around $x = v$)

$$f(x) \approx f(v) + \nabla f(v)^T (x - v) + \frac{1}{2} (x - v)^T \underbrace{\nabla^2 f(v)}_{\substack{d \times d \text{ matrix} \\ \text{Hessian}}} (x - v)$$

Higher Order Approximations

Maxima, minima and saddle points

If $f(x)$ is minimised at v

$$\Downarrow$$

$$\nabla f(v) = 0$$

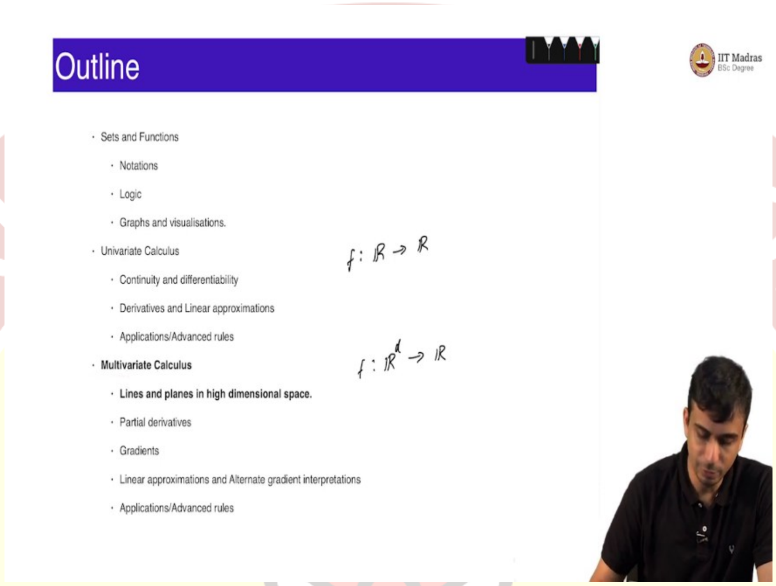
$\{v : \nabla f(v) = 0\} \rightsquigarrow$ Critical point

IIT Madras

ONLINE DEGREE

Machine Learning Foundations
Professor Harish Guruprasad Ramaswamy
Department of computer Sciences & Engineering
Indian Institute of Technology, Madras
Multivariate Calculus: Lines and Planes in Higher Dimensional Space

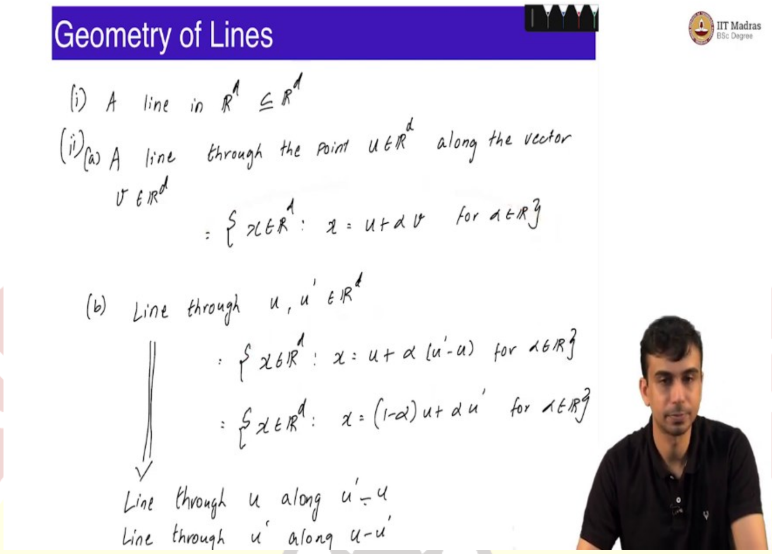
(Refer Slide Time: 00:16)



Hello everyone, and welcome to another lecture of Machine Learning Foundations. We will continue with our calculus recap. We have been seeing about univariate function, univariate calculus, which is about functions f from R to R, the domain and codomain are both R. We will generalize most of these two multivariate functions which are functions from Rd to R, the codomain is still R, the real value, just that the input domain is now d dimensional.

And this brings us lots of intuitive geometric ideas for it, for explaining multivariate calculus ideas. Still, we will be doing only mostly differential calculus and most of these you should have seen in your Mathematics for Data Science 2.

(Refer Slide Time: 01:17)



Before we actually get × the calculus, we should first get the basic geometry of high dimensional space. And the basic sets for doing that are lines and planes. So, what is a line, we will be working with the Rd now, let us, line in Rd is first of all a subset of Rd, line in Rd is a subset of Rd. A line is parameterized by two vectors, a line through the point U in Rd along the direction v in Rd is simply given by the set of all d dimensional vectors such that x can be written as u + α × v for α reals.

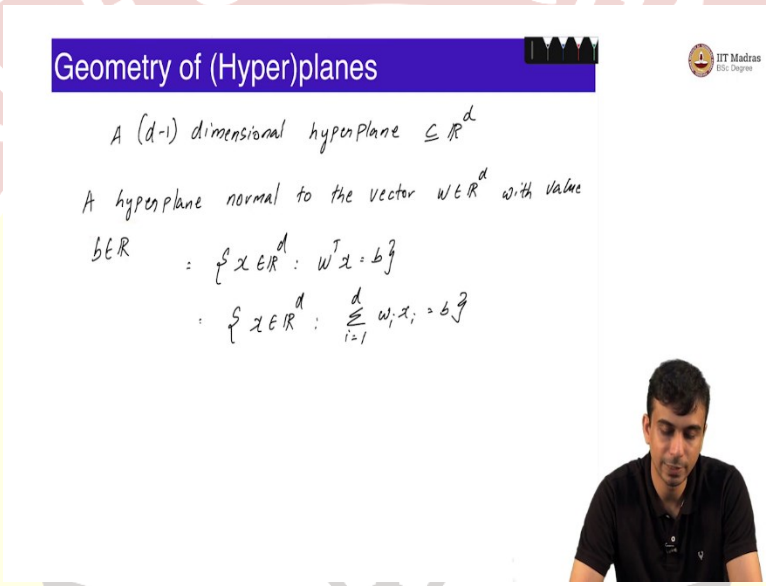
We will see some examples soon. But the idea is pretty clear it is a line through the point along the direction is given by this set. Some× you might also see another equivalent version that is a line through two points, it is u and u dash. So, that is another definition this, we will call this definition 2 a and 2 b is a line through the point, through the points u and u dash is simply given by the set of all x in Rd such that x can be written as u + α × u dash - u for α in real. Which is also written as set of all x in Rd such that x = 1 - α× u + α × u dash for α reals.

So, as you vary α you get different 1 - α u + α u dash and the union of every such element forms the line. You can see from this expression of the line through u dash and line through u along the vector v that this line through u and u dash is exactly the same as the line through u along the direction u dash - u. Similarly, line through all of these things are clearly the same.

So, this is exactly equal to a line through u along u dash - u. This is also equal to line through u dash along u - u dash. All of these three sets are the same really, line through u and u dash is the same as line through u along u dash - u, and it is the same as line through u dash along u - u dash, all of these are the same subset of R. You can, this is a simple exercise to prove.

In fact, it is kind of obvious to see from this notation that u + α × u dash - u can just be u dash - u as a vector indicating direction, if you do that, it is exactly the same as this. So that is a line is a subset of Rd. And it does look like line, we will see some examples soon.

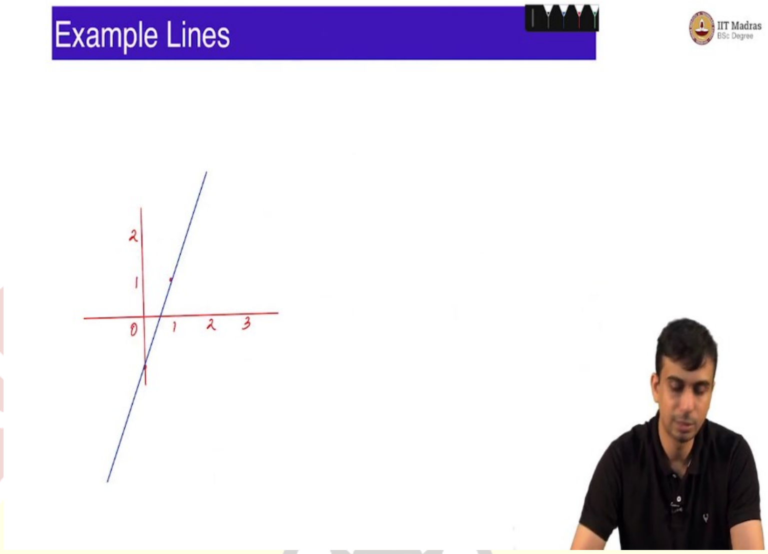
(Refer Slide Time: 05:25)



The next basic set in d dimensional space is a hyperplane or just a plane. A plain, hyperplane is typically a d - 1 dimensional hyperplane. This the default dimension of the hyperplane, this is also a subset of Rd. We will say that a hyperplane normal to the vector W in Rd with value b in R is given by the set of all x in Rd such that W transpose X = b which will set of x in Rd such that the sigma i equal to 1 to d Wi xi = b. So, this the definition of a hyperplane. There are other definitions also. But those are slightly more complicated.

So, this is the standard definition of hyperplane, which is defined in terms of a vector normal to the plane and value b, we will explain what these terms mean soon. But as of now, this is parametrized by a vector w and a value b.

(Refer Slide Time: 07:13)

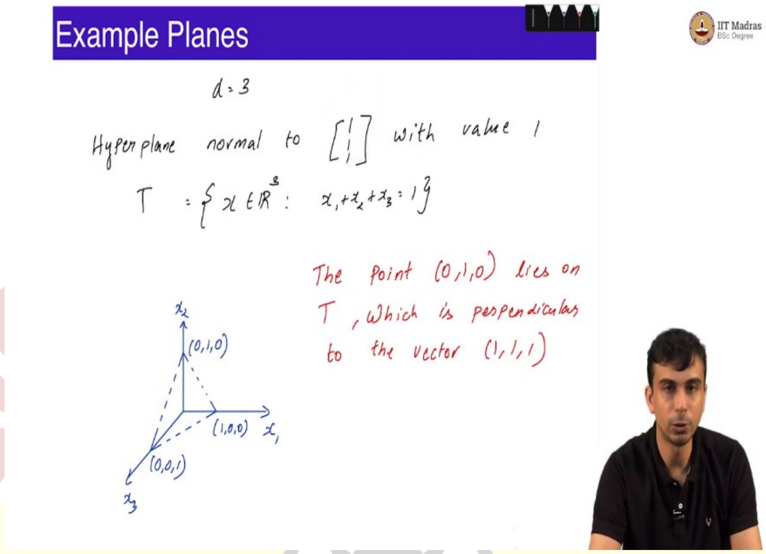


So, here are a couple of simple examples. For lines first. So, an example line here, this is an example line, a line through consider this set, this a line through the point 1, , 1 along the direction, 1, , 2. So, this would be the set of all x in R2 such that x = 1, 1 + α × 1, 2, where α varies over the real line.

So, visualization would be something like this, it does pass through the point 1, , 1 and it has all scalar multiples of 1, , 1 + α × 1, , 2, all scalar multiples of 1, , 2 can be added to the point 1, , 1. And that would correspond to this blue set. And you can see why it is called a line because it does really look like a line when plotted. So that is a line.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS • सिद्धिर्भवति कर्मजा •

(Refer Slide Time: 08:28)



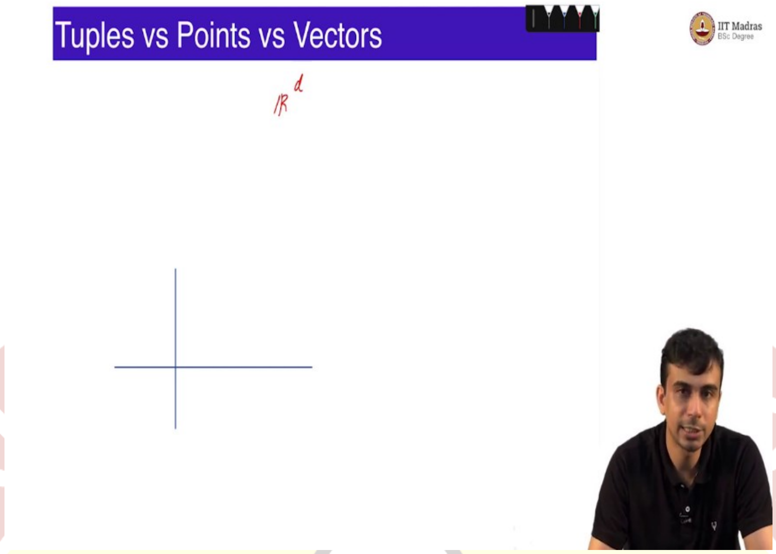
Now we will take a look at some example planes. Here is a simple example plane with three dimensions, d = three, a hyperplane. That is normal to the vector 1, 1, 1, normal, perpendicular, orthogonal are pretty much the same. Normal means perpendicular. Normal with value 1, what is that, it is simply the set of all x in R3 such that x1 + x2 + x3 = 1.

So, visualization of that would, in imagine a 3d, 3 axes given by x1, x2, x3. Here, x1 + x2 + x3 = 1 would correspond to a plane, which a part of it would be this, it could pass through the point 0, 1, 0 1, 0, 0, 0, 0, 1 it will go beyond the, beyond the axis as well. But the visible part, if you can imagine a 3d corner of your room and say imagine a corner of your room a paper, a piece of paper floating there would look like this, corresponding to the hyperplane, would look like this.

So here, we will make a distinction between points, vectors and tuples. So, this particular set, we will call it this the hyperplane T. You can say a statement like this. The point 0, , 1, , 0 lies on T, which is perpendicular or normal, perpendicular to the vector 1, , 1, , 1.

So why is this particular plane called normal or perpendicular to the vector 1, , 1, , 1. Well, what you could do is you could imagine a vector or a ray originating at the origin 0, , 0, , 0 and along 1, , 1, , 1. So that is kind of a ray, that would cut this, that would pierce this plane at 90 degrees. That is the reason why this particular plane is said to be normal, or perpendicular to the vector 1, , 1, , 1.

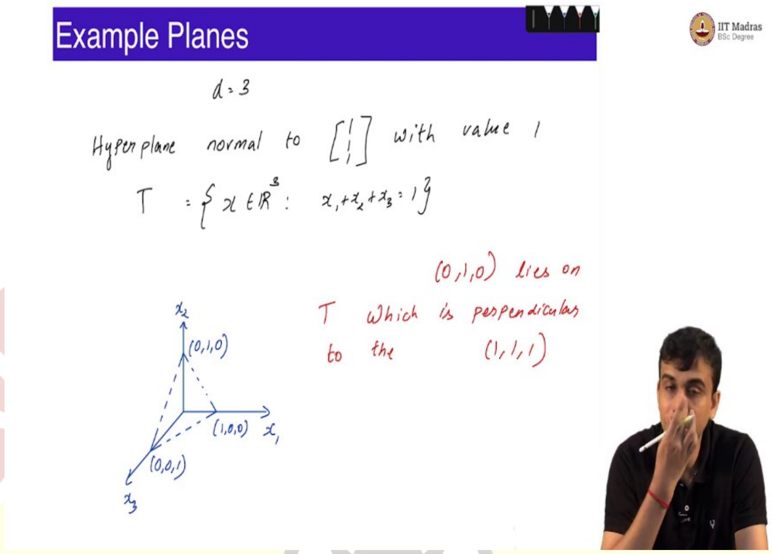
(Refer Slide Time: 11:27)



So, this makes the definition I mean, I have been using vectors, points and tuples interchangeably. Because all of these are represented by Rd. Rd is, a tuple is simply what you would store in, let us say, you are storing a d dimensional vector in a programming language, you would store it in a, as a tuple. So that is, that is how represented, so points and vectors are both represented as tuples.

But they have different meaning though, a point corresponds to, if you take physics, a point corresponds to a location and a vector corresponds to a direction. You can say that I am in Chennai, and I am going, I am going towards Bangalore. So that is a valid, so you can imagine a line through Chennai along the direction, given by Bangalore - Chennai. So that is, the that that is an example line. And directions are essentially vectors and, points are locations. So that is the, there is a difference both of these are represented by a vector in, by tuple in Rd, but there is a difference between point and vectors.

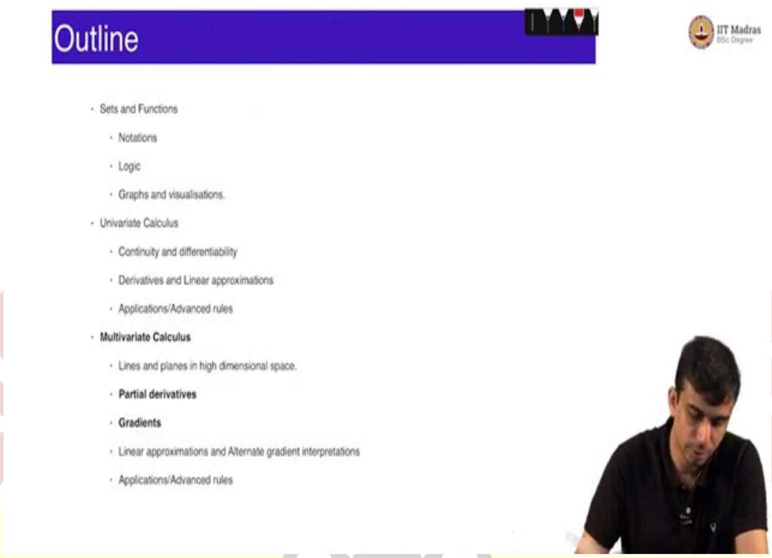
(Refer Slide Time: 12:50)



From context you should be clear whether a tuple is used as a point or as a vector. For example, take this point. So, some× you might not even see, you might say 0, , 1, , 0 lies on T which is perpendicular to 1, , 1, , 1. You can say this and you would not be wrong, even though I have not said the word point here. And the vector here, this would be a perfectly valid statement, because from context, it is clear that 0, , 1, , 0 is used as a point and 1, , 1, , 1 is used as a vector.

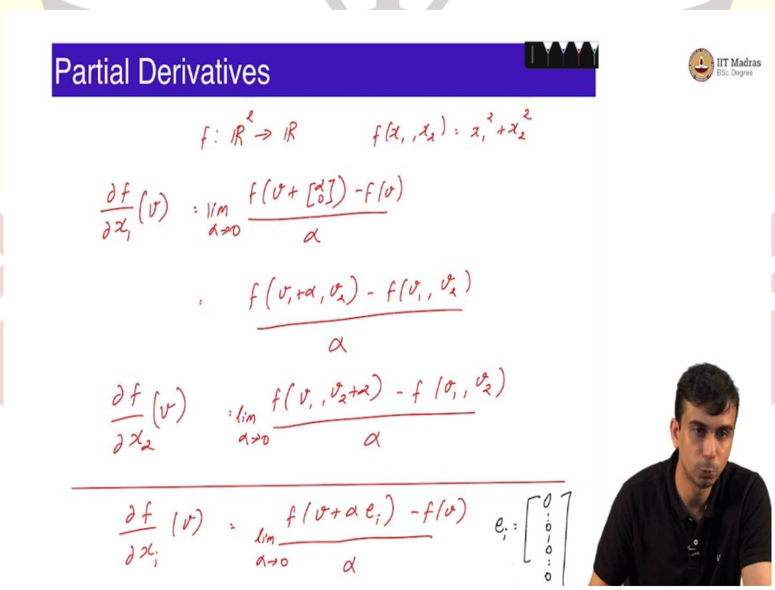
Why is that because 1, , 1, , 1 as a point makes no sense, because you cannot be perpendicular to a point you can be perpendicular to a vector, but you cannot be perpendicular to a point. And similarly, a vector cannot lie on a plane, only points can lie on a plane. But we will be using all of this interchangeably, from context it will be clear whether a particular element in Rd is used as a point or as a vector. Algebraically, it is, algebraically it should to be very clear. But geometrically, you must understand whether a particular vector is, a particular tuple is used as a vector or as a part.

(Refer Slide Time: 14:01)



So, with that, I think that gives the basic geometry tools that are required for understanding multivariate functions. And now we can move on to the basic tool that is, the basic building block of multivariate calculus, which is, which are partial derivatives and gradients.

(Refer Slide Time: 14:22)



So, what is a partial derivative. You have seen this perhaps in your school already. Let us say you can imagine a function f from R2 to R. The simplest multivariate function you can think of which is a two dimensional function. Let us say f of x1, , x2 = x1 square + x2 square, it is an example function. The partial derivative of a function with respect to let us say one of its arguments, what it means is you view the rest of the arguments as constants, if that is the case, this is still a one dimensional function.

If you fix the other variables, it becomes a one dimensional function, and then you can do the derivative. So that is what the partial derivative means, ∂ f by ∂ x1. So, I am going to use the sound of ∂ to denote this, I might see it called partial or something like that, but I am going to use ∂ because I find it much more convenient than partial, do that. So, ∂ f by ∂ x1 is simply F of ∂ F by ∂ x1 evaluated at a point v.

So, that is simply equal to f of v + I am just going to take d equal to 2 in this case, where we + α 0 - f of v by α. Why is this correct? This is correct because I am fixing the second variable, this is with respect to the first variable x1, so second variable is fixed. So, F of v + α or in other words I will write this even simpler, this is F of v1 + α, , v2 - f of v1, , v2 divided by α.

So, for the purposes of calculation of the partial derivative with respect to the first variable, the second variable is kept constant. So, it is viewed as a constant it is not changing at all, only the first variable is changing. F of v1 + α, and F of v1 + α - F of v1 by α, the second variable is kept constant at v2. So, this is the partial derivative Similarly, you can do the partial derivative of this with respect to the second variable also, what is ∂ f by ∂ x2 evaluated at v. What is that, that is f of v1, , v2 + α - f of v1, , v2 divided by α.

In general, if you have a d dimensional function, how would you do this. If you have a d dimensional function for a general d dimensional function you can still write it this way, that is ∂ f by let us say ∂ xi the ith argument. Let us say f has d arguments x1 to xd ∂ f by ∂ xi evaluated v what would that be, that would be f of v + α × ei - f of v divided by α.

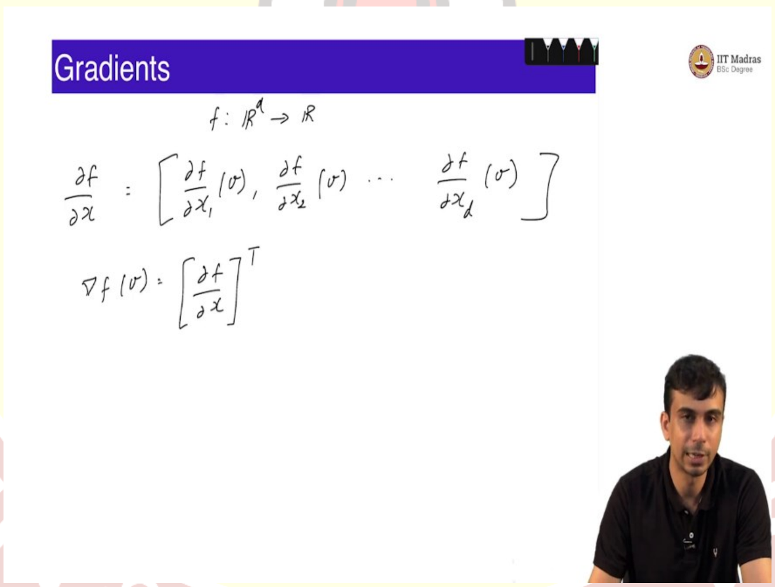
All of this is with limit α approaching 0. What is ei, ei is simply the coordinate vector with the ith coordinate being equal to 1 and the rest being 0. You can easily see that is what is happening here for, to get the first derivative with respect to the first variable you just take v + α × 1, 0, to get the derivative with respect to the second variable you take v + α × 0, 1.

In general, if you have d variables, if you take you want the partial derivative with respect to the ith variable, (())(18:41) × you take f of v + α × ei, where ei is the coordinate vector which has 0, where ei is a d dimensional vector, it is 0 everywhere else and 1 in the ith position, with there is just a single 1 which is in the ith position.

This is clearly generalizing the case of two variables that we have here. So, all of this, if you had, I mean for the, let us go back to the two variable case and ∂ f by ∂ x1 v. If the second variable is fixed at v2, f is just a function of the first variable, around v1. And in that case, this exactly reduces to the standard derivative expression.

In general, however, this is the general expression for the partial derivative or the, the ith argument you just consider F of v + α × ei - f of v by α, where you let the limit of α approach 0. So, this is for every, we have a d dimensional function, you can evaluate the partial derivative at, you can evaluate the partial derivative of a function at a saddle point, at a point v with respect to the several variables x1 to xd.

(Refer Slide Time: 20:20)

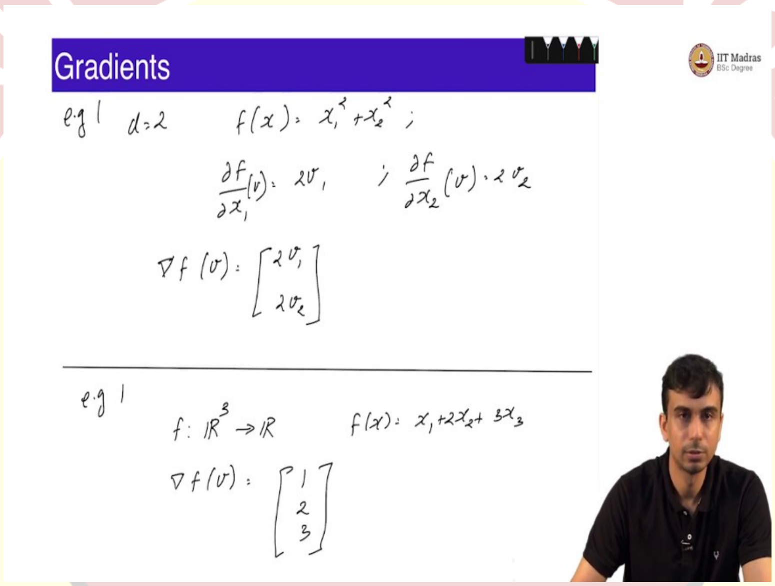


So, now, that means that you have a function f from Rd to R. You have for a given point, for any given point v, you have these things, which is ∂ f by ∂ x1 evaluated at v, ∂ f by ∂ x2 evaluated at v, and so on. ∂ f by ∂ xd evaluated at v. You can imagine putting all of these in a row vector and this you can just call it us ∂ f by ∂ x.

So, this f is a function, this x is a vector. So that we can take a derivative of a function with respect to a vector. By that we mean this particular vector, where you take you package in all your partial derivatives × one row vector. Often, you would also use a gradient, which is a gradient of f at v is simply ∂ f by ∂ x transpose.

Why is that, because the gradient is typically written as a column vector. But derivative of f with respect to x, where x is a d dimensional vector is written as a row vector. Otherwise, it is the same thing, it is the gradient is also a collection of partial derivatives, it is the d dimensional vector which collects partial derivatives.

(Refer Slide Time: 21:44)

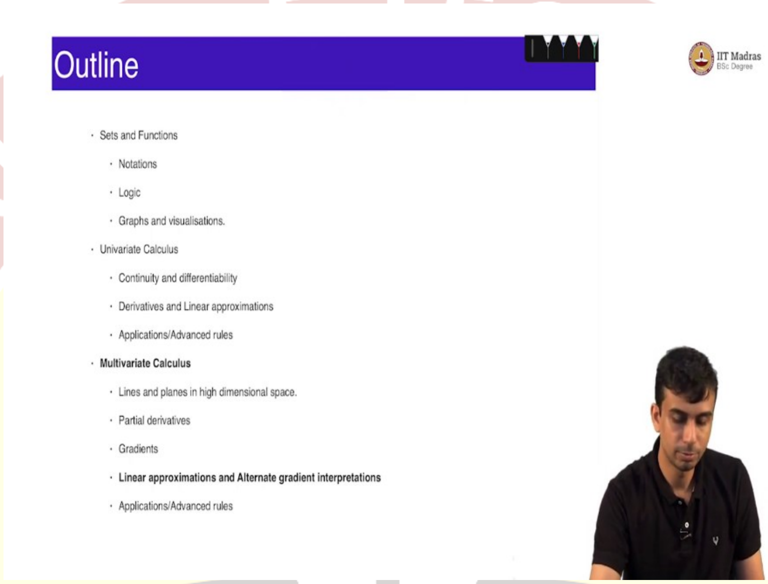


So, we can have some simple functions and take the derivatives. Let us use the same function which is f of x = x1 square + x2 square. In this case, let us say d = 2, it is a 2 dimensional function, f of x equal x1 square + x2 square. What is ∂ f by ∂ x1, well, that is simply equal to 2 x1 ∂ f by ∂ x2 is simply equal to 2x 2. ∂ f by ∂ x1 evaluated at v is simply 2 v1, 2 v2, so ∂ f, I think it is better to write ∂ f by x1 evaluated at v = 2 v1, ∂ f by ∂ x2 evaluated at v is equal 2 v2.

So, the gradient of f evaluated v is a column vector given by 2 v1 2 v2, you can have other functions, we will have. This is the first example. Another example might be, let us say you have a function f from R3 to R as follows. F of x = let us say x1 + 2x 2 + 3x 3. What is the gradient of f with respect to, what is the gradient of f evaluated at some point v? Well, that is the vector given by ∂ f by ∂ x 1, ∂ f and ∂ f by ∂ x2 ∂ f by ∂ x3.

And if you take the, if you view x2 and x3 as constants, and you take the derivative of f with respect x1, you get simply 1. Similarly, if you take the derivative with respect to the second variable is x 2, you get 2, you get for the third variable you get 3. So, the gradient happens to be a constant. And these kind of functions where the gradient is a constant is called a linear function. f of x equals x 1 + 2x 2 + 3x 3 it is linear in all its arguments. And that is the reason why the gradient is a constant.

(Refer Slide Time: 24:07)



So, with that, I think we have seen the basic tool, the basic partial derivatives and gradients that are required for multivariate calculus. We will see some more applications of these gradients and partial derivatives soon. But we will wrap this video for now.

Outline

- Sets and Functions
    - Notations
    - Logic
    - Graphs and visualisations.
- Univariate Calculus
    - Continuity and differentiability
    - Derivatives and Linear approximations
    - Applications/Advanced rules
- **Multivariate Calculus**
    - **Lines and planes in high dimensional space.**
    - Partial derivatives
    - Gradients
    - Linear approximations and Alternate gradient interpretations
    - Applications/Advanced rules

$f : \mathbb{R} \to \mathbb{R}$

$f : \mathbb{R}^d \to \mathbb{R}$

Geometry of Lines

(i) A line in $\mathbb{R}^d \subseteq \mathbb{R}^d$

(ii) (a) A line through the point $u \in \mathbb{R}^d$ along the vector $v \in \mathbb{R}^d$

$$= \{ x \in \mathbb{R}^d : x = u + \alpha v \text{ for } \alpha \in \mathbb{R} \}$$

(b) Line through $u, u' \in \mathbb{R}^d$

$$= \{ x \in \mathbb{R}^d : x = u + \alpha (u' - u) \text{ for } \alpha \in \mathbb{R} \}$$

$$= \{ x \in \mathbb{R}^d : x = (1-\alpha) u + \alpha u' \text{ for } \alpha \in \mathbb{R} \}$$

$\Downarrow$

Line through u along $u' - u$

Line through u' along $u - u'$

Geometry of (Hyper)planes

A $(d-1)$ dimensional hyperplane $\subseteq \mathbb{R}^d$

A hyperplane normal to the vector $w \in \mathbb{R}^d$ with value $b \in \mathbb{R}$

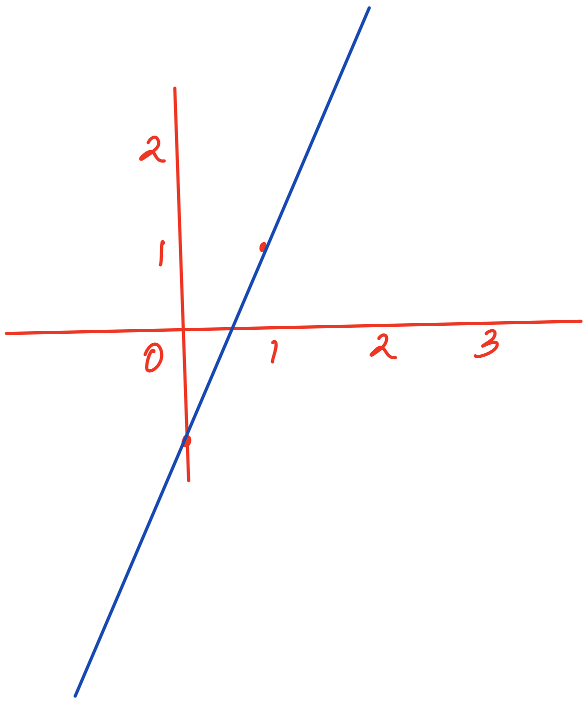
$$= \{x \in \mathbb{R}^d : w^T x = b\}$$

$$= \{x \in \mathbb{R}^d : \sum_{i=1}^{d} w_i x_i = b\}$$

Example Lines

Line through $\begin{pmatrix}1\\1\end{pmatrix}$ along $\begin{pmatrix}1\\2\end{pmatrix}$

$\left\{x \in \mathbb{R}^2 : x = \begin{bmatrix}1\\1\end{bmatrix} + \alpha \begin{bmatrix}1\\2\end{bmatrix} : \alpha \in \mathbb{R}\right\}$

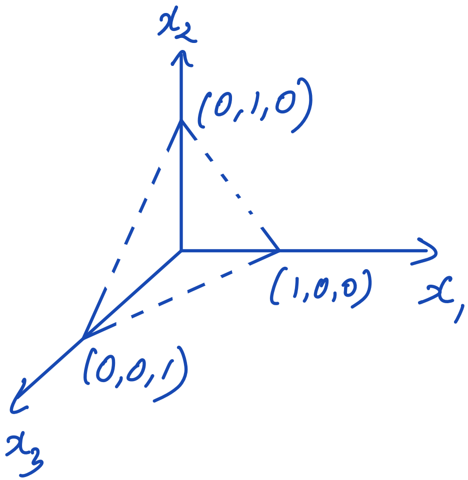


Example Planes

$d = 3$

Hyperplane normal to $\begin{bmatrix}1\\1\\1\end{bmatrix}$ with value 1

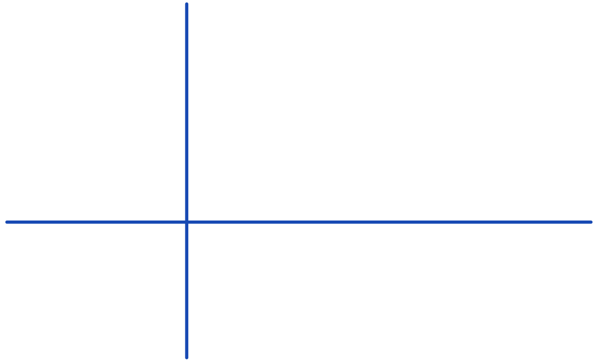
$T = \{ x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 1 \}$



$(0,1,0)$ lies on T which is perpendicular to the $(1,1,1)$

Tuples vs Points vs Vectors

$\mathbb{R}^d$



Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - **Partial derivatives**
  - **Gradients**
  - Linear approximations and Alternate gradient interpretations
  - Applications/Advanced rules

Partial Derivatives

$$f : \mathbb{R}^2 \to \mathbb{R} \qquad f(x_1, x_2) = x_1^2 + x_2^2$$

$$\frac{\partial f}{\partial x_1}(v) = \lim_{\alpha \to 0} \frac{f\left(v + \begin{bmatrix} \alpha \\ 0 \end{bmatrix}\right) - f(v)}{\alpha}$$

$$= \lim_{\alpha \to 0} \frac{f(v_1 + \alpha, v_2) - f(v_1, v_2)}{\alpha}$$

$$\frac{\partial f}{\partial x_2}(v) = \lim_{\alpha \to 0} \frac{f(v_1, v_2 + \alpha) - f(v_1, v_2)}{\alpha}$$

$$\frac{\partial f}{\partial x_i}(v) = \lim_{\alpha \to 0} \frac{f(v + \alpha e_i) - f(v)}{\alpha} \qquad e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Gradients

$$f : \mathbb{R}^d \to \mathbb{R}$$

$$\frac{\partial f}{\partial x}(v) = \left[\frac{\partial f}{\partial x_1}(v), \ \frac{\partial f}{\partial x_2}(v) \ \cdots \ \frac{\partial f}{\partial x_d}(v) \right]$$

$$\nabla f(v) = \left[\frac{\partial f}{\partial x} \right]^T$$

Gradients

e.g 1 $\quad d=2 \qquad f(x) = x_1^2 + x_2^2$;

$$\frac{\partial f}{\partial x_1}(v) = 2v_1 \quad ; \quad \frac{\partial f}{\partial x_2}(v) = 2v_2$$

$$\nabla f(v) = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix}$$

---

e.g 1

$f : \mathbb{R}^3 \to \mathbb{R} \qquad f(x) = x_1 + 2x_2 + 3x_3$

$$\nabla f(v) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Outline

- Sets and Functions
  - Notations
  - Logic
  - Graphs and visualisations.
- Univariate Calculus
  - Continuity and differentiability
  - Derivatives and Linear approximations
  - Applications/Advanced rules
- **Multivariate Calculus**
  - Lines and planes in high dimensional space.
  - Partial derivatives
  - Gradients
  - **Linear approximations and Alternate gradient interpretations**
  - Applications/Advanced rules

Machine Learning - Foundations

Tutorial - Week 2

IIT Madras Online Degree

Outline

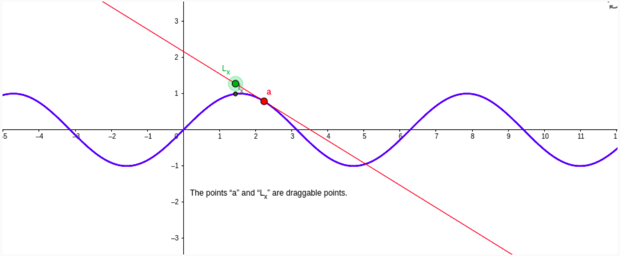
1. Linear approximation
2. Higher order approximations
3. Multivariate linear approximation
4. Directional derivatives

Linear approximation

Linear approximation (Linearization)

Def:

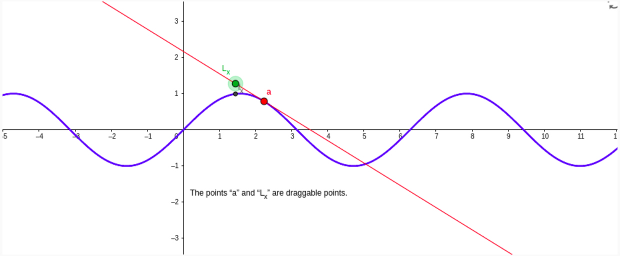
Approximation of any function using a linear function .



Linear approximation (Linearization)

Def:

Approximation of any function using a linear function .



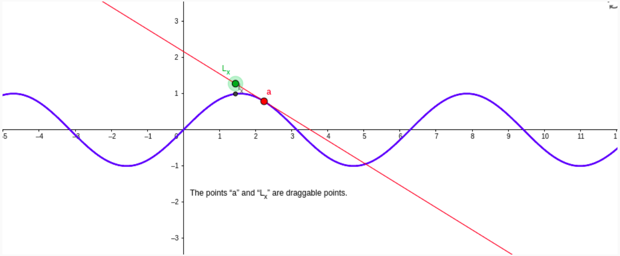
Need:

- Linear functions are easier to work with.

Linear approximation (Linearization)

Def:

Approximation of any function using a linear function .



Need:

- Linear functions are easier to work with.
- Finding approximate values of functions at certain points when exact values are not known.

The equation

The linear approximation $L(x)$ of a function $f(x)$ at point a is given by:

$$\boxed{L(x) = f(a) + f^{'}(a)(x - a)}$$

The equation

The linear approximation $L(x)$ of a function $f(x)$ at point a is given by:

$$\boxed{L(x) = f(a) + f^{'}(a)(x - a)}$$

This is indeed the equation of a tangent line:

$$y - y_1 = m(x - x_1)$$
$$y = y_1 + m(x - x_1)$$

The equation

The linear approximation $L(x)$ of a function $f(x)$ at point a is given by:

$$\boxed{L(x) = f(a) + f^{'}(a)(x - a)}$$

This is indeed the equation of a tangent line:

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y &= y_1 + m(x - x_1) \end{aligned}$$

If $x_1 = a$, $y_1 = f(a)$ and $m = f'(a)$, we get,

$$y = f(a) + f'(a)(x - a)$$

Problem 1

Compute the approximate value of $\sqrt{50}$.

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

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

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$\begin{aligned} f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\ f(49) &= \sqrt{49} = 7 \end{aligned}$$

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14}
\end{aligned}
$$

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

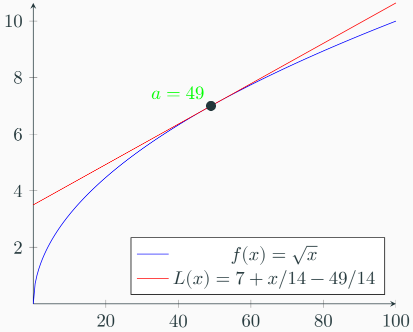
$$
\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\
L(x) &= f(49) + f'(49)(x - 49)
\end{aligned}
$$

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\
L(x) &= f(49) + f'(49)(x-49) \\
L(x) &= 7 + \frac{1}{14}(x-49)
\end{aligned}
$$

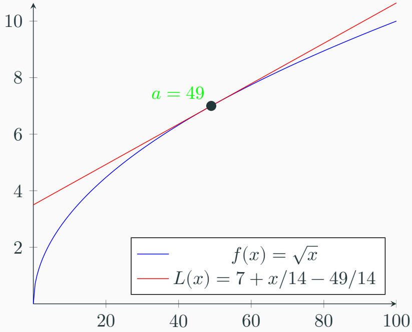


Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\
L(x) &= f(49) + f'(49)(x - 49) \\
L(x) &= 7 + \frac{1}{14}(x - 49)
\end{aligned}
$$



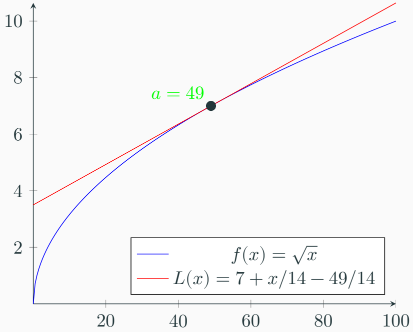
Approximate value of $\sqrt{50} = L(50) = 7 + \frac{1}{14}(50 - 49) = 7 + \frac{1}{14} =$ **7.071**

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\
L(x) &= f(49) + f'(49)(x-49) \\
L(x) &= 7 + \frac{1}{14}(x-49)
\end{aligned}$$



Approximate value of $\sqrt{50} = L(50) = 7 + \frac{1}{14}(50-49) = 7 + \frac{1}{14} =$ **7.071**

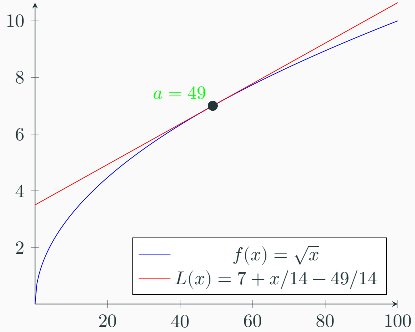
- **Note 1:** Actual value of $\sqrt{50}$ (up to 3 decimal places) is 7.071.

Problem 1

Compute the approximate value of $\sqrt{50}$.

The closest known value to $\sqrt{50}$ is $\sqrt{49}$, so we set $f(x) = \sqrt{x}$ and $a = 49$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}x^{-\frac{1}{2}} \\
f(49) &= \sqrt{49} = 7 \\
f'(49) &= \frac{1}{(2)(\sqrt{49})} = \frac{1}{14} \\
L(x) &= f(49) + f'(49)(x - 49) \\
L(x) &= 7 + \frac{1}{14}(x - 49)
\end{aligned}
$$



Approximate value of $\sqrt{50} = L(50) = 7 + \frac{1}{14}(50 - 49) = 7 + \frac{1}{14} =$ **7.071**

- **Note 1:** Actual value of $\sqrt{50}$ (up to 3 decimal places) is 7.071.
- **Note 2:** $L(100)$ gives 10.64 while the actual value of $\sqrt{100}$ is 10.

Problem 2

Compute the approximate value of $e^{0.017}$.

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$f'(x) = e^x$$

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \end{aligned}$$

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \\ f'(0) &= 1 \end{aligned}$$

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

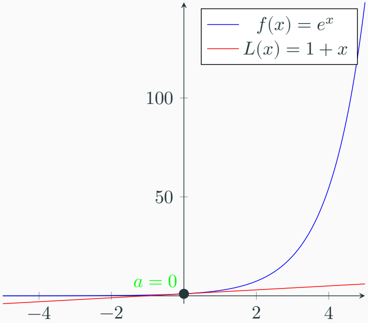
$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \\ f'(0) &= 1 \\ L(x) &= f(0) + f'(0)(x - 0) \end{aligned}$$

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

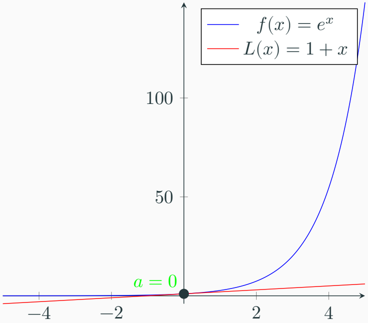
$$
\begin{aligned}
f'(x) &= e^x \\
f(0) &= 1 \\
f'(0) &= 1 \\
L(x) &= f(0) + f'(0)(x - 0) \\
L(x) &= 1 + 1(x) = 1 + x
\end{aligned}
$$



Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \\ f'(0) &= 1 \\ L(x) &= f(0) + f'(0)(x-0) \\ L(x) &= 1 + 1(x) = 1 + x \end{aligned}$$



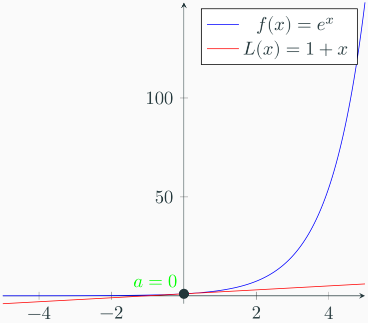
Approximate value of $e^{0.017} = L(0.017) = 1 + 0.017 = 1.017$

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$
\begin{aligned}
f'(x) &= e^x \\
f(0) &= 1 \\
f'(0) &= 1 \\
L(x) &= f(0) + f'(0)(x-0) \\
L(x) &= 1 + 1(x) = 1 + x
\end{aligned}
$$



Approximate value of $e^{0.017} = L(0.017) = 1 + 0.017 = 1.017$

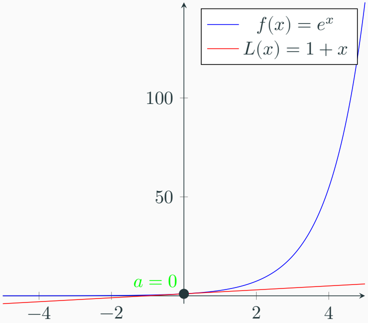
- **Note 1:** Actual value of $e^{0.017}$ is also 1.017.

Problem 2

Compute the approximate value of $e^{0.017}$.

The closest known value to $e^{0.017}$ is e^0, so we set $f(x) = e^x$ and $a = 0$.

$$\begin{aligned} f'(x) &= e^x \\ f(0) &= 1 \\ f'(0) &= 1 \\ L(x) &= f(0) + f'(0)(x-0) \\ L(x) &= 1 + 1(x) = 1 + x \end{aligned}$$



Approximate value of $e^{0.017} = L(0.017) = 1 + 0.017 = 1.017$

- **Note 1:** Actual value of $e^{0.017}$ is also 1.017.
- **Note 2:** $L(1)$ gives 2 while the actual value of e is 2.718.

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$f'(x) \quad = \quad \frac{1}{2}(x+4)^{-\frac{1}{2}}$$

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$\begin{aligned} f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f(5) &= \sqrt{5+4} = 3 \end{aligned}$$

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$
\begin{aligned}
f^{'}(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\
f(5) &= \sqrt{5+4} = 3 \\
f^{'}(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6}
\end{aligned}
$$

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

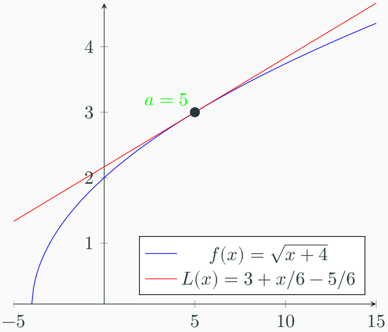
$$\begin{aligned} f^{'}(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f(5) &= \sqrt{5+4} = 3 \\ f^{'}(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\ L(x) &= f(5) + f^{'}(5)(x-5) \end{aligned}$$

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\
f(5) &= \sqrt{5+4} = 3 \\
f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\
L(x) &= f(5) + f'(5)(x-5) \\
L(x) &= 3 + \frac{1}{6}(x-5) = 3 + \frac{x-5}{6}
\end{aligned}
$$

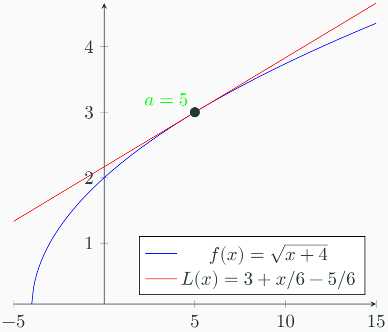


Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$\begin{aligned} f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f(5) &= \sqrt{5+4} = 3 \\ f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\ L(x) &= f(5) + f'(5)(x-5) \\ L(x) &= 3 + \frac{1}{6}(x-5) = 3 + \frac{x-5}{6} \end{aligned}$$



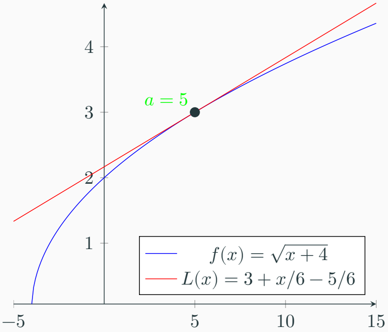
Approximate value of $f(6) = L(6) = 3 + \frac{6-5}{6} = 3 + \frac{1}{6} = \frac{19}{6} = 3.1666$

Problem 3

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\
f(5) &= \sqrt{5+4} = 3 \\
f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\
L(x) &= f(5) + f'(5)(x-5) \\
L(x) &= 3 + \frac{1}{6}(x-5) = 3 + \frac{x-5}{6}
\end{aligned}
$$



Approximate value of $f(6) = L(6) = 3 + \frac{6-5}{6} = 3 + \frac{1}{6} = \frac{19}{6} = 3.1666$

- **Note:** Actual value of $f(6) = \sqrt{10}$ is 3.1622. **Why?**

Higher order approximations

Higher order approximations

Linear Approximation

$$\boxed{L(x) = f(a) + f^{'}(a)(x - a)}$$

Higher order approximations

Linear Approximation

$$\boxed{L(x) = f(a) + f^{'}(a)(x-a)}$$

Quadratic Approximation

$$\boxed{L(x) = f(a) + f^{'}(a)(x-a) + \frac{f^{''}(a)}{2}(x-a)^2}$$

Higher order approximations

Linear Approximation

$$\boxed{L(x) = f(a) + f^{'}(a)(x-a)}$$

Quadratic Approximation

$$\boxed{L(x) = f(a) + f^{'}(a)(x-a) + \frac{f^{''}(a)}{2}(x-a)^2}$$

Higher-order Approximations

$$\boxed{\begin{gathered} L(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2}(x-a)^2 + \\ + \frac{f^{(3)}(a)}{3 \cdot 2}(x-a)^3 + \frac{f^{(4)}(a)}{4 \cdot 3 \cdot 2}(x-a)^4 \ldots \end{gathered}}$$

Problem 4

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

Problem 4

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

Problem 4

Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$\begin{aligned} f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\ f''(x) &= -\frac{1}{4}(x+4)^{-\frac{3}{2}} \\ f'''(x) &= \frac{3}{8}(x+4)^{-\frac{5}{2}} \end{aligned}$$

Problem 4

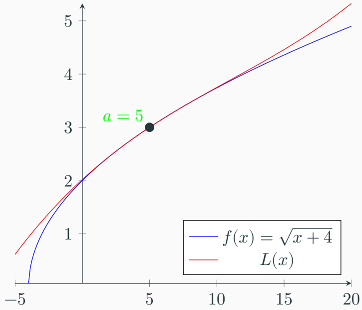
Let $f(x) = \sqrt{x+4}$, what is $f(6)$?

The closest known value to $\sqrt{x+4}$ is $\sqrt{9}$ with $x = 5$, so we set $f(x) = \sqrt{x+4}$ and $a = 5$.

$$
\begin{aligned}
f'(x) &= \frac{1}{2}(x+4)^{-\frac{1}{2}} \\
f''(x) &= -\frac{1}{4}(x+4)^{-\frac{3}{2}} \\
f'''(x) &= \frac{3}{8}(x+4)^{-\frac{5}{2}} \\
f(5) &= \sqrt{5+4} = 3 \\
f'(5) &= \frac{1}{(2)(\sqrt{9})} = \frac{1}{6} \\
f''(5) &= -\frac{1}{108} \\
f'''(5) &= \frac{1}{(24)(27)}
\end{aligned}
$$

$$
\begin{aligned}
L(x) &= f(5) + f^{'}(5)(x-5) + \frac{f^{''}(5)}{2}(x-5)^2 + \frac{f^{'''}(5)}{(3)(2)}(x-5)^3 + ... \\
L(x) &= 3 + \frac{1}{6}(x-5) - \frac{1}{(108)(2)}(x-5)^2 + \frac{1}{(24)(27)(6)}(x-5)^3
\end{aligned}
$$

$$
\begin{aligned}
L(x) &= f(5) + f^{'}(5)(x-5) + \frac{f^{''}(5)}{2}(x-5)^2 + \frac{f^{'''}(5)}{(3)(2)}(x-5)^3 + \dots \\
L(x) &= 3 + \frac{1}{6}(x-5) - \frac{1}{(108)(2)}(x-5)^2 + \frac{1}{(24)(27)(6)}(x-5)^3
\end{aligned}
$$

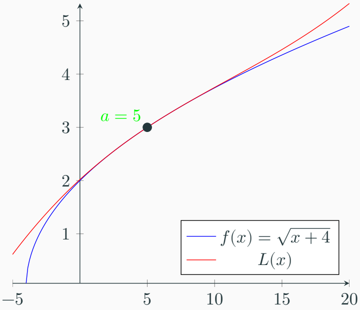


Approximate value of

$f(6) = L(6) = 3 + \frac{1}{6}(6-5) - \frac{1}{(108)(2)}(6-5)^2 + \frac{1}{(24)(27)(6)}(6-5)^3 = 3.1622$

$$L(x) = f(5) + f^{'}(5)(x-5) + \frac{f^{''}(5)}{2}(x-5)^2 + \frac{f^{'''}(5)}{(3)(2)}(x-5)^3 + \ldots$$

$$L(x) = 3 + \frac{1}{6}(x-5) - \frac{1}{(108)(2)}(x-5)^2 + \frac{1}{(24)(27)(6)}(x-5)^3$$



Approximate value of

$f(6) = L(6) = 3 + \frac{1}{6}(6-5) - \frac{1}{(108)(2)}(6-5)^2 + \frac{1}{(24)(27)(6)}(6-5)^3 = 3.1622$

Multivariate linear approximation

Linear approximation of functions involving multiple variables

The linear approximation of a function f of two variables x and y in the neighborhood of (a, b) is:

$$\boxed{L(x,y) = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b)}$$

Problem 5

Find the linearization of $f(x, y) = xe^{xy}$ at $(1, 0)$. Use it to approximate $f(1.1, -0.1)$.

Problem 5

Find the linearization of $f(x, y) = xe^{xy}$ at $(1, 0)$. Use it to approximate $f(1.1, -0.1)$.

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy} \\ \frac{\partial f}{\partial y}(x, y) &= xe^{xy}x = x^2e^{xy} \end{aligned}$$

Problem 5

Find the linearization of $f(x,y) = xe^{xy}$ at $(1,0)$. Use it to approximate $f(1.1,-0.1)$.

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy} \\
\frac{\partial f}{\partial y}(x,y) &= xe^{xy}x = x^2e^{xy}
\end{aligned}$$

Here $(a,b) = (1,0)$.

Problem 5

Find the linearization of $f(x,y) = xe^{xy}$ at $(1,0)$. Use it to approximate $f(1.1,-0.1)$.

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy}\\ \frac{\partial f}{\partial y}(x,y) &= xe^{xy}x = x^2e^{xy}\end{aligned}$$

Here $(a,b) = (1,0)$.

$$f(1,0) = e^0 = 1$$

Problem 5

Find the linearization of $f(x,y) = xe^{xy}$ at $(1,0)$. Use it to approximate $f(1.1,-0.1)$.

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy} \\ \frac{\partial f}{\partial y}(x,y) &= xe^{xy}x = x^2e^{xy} \end{aligned}$$

Here $(a,b) = (1,0)$.

$$\begin{aligned} f(1,0) &= e^0 = 1 \\ \frac{\partial f}{\partial x}(a,b) &= \frac{\partial f}{\partial x}(1,0) = e^0 = 1 \end{aligned}$$

Problem 5

Find the linearization of $f(x,y) = xe^{xy}$ at $(1,0)$. Use it to approximate $f(1.1,-0.1)$.

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= xe^{xy}y + e^{xy} = xye^{xy} + e^{xy}\\ \frac{\partial f}{\partial y}(x,y) &= xe^{xy}x = x^2e^{xy}\end{aligned}$$

Here $(a,b) = (1,0)$.

$$\begin{aligned}f(1,0) &= e^0 = 1\\ \frac{\partial f}{\partial x}(a,b) &= \frac{\partial f}{\partial x}(1,0) = e^0 = 1\\ \frac{\partial f}{\partial y}(a,b) &= \frac{\partial f}{\partial y}(1,0) = e^0 = 1\end{aligned}$$

$$L(x, y) \quad = \quad f(1,0) + \frac{\partial f}{\partial x}(1,0)(x-1) + \frac{\partial f}{\partial y}(1,0)(y-0)$$

$$
\begin{aligned}
L(x,y) &= f(1,0) + \frac{\partial f}{\partial x}(1,0)(x-1) + \frac{\partial f}{\partial y}(1,0)(y-0) \\
&= 1 + 1(x-1) + 1(y) \\
&= x + y
\end{aligned}
$$

$$
\begin{aligned}
L(x,y) &= f(1,0) + \frac{\partial f}{\partial x}(1,0)(x-1) + \frac{\partial f}{\partial y}(1,0)(y-0) \\
&= 1 + 1(x-1) + 1(y) \\
&= x + y \\
f(1.1,-0.1) &= L(1.1,-0.1) \\
&= 1.1 - 0.1 = 1
\end{aligned}
$$

$$
\begin{aligned}
L(x, y) &= f(1,0) + \frac{\partial f}{\partial x}(1,0)(x-1) + \frac{\partial f}{\partial y}(1,0)(y-0) \\
&= 1 + 1(x-1) + 1(y) \\
&= x + y \\
f(1.1, -0.1) &= L(1.1, -0.1) \\
&= 1.1 - 0.1 = 1
\end{aligned}
$$

The actual value of f(1.1, -0.1) = 1.1e$^{-0.11} = \frac{1.1}{1.11628} = 0.98542$

Directional derivatives

Directional Derivative

- $f_x(x,y) = \frac{\partial f}{\partial x}(x,y)$

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$ = Rate of change of f as we vary y (keeping x fixed).

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$ = Rate of change of f as we vary y (keeping x fixed).
- Directional derivative of $f(x, y)$

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$ = Rate of change of f as we vary y (keeping x fixed).
- Directional derivative of $f(x, y)$ = Rate of change of f if we allow both x and y to change simultaneously (in some direction (u)).

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$ = Rate of change of f as we vary y (keeping x fixed).
- Directional derivative of $f(x, y)$ = Rate of change of f if we allow both x and y to change simultaneously (in some direction (u)).

$$
\begin{aligned}
D_{\vec{u}} f(x, y) &= \nabla f \cdot u \\
&= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \cdot [u_1, u_2] \\
&= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y}
\end{aligned}
$$

Directional Derivative

- $f_x(x, y) = \frac{\partial f}{\partial x}(x, y)$ = Rate of change of f as we vary x (keeping y fixed).
- $f_y(x, y) = \frac{\partial f}{\partial y}(x, y)$ = Rate of change of f as we vary y (keeping x fixed).
- Directional derivative of $f(x, y)$ = Rate of change of f if we allow both x and y to change simultaneously (in some direction (u)).

$$
\begin{aligned}
D_{\vec{u}} f(x, y) &= \nabla f \cdot u \\
&= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \cdot [u_1, u_2] \\
&= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y}
\end{aligned}
$$

Directional derivative can be considered to be a weighted sum of partial derivatives.

Problem 6

Find the derivative of $f(x, y) = xcos(y)$ in the direction of $\vec{u} = [2, 1]$.

Problem 6

Find the derivative of $f(x, y) = xcos(y)$ in the direction of $\vec{u} = [2, 1]$.

$$\frac{\partial f}{\partial x} = \cos(y)$$
$$\frac{\partial f}{\partial y} = -x \sin(y)$$

Problem 6

Find the derivative of $f(x, y) = xcos(y)$ in the direction of $\vec{u} = [2, 1]$.

$$\frac{\partial f}{\partial x} = \cos(y)$$
$$\frac{\partial f}{\partial y} = -x\sin(y)$$

Unit vector in the direction of $\vec{u}$ = $\left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]$

Problem 6

Find the derivative of $f(x, y) = xcos(y)$ in the direction of $\vec{u} = [2, 1]$.

$$\frac{\partial f}{\partial x} = \cos(y)$$
$$\frac{\partial f}{\partial y} = -x\sin(y)$$

Unit vector in the direction of $\vec{u}$ = $\left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]$

$$\begin{aligned} D_{\vec{u}} f(x, y) &= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y} \\ &= \frac{2}{\sqrt{5}} \cos(y) - \frac{1}{\sqrt{5}} x \sin(y) \end{aligned}$$

Problem 7

Find the derivative of $f(x, y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2, -3)$.

Problem 7

Find the derivative of $f(x, y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2, -3)$.

$$\frac{\partial f}{\partial x} = 2x - y$$

$$\frac{\partial f}{\partial y} = -x$$

Problem 7

Find the derivative of $f(x, y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2, -3)$.

$$\frac{\partial f}{\partial x} = 2x - y$$
$$\frac{\partial f}{\partial y} = -x$$

$\vec{u}$ is already a unit vector. $u_1 = 0.6, u_2 = 0.8$.

Problem 7

Find the derivative of $f(x,y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2,-3)$.

$$\frac{\partial f}{\partial x} = 2x - y$$
$$\frac{\partial f}{\partial y} = -x$$

$\vec{u}$ is already a unit vector. $u_1 = 0.6, u_2 = 0.8$.

$$\begin{aligned} D_{\vec{u}} f(x,y) &= 0.6(2x - y) + 0.8(-x) \\ &= 0.6(2x - y) - 0.8x \end{aligned}$$

Problem 7

Find the derivative of $f(x,y) = x^2 - xy$ in the direction of $\vec{u} = 0.6i + 0.8j$ at the point $(2,-3)$.

$$\frac{\partial f}{\partial x} = 2x - y$$
$$\frac{\partial f}{\partial y} = -x$$

$\vec{u}$ is already a unit vector. $u_1 = 0.6, u_2 = 0.8$.

$$\begin{aligned} D_{\vec{u}} f(x,y) &= 0.6(2x-y) + 0.8(-x) \\ &= 0.6(2x-y) - 0.8x \end{aligned}$$

$$\begin{aligned} D_{\vec{u}} f(2,-3) &= 0.6(2(2)+3) - 0.8(2) \\ &= 0.6(7) - 1.6 \\ &= 4.2 - 1.6 \\ &= 2.6 \end{aligned}$$

Thank you.