Lecture 5: Differentiation

Advanced Microeconomics I, ITAM

Xinyang Wang*

A function is differentiable at a point if and only if it has a local affine approximation.

1 One Variable

1.1 Derivative

Given a function $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}$ is open, we say f is *differentiable at* $x_0 \in S$ if

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon) - f(x_0)}{\varepsilon} \text{exists.}$$

In case the limit exists, we write $f'(x_0)$ to be the limit. If f is differentiable at all $x_0 \in S$, we say f is *differentiable*.

Remark.

1. *$f'(x_0)$ gives the slope of the tangent line at x_0.*

*Please email me at xinyang.wang@itam.mx for typos or mistakes. Version: February 2, 2021.

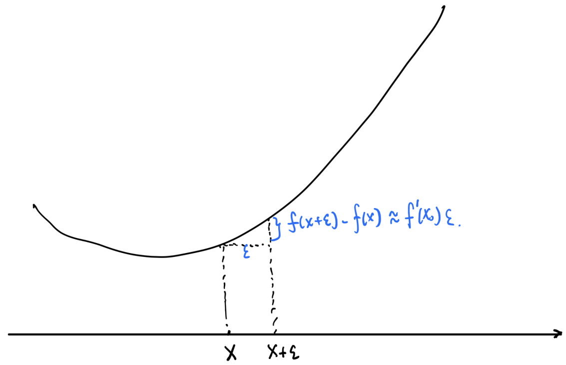

Figure 1: Geometric interpretation of the derivative

It is worth noting that the tangent line at $(x_0, f(x_0))$ *is given by the equation* $y - f(x_0) = f'(x_0)(x - x_0)$.

2. *When a function* f *is differentiable at* x_0*, it admits a local affine approximation:*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$$

Here, $o(x - x_0)$ *is a function* $r : \mathbb{R} \to \mathbb{R}$ *such that* $\lim_{x \to x_0} \frac{r(x)}{x - x_0} = 0$. *Sometimes, this approximation is called a Taylor expansion.*

3. *Once we know a function is differentiable, we can discuss its higher order differentiability inductively: If a function* $f : \mathbb{R} \to \mathbb{R}$ *is* $(n-1)-$ *times differentiable with* $(n-1)$*-th derivative* $f^{(n-1)}$*, then we say the function is* n*-times differentiable at* x_0 *if the limit*

$$\lim_{x \to x_0} \frac{f^{(n-1)}(x) - f^{(n-1)}(x_0)}{x - x_0}$$

exists. If it exists, we denote the n*-th derivative of function* f *at* x_0, $f^{(n)}(x_0)$*, to be this limit. If the function is* n*-times differentiable at all points in its domain, we say the function is* n*-times differentiable.*

4. *A differentiable function must be continuous. The converse may not be true.*

   e.g. $f = |x|$

   A twice differentiable function must be continuously differentiable (i.e. its derivative as *a function on $\mathbb{R}$ is continuous). The converse may not be true.*

e.g. $f = |x|x$

and so on...

5. *When $f'(x) > 0$ at a point x_0, f is increasing in a neighborhood of x_0.*

   When $f'(x) < 0$ at a point x_0, f is decreasing in a neighborhood of x_0.

   When $f'(x) = 0$ at a point x_0, f can be either increasing or decreasing in a neighborhood of x_0. e.g. $f = \pm x^3$.

1.2 Computation

Some frequently used computational rule is given as follows. Functions f, g below are real-valued functions on $\mathbb{R}$.

- $(x^\alpha)' = \alpha x^{\alpha-1}$, in a domain that $x^\alpha, x^{\alpha-1}$ are well defined
- $(\log x)' = 1/x$
- $(e^x)' = e^x$
- $(\sin x)' = \cos x$
- $(\cos x)' = -\sin x$
- Differentiation is a linear operator: $(af + g)' = af' + g'$
- Product Rule: $(fg)' = f'g + fg'$

  Prove it as an exercise.

- Chain Rule: $(f(g(x)))' = f'(g(x))g'(x)$

  A not entirely wrong way to remember the chain rule is by observing the following equation, where Δg terms cancel out:

$$\frac{\Delta f \circ g}{\Delta x} = \frac{\Delta f \circ g}{\Delta g} \cdot \frac{\Delta g}{\Delta x}$$

Exercise.

- *Use the product rule to prove* $(\frac{f}{g})' = \frac{f'g - fg'}{g^2}$.

  Hint: $(\frac{f}{g}) = f \times g^{-1}$.

- *Prove* $(f^{-1})'(y) = \frac{1}{f' f^{-1}(y)}$.

  Hint: $f \circ f^{-1} = id$, *so* $(f \circ f^{-1})' = 1$.

- *Compute* $(a^x)', (x^x)'$.

  Hint: $a^x = e^{x \log a}$.

- *Prove* $f(x) = \begin{cases} x^2 \sin(\frac{1}{x}) & x \neq 0 \\ 0 & x = 0 \end{cases}$ *is differentiable everywhere, but* f' *is not continuous.*

2 Multi variables

2.1 Differentiability

As we have remarked at the beginning of this notes, a function is differentiable at a point if it admits an affine approximation at that point. Recall that in one dimension, a real-valued function is differentiable at x_0 if $f(x) \approx f(x_0) + A(x - x_0)$ for some number A. Similarly, in higher dimensions, for a function $f : \mathbb{R}^n \to \mathbb{R}^m$, a real-valued function is differentiable at x_0 if $f(x) \approx f(x_0) + A(x - x_0)$ for some A. Here, the vector $(x - x_0)$ is in $\mathbb{R}^n$, and $f(x) \in \mathbb{R}^m$. Therefore, to make the expression make sense, A should be an $m \times n$ matrix.

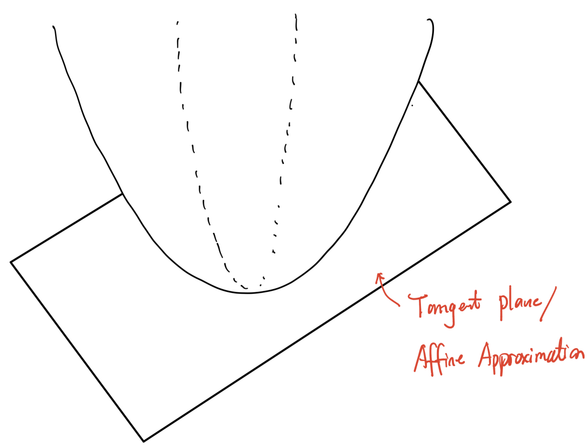

Figure 2: Affine approximation of a differentiable function

Formally, for a function $f : S \to \mathbb{R}^m$, where $S \subset \mathbb{R}^n$ is open, is *differentiable at* $x_0 \in S$ if there is an $m \times n$ matrix A such that

$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - A(x - x_0)\|}{\|x - x_0\|} = 0$$

i.e. $f(x) = f(x_0) + A(x - x_0) + o(\|x - x_0\|)$. Here, the matrix A is the derivative of f at x_0. If f is differentiable everywhere in S, we say f is *differentiable*.

We note that this definition, based on existence, is not very operational, as given a function f, we can not write out its derivative directly. Fortunately, we have the following theorem:

Theorem. *If f is differentiable at x_0, then the derivative of f at x_0 is given by the Jacobian matrix of f at x_0.*

2.2 Jacobian Matrix

2.2.1 Jacobian Matrix of Real-valued Functions

In this subsection, we define the Jacobian Matrix of a funcion.

First, for a real-valued function f on an open set $S \subset \mathbb{R}^n$, we could define the *j-th partial derivative* of f at x by

$$\frac{\partial}{\partial x_j} f(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon e_j) - f(x)}{\varepsilon}$$

if it exists. Here, e_j is the unit vector having an one at its j-th coordinate.

If a function f having j-th partial derivative for all $1 \leq j \leq n$, we are able to define its *gradient* at x_0 by

$$\nabla f(x_0) = \begin{pmatrix} \frac{\partial}{\partial x_1} f \\ \vdots \\ \frac{\partial}{\partial x_n} f \end{pmatrix} (x_0) \in \mathbb{R}^n$$

And the *Jacobian matrix* of f at x_0 is defined by

$$Df(x_0) = \begin{pmatrix} \frac{\partial}{\partial x_1} f & \dots & \frac{\partial}{\partial x_n} f \end{pmatrix} (x_0)$$

One desire at this time might be to give the gradient a more straightforward understanding. For this purpose, we define directional derivative:

For a real-valued function f on an open set $S \subset \mathbb{R}^n$, we could define its *directional derivative at x with respect to v*, where v is a nonzero n-dimensional vector, by

$$D_v f(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon}$$

if it exists.

Intuitively, the directional derivative of f at x with respect to v is just the one-dimensional derivative of f on the one dimensional subset $\{x + tv : t \in \mathbb{R}\}$ at x. It is clear that the j-th partial derivative of f at x is just the directional derivative of f at x with respect to vector e_j. Recall that the derivative of a function is an local affine approximation, therefore, we have the following proposition:

Proposition. *If $D_v f(x)$ and $Df(x)$ are well-defined, $D_v f(x) = Df(x)v$.*

Proof. Prove it by definition. ■

Remark.
- *By this proposition, we know that $D_{cv} f(x) = cD_v f(x)$, for any real number c.*
- *By this proposition, we know that the gradient vector at a point in the domain specifies the steepest growth direction of a function. See the figure below.*
- *Even if the Jacobian of a function $Df(x)$ exists, $D_v f(x)$ may not be well-defined. See the example below.*

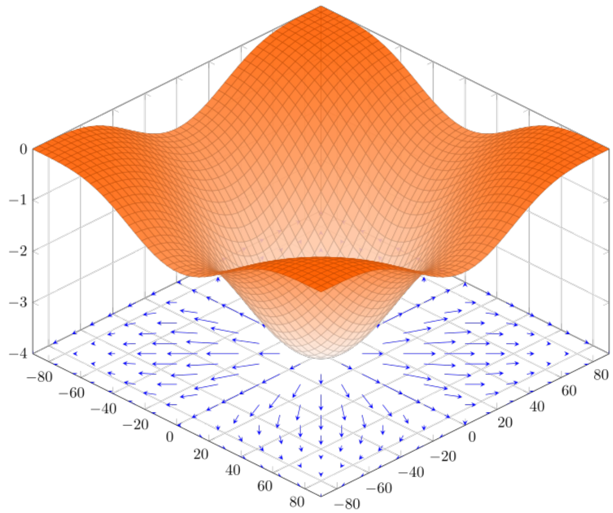

Figure 3: The gradient of $f(x, y) = -(\cos^2 x + \cos^2 y)^2$

Example. *For* $f : \mathbb{R}^2 \to \mathbb{R}$*,*

$$f(x, y) = \begin{cases} \frac{x^3}{x^2+y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$$

Prove that $Df(0, 0)$ *exists, but* f *is not differentiable at* $(0, 0)$*.*

Proof. We prove by contradiction. It is routine to compute that $\frac{\partial}{\partial x_1} f(0, 0) = 1$, $\frac{\partial}{\partial x_2} f(0, 0) = 0$. By the theorem from last section, we know when f is differentiable, then its derivative at $(0, 0)$ is given by its Jacobian $(1, 0)$. However, we plug it back to check the definition

$$\frac{\|f(x, y) - f(0, 0) - (1, 0) \cdot (x, y)\|}{\|(x, y) - (0, 0)\|}$$

should goes to 0, as $(x, y) \to (0, 0)$. However, let $x = y = \varepsilon$, i.e. along the direction $(1, 1)$, we know the fraction does not converge to 0. Therefore, we have a contradiction. ■

Going one step further, we notice that in the above example, $\frac{\partial}{\partial x_1} f$ is not a continuous function. Indeed, if all partial derivatives are continuous. we know the function is differentiable.

Theorem. *If all partial derivatives of a function, not necessarily real-valued, are continuous at a point, we know the function is differentiable at that point.*

2.2.2 Jacobian Matrix of Vector-valued Functions

First, for a vector-valued function $f : S \to \mathbb{R}^m$ on an open set $S \subset \mathbb{R}^n$, we note we could treat it m real-valued function, in the sense that

$$f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

If all partial derivatives of the functions f_1,...,f_m at x exist, we define the Jacobian matrix of f at x by

$$Df(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \cdots & \frac{\partial}{\partial x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x) & \cdots & \frac{\partial}{\partial x_n} f_m(x) \end{pmatrix}$$

We note, if f is differentiable, it admits a local affine transformation in the form

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + o(\|x - x_0\|)$$

2.3 Tangent Planes and Normal Vectors

Given a differentiable function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is open, the tangent plane of f at x_0 is given by the set

$$\{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y - f(x_0) = Df(x_0)(x - x_0)\}$$

In addition, for a plane defined by

$$G(x) = 0$$

for some affine function $G : \mathbb{R}^{n+1} \to \mathbb{R}$, the normal vector of the plane is given by $\nabla G(x_0)$. (Why?)

Example 1. *For a real valued function f on $\mathbb{R}^2$ defined by $f(x, y) = x^2 + 2y^2$, at $(x_0, y_0) =$*

$(1,1)$, *its Jacobian matrix is*

$$Df(x_0, y_0) = (2x_0, 4y_0) = (2, 4)$$

Therefore, the tangent plane at $(1,1)$ *is given by*

$$z - 3 = (2,4) \cdot (x - x_0, y - y_0) = 2(x-1) + 4(y-1)$$

That is, the tangent plane is given by $G = 0$ *where*

$$G = 2x + 4y - z - 3$$

Consequently, the normal vector to the tangent plane is $(2, 4, -1)$.

2.4 Hessian Matrix

For a real-valued function f on an open set $S \subset \mathbb{R}^n$, its *Hessian matrix* at x is given by

$$Hf(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f & \cdots & \frac{\partial^2}{\partial x_n^2} f \end{pmatrix} (x)$$

if all entries are well-defined. When f is twice differentiable, we have the approximation:

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T Hf(x_0)(x - x_0) + o(\|x - x_0\|^2)$$

That is, at some point x_0, the function f admits a local quadratic approximation.

Remark. *(Young's theorem) If all mixed derivatives* $\frac{\partial^2}{\partial x_i \partial x_j} f$ *are continuous in the domain, then* H *is a symmetric matrix.*

As we have remarked, the positive definiteness of the Hessian matrix is equivalent to the convexity of a function:

Theorem. *Given a twice differentiable function* $f : S \to \mathbb{R}$, *where* S *is an open set in* $\mathbb{R}^n$, *then we have*

1. f *is convex* $\iff f(x) \geq f(x_0) + Df(x_0)(x - x_0), \forall x, x_0 \in S$
2. f *is convex* $\iff S$ *is convex and* $Hf(x)$ *is positive semi-definite for all* $x \in S$
3. *If* $Hf(x)$ *is positive definite for all* $x \in S$*, then* f *is strictly convex*

Remark. *We note the converse statement of 3 is not ture. For instance,* $f(x, y) = x^4 + y^4$ *is strictly convex, but its Hessian is* 0 *at the origin.*