Columbia MA Math Camp

Differential Calculus

Vinayak Iyer

July 22, 2020

Table of Contents

Derivative

Continuous fns $\not\Rightarrow$ Differentiable
$\Leftarrow$ ✓

The fundamental concept in calculus is that of the **derivative**, which is simply a rate of change. Consider a function $f : \mathbb{R} \to \mathbb{R}$. The quantity

$$\frac{f(x_0 + h) - f(x_0)}{h}$$

tells us the average rate of change of f between x_0 and $x_0 + h$.

The big idea with derivatives is simply that we let h go to 0.

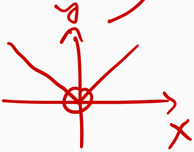
Definition 1.1
Let $A \subset \mathbb{R}$. A function $f : A \to \mathbb{R}$ is said to be differentiable at x_0 iff the limit :

$$f'(x_0) \equiv \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists.

We define the derivative of f at x_0 as $f'(x_0)$

$y = |x|$



$\Rightarrow g(x_0) = mx+b$ where $m = f'(x_0)$

There's another interpretation of the derivative: it is the best linear approximation of a function.

- Suppose you wanted to approximate $f(x)$ by a linear function $g(x) = mx + b$ around the point x_0
- A good approximation should have the following properties:
  - The functions should agree at x_0: $g(x_0) = f(x_0)$. So $g(x_0 + h) = f(x_0) + mh$
  - The per-unit error should be small near x_0:

$$\lim_{h \to 0} \frac{f(x_0 + h) - g(x_0 + h)}{h} = 0$$

when $h=0$

$\Rightarrow g(x_0) = f(x_0)$

These properties combined imply $\boxed{m = f'(x_0)}$

Differentiate a function $f(x)$ wrt x.

Common functions :

- $\frac{d}{dx}c = 0$
- $\frac{d}{dx}x^n = nx^{n-1}$
- $\frac{d}{dx}e^x = e^x$
- $\frac{d}{dx}\log x = \frac{1}{x}$

Combining derivatives :

- $\frac{d}{dx}(f(x) + g(x)) = f'(x) + g'(x)$
- $\frac{d}{dx}(\alpha f(x)) = \alpha f'(x)$ α is a constant
- $\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$ (product rule)
- $\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$ (chain rule)

Quotient Rule :-

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'g - g'f}{(g(x))^2}$$

$f' = f'(x)$

Chain Rule example

- Consider a consumer whose utility u is only directly dependent on consumption c: $u = u(c)$
- However, consumption depends on the consumer's wealth: $c = c(w)$
- Therefore $u = u(c(w))$. We can capture the dependencies in a graph

How much does utility change if I get $1 more?

$$\begin{array}{c} \text{Utility} \\ \downarrow \\ \text{Consumption} \\ \downarrow \\ \text{Wealth} \end{array}$$

How many units does consn change if I increase w by $1

- The chain rule tells us:

Effect on utility

$$\frac{du}{dw} = \frac{du}{dc}\frac{dc}{dw}$$

Why does this make sense..?

$\frac{dU}{dC}$ = How much does utility change if I increase consumption by 1 unit.

$\frac{dU}{dC} = 2x$

Increase w by $1

Table of Contents

Utility:-

$u(x_1, x_2 \dots x_n) : \mathbb{R}^n \to \mathbb{R}$

- We will be working with functions from $\mathbb{R}^n$ to $\mathbb{R}$ a lot
- For functions of two variables (or more), one common tool we'll use is **level curves**. This is the graph of the equation $f(x, y) = c$
- Example: let $a \in \mathbb{R}^2$ and consider the function $f(x, y) = a_1 x + a_2 y$. The graph of this function is called a **hyperplane**

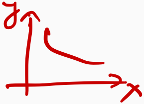
$f : \mathbb{R}^2 \to \mathbb{R}$

$z = f(x, y)$

Give me an example of this!

Fix $z = c$

$f(x, y) = c$



Indifference Curves!

Notion of Linear Approximation

In the single-variable setting, $f'(x)$ was the term in the **best linear approximation** of f

$$f(y) \approx \underbrace{f(x) + f'(y)(y - x)}$$

— Linear function $g(x)$

where "best" meant the relative error goes to 0. That is, $f'(x)$ is the value m such that

— Approximation

$$\lim_{y \to x} \frac{f(y) - \overbrace{(f(x) + m(y - x))}}{y - x} = 0$$

Notion of Linear Approximation (cont.)

$f: \mathbb{R}^2 \to \mathbb{R}^2$ $\quad f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix}$

If $f : \mathbb{R}^n \to \mathbb{R}^m$, then a *linear approximation* of f is:

$y \in \mathbb{R}^n$ ($n \times 1$ vector)

$$\underbrace{f(\mathrm{y})}_{m\times 1} = \underbrace{f(\mathrm{x})}_{m\times 1} + \underbrace{\mathrm{A}}_{m\times n}\overbrace{(\mathrm{y} - \mathrm{x})}^{n\times 1 \text{ vector}}$$

$f(y) \in \mathbb{R}^m$ is an $m \times 1$ vector

where A is a $m \times n$ matrix. Thus we will define the **derivative of f at x** or the **Jacobian of f at x** as the matrix A such that

$$\lim_{y \to x} \frac{f(y) - (f(x) + \mathrm{A}(y - x))}{\|y - x\|} \to 0$$

A is the derivative of f at x

$A = f'(x)$ $\quad m \times n$ matrix

Let us introduce the notion of a very useful concept of a **partial derivative** :

Definition 2.1
For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, it's partial derivative *of the* $\boxed{ith}$ *coordinate wrt to the* $\boxed{jth}$ *argument at $x \in \mathbb{R}^n$ is :*

$$\frac{\partial f_i}{\partial x_j}(x) := \frac{d}{\boxed{dt}} f_i(x + te_j)|_{t=0}$$

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

What's the partial derivative of $f(x) = x_1^2 + x_1x_2$?

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

We want $\frac{\partial f_i(x)}{\partial x_j}$ $\qquad f_i(x)$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_j \\ x_n \end{pmatrix}$$

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

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

$$\frac{\partial f}{\partial x_1} = ? \quad 2x_1 + x_2$$

(Assume x_2 is constant)

$$\frac{\partial f}{\partial x_2} = x_1$$

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

$$f(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix}$$

$$\frac{\partial f_1(x_1, x_2)}{\partial x_2} = \frac{d}{dt}\left(f_1(x_1, x_2 + t)\right)\Big|_{t=0}$$

$$x + te_j = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_j + t \\ \vdots \\ x_n \end{pmatrix}$$

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

$m \times n$ matrix

$x \in \mathbb{R}^n$

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

Properties :

- If $f : \mathbb{R}^n \to \mathbb{R}^m$, then $f'(x)$ is an $m \times n$ matrix also known as the **Jacobian**.
  - If f is real-valued, the column vector $f'(x)^T$ is called the **gradient** of f or sometimes denoted as $\nabla f(x)$

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

$f'(x) = 1 \times n$ vector $\left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n} \right)$

$n \times 1$ vector

$$(f'(x)^T) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

- The derivative is **linear**. If $f, g : \mathbb{R}^n \to \mathbb{R}^m$ and $\alpha \in \mathbb{R}$:
  - $(f+g)'(x) = f'(x) + g'(x)$
  - $(\alpha f)'(x) = \alpha f'(x)$

- If f is differentiable at x, f is continuous at x

$\nabla f(x) = 0$

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The Chain Rule

- Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$.
- Define $h : \mathbb{R}^n \to \mathbb{R}^k$ by $h(x) = g(f(x))$. $\Longleftrightarrow v(w) = u(x(w))$

$x(w) : \mathbb{R} \to \mathbb{R}^2$
$u : \mathbb{R}^2 \to \mathbb{R}$
$v : \mathbb{R} \to \mathbb{R}$

- If f is *differentiable* at x and g is *differentiable* at $f(x)$, then

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

- (Heuristic proof) two linear approximations:

$$\begin{aligned} h(y) &= g(f(y)) \\ &\approx g\left[f(x) + f'(x)(y-x)\right] \\ &\approx g(f(x)) + g'(f(x))f'(x)(y-x) \\ &= h(x) + \underbrace{g'(f(x))f'(x)}_{h'(x)}(y-x) \end{aligned}$$

- Full proof in FMEA, page 96

What does $f'(x)$ look like?

- Consider the partial derivative wrt x_i i.e. :

$$\frac{\partial f(x)}{\partial x_i} = i\text{-th column of } f'(x)$$

1st column of $f'(x)$

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} \\ \vdots \\ \frac{\partial f_m}{\partial x_1} \end{pmatrix}$$

- This allows us to calculate the Jacobian

$$f'(x) = \underbrace{\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}}$$

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

$u(x): \mathbb{R}^2 \to \mathbb{R}$

$v(w) = \mathbb{R} \to \mathbb{R}$

Returning to the chain rule, consider the following example:

- Utility u depends on consumption c and hours worked h
- However c and h depend on the going wage w. Define $x(w) : \mathbb{R} \to \mathbb{R}^2$ by $\boxed{x(w) = (c(w), h(w))}$.
- Define $v(w) = u(x(w))$. The chain rule says $v'(w) = u'(x(w))x'(w)$:

$v(w) = u(x(w))$

$$\begin{aligned} v'(w) &= \underbrace{\begin{pmatrix} \frac{\partial u}{\partial c} & \frac{\partial u}{\partial h} \end{pmatrix}}_{1\times 2} \underbrace{\begin{pmatrix} \frac{\partial c}{\partial w} \\ \frac{\partial h}{\partial w} \end{pmatrix}}_{2\times 1} \\ &= \frac{\partial u}{\partial c}\frac{\partial c}{\partial w} + \frac{\partial u}{\partial h}\frac{\partial h}{\partial w} \end{aligned}$$

$u'(x(w)) = \begin{pmatrix} \frac{\partial u}{\partial c} & \frac{\partial u}{\partial h} \end{pmatrix}$

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

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

$x'(w) = \begin{pmatrix} \frac{\partial x_1}{\partial w} \\ \frac{\partial x_2}{\partial w} \end{pmatrix}$

How much does utility change if

w increases by \$1.

$$u\left(c(w), h(w) \right)$$

$$\frac{\partial U}{\partial w} = \underbrace{\frac{\partial u}{\partial c}\frac{\partial c}{\partial w}}_{\text{How much utility changes due to a \$1 wealth increase through the Cons channel!}} + \underbrace{\frac{\partial u}{\partial h}\frac{\partial h}{\partial w}}_{\text{Effect on utility of wealth through the hours worked channel!}}$$

Being comfortable taking vector derivatives in one step can save you a **lot** of algebra (especially in econometrics). You should know these identities by heart :

- Let $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(x) = Ax$ where A is an $m \times n$ matrix:

$$f'(x) = A$$

- Let $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = x'Ax$ where A is an $n \times n$ matrix:

$$f'(x) = x'(A + A')$$

  If A is symmetric, $f'(x) = 2x'A$

- If $f, g : \mathbb{R}^n \to \mathbb{R}$ and $h(x) = f(x)g(x)$, then

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

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

$f(x) = cx$

$f'(x) = c$

generalization of

x transpose

A' refers to A transpose

$f(x) = ax^2$

$f'(x) = 2ax$

$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_n \end{pmatrix}$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots \beta_n x_n + \epsilon$

OLS: $\min_{\beta} \| y - X\beta \|$

derivative wrt β vector

$f : \mathbb{R}^n \to \mathbb{R}$ $\quad f'(x)$

$1 \times n$ vector

$f(x) = x^T A x$

$$f'(\underline{x}) = \left(\underbrace{\frac{\partial f}{\partial x_1}} \cdots \cdots \frac{\partial f}{\partial x_n} \right)$$

$$f(x) = x^T A x = \underbrace{\sum_i \sum_j^n a_{ij} x_i x_j}$$

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1} \left(a_{11} x_1^2 + a_{12} x_1 x_2 + a_{13} x_1 x_3 \cdots a_{1n} x_1 x_n \right.$$
$$\left. + a_{21} x_2 x_1 \quad a_{31} x_3 x_1 \right) + \cdots$$

$$= 2 a_{11} x_1 + 2 a_{12} x_2 \cdots 2 a_{1n} x_n$$

$$\frac{\partial f}{\partial x_2} = 2 a_{22} x_2 + 2 a_{21} x_1 \cdots 2 a_{2n} x_n$$

$1 \times n$

$$f'(x) = 2 x' A_{n \times n}$$

$$= 2 \left(x_1 \cdots x_n \right) \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots \\ \vdots & & & \\ \vdots & - & \cdots & a_{nn} \end{pmatrix}$$

$$= 2 x_1 a_{11} + 2 x_2 a_{21} \cdots 2 x_n a_{n1}$$

If f is differentiable $\Rightarrow$ partial derivatives exist & Jacobian is matrix of the partials

- We have seen that if f is differentiable, its partials exist and the Jacobian is just the matrix of partial derivatives.
- What if we only know that the partials exist? Is that enough for differentiability?
- Sadly, the answer is no (idea: function could behave nicely along the axes, but misbehave along other directions)
- However, if the partials are also continuous, then the derivative exists. Almost every function we work with in economics will have continuous partial derivatives

Table of Contents

Higher Derivatives: Single Variable

For a function $f : \mathbb{R} \to \mathbb{R}$, the **second derivative** of f at x is the derivative of f' at x

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

The second derivative measures the *change in the slope per unit change in* x:

- $f''(x) > 0$ means the derivative is (locally) increasing in x
- $f''(x) < 0$ means the derivative is (locally) decreasing in x

Can keep going to third, fourth derivatives, etc. (not commonly used)

Taylor Series: Single Variable

Suppose you want to approximate f by a polynomial around the point x_0

$$h(x) \equiv a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots + a_n(x - x_0)^n$$

Two intuitive criteria for a “good” approximation are:

- $h(x_0) = f(x_0)$, which implies $a_0 = f(x_0)$
- The first n derivatives of h should match those of f at x_0

Differentiating repeatedly gives $h^k(x_0) = k!a_k$. Thus the **Taylor series expansion of order n of f around x_0** is

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + ... + \frac{1}{n!}f^n(x_0)(x - x_0)^n$$

Taylor Series: Exact Form of Remainder

Proposition 3.1
Let f be $k+1$ times differentiable on $[a,x]$. Then

$$f(x) = f(a) + f'(a)(x-a) + \ldots + \frac{f^k(a)}{k!}(x-a)^k + \frac{f^{k+1}(\zeta)}{(k+1)!}(x-a)^{k+1}$$

for some $\zeta \in (a,x)$.

Taylor Series: Single Variable (cont.)

- Taylor series are useful tools for many proofs in econometrics
- Approximation methods are used frequently in economics to help simplify nonlinear equations
- Accuracy of the approximation depends on distance from x_0

[Show picture]

Second Derivations: Multiple Variables

- For a function $f : \mathbb{R}^n \to \mathbb{R}$, we define second derivatives similarly, as the derivative of ∇f.
- Evidently, the second derivative of f is an $n \times n$ matrix, called the **Hessian** of f at x
- The form of the Hessian is

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

Schwarz's Theorem

Worried about remembering the order of differentiation for the Hessian? In most cases, there's no need:

Theorem 3.1
(Schwarz) If f is twice continuously differentiable, the Hessian matrix is symmetric.

Taylor Series for Multivariable Functions

- We will consider second-order expansions, as the notation is cumbersome beyond that point (and expansions beyond two orders are rare in economics).
- Let's take $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose we want to approximate f around (x_1^*, x_2^*) with a second-order polynomial

$$\begin{aligned} h(x_1, x_2) &= a_0 + a_1(x_1 - x_1^*) + a_2(x_2 - x_2^*) \\ &+ a_{11}(x_1 - x_1^*)^2 + a_{12}(x_1 - x_1^*)(x_2 - x_2^*) + a_{22}(x_2 - x_2^*)^2 \end{aligned}$$

- We again require that $h(x_0, y_0) = f(x_0, y_0)$ and that all first and second order derivatives of h match f

Taylor Series for Multivariable Functions (cont.)

- Differentiating and matching terms gives

$$\begin{aligned} h(x_1, x_2) &= f(x_1^*, x_2^*) + \frac{\partial f(x_1^*, x_2^*)}{\partial x_1}(x_1 - x_1^*) + \frac{\partial f(x_1^*, x_2^*)}{\partial x_2}(x_2 - x_2^*) \\ &+ \frac{1}{2}\frac{\partial^2 f(x_1^*, x_2^*)}{\partial x_1^2}(x_1 - x_1^*)^2 + \frac{\partial^2 f(x_1^*, x_2^*)}{\partial x_1 \partial x_2}(x_1 - x_1^*)(x_2 - x_2^*) \\ &+ \frac{1}{2}\frac{\partial^2 f(x_1^*, x_2^*)}{\partial x_2^2}(x_2 - x_2^*)^2 \end{aligned}$$

- Much cleaner to write in matrix form:

$$h(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T H(x^*)(x - x^*)$$

- This formula holds for functions with more than 2 variables, and is known as the **second order Taylor series expansion of f around x^***

Table of Contents

Motivation

- Many economic analyses introduce equations of the form $f(x, y) = 0$, where x is a vector of "exogenous" variables and y a vector of "endogenous" variables
- We are frequently interested in understanding the impact of x on y, namely $y'(x)$
- However, the equations are complicated and it may not be possible to solve explicitly for $y(x)$ in order to take derivatives
- The implicit function theorem gives us a way to do such comparative statics even in the absence of a closed form solution

IFT: Two Variables

- Let $f : \mathbb{R}^2 \to \mathbb{R}$
- Assume for every x there is a unique y that satisfies $f(x, y) = 0$. Write $y = y(x)$
- Differentiate the expression $f(x, y(x)) = 0$ with respect to x and apply the chain rule:

$$f_x(x, y(x)) + f_y(x, y(x))y'(x) = 0$$

- So long as $f_y(x, y(x)) \neq 0$, we can solve for $y'(x)$:

$$y'(x) = -\frac{f_x(x, y(x))}{f_y(x, y(x))}$$

IFT: Many variables

- Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$
- Assume for every x there is a unique y that satisfies $f(x, y) = 0$. Write $y = y(x)$
- Now differentiate the expression $f(x, y(x)) = 0$ with respect to x and apply the chain rule:

$$f_x(x, y(x)) + f_y(x, y(x))y'(x) = 0$$

- So long as $f_y(x, y(x))$ is invertible, we have

$$y'(x) = -\left(f_y(x, y(x))^{-1}\right) f_x(x, y(x))$$

- So long as f is continuously differentiable and $det(f_y(x, y(x))) \neq 0$, the above formula is correct. See FMEA page 84 for a full statement.

IFT: Supply and Demand Example

- Let $\theta \in \mathbb{R}^n$ be a vector of variables that affect supply and demand.
- Market clearing implies

$$Q^s(\theta, p) = Q^d(\theta, p)$$

  for all θ

- To put it into our format:

$$\underbrace{Q^s(\theta, p) - Q^d(\theta, p)}_{f(\theta,p):\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}} = 0$$

- This implicitly defines p as a function of θ. Differentiating gives:

$$Q^s_\theta(\theta, p(\theta)) + Q^s_p(\theta, p(\theta))p'(\theta) = Q^d_\theta(\theta, p(\theta)) + Q^d_p(\theta, p(\theta))p'(\theta)$$

IFT: Supply and Demand Example (cont.)

- Solving for $p'(\theta)$ gives:

$$p'(\theta) = \frac{Q^d_\theta - Q^s_\theta}{Q^s_p - Q^d_p}$$

- The denominator is positive
- Therefore the sign of $p'(\theta)$ depends on the sign of $Q^d_\theta - Q^s_\theta$
- If demand reacts more strongly than supply to changes in θ, price increases