ECON 508-A

Multivariate Calculus Review

Juan Ignacio Vizcaino

Washington University in St. Louis

July 29 - August 16 2019

The material here is based on the slides on Multivariate Calculus prepared by Carmen Aston-Figari.

All errors are mine.

Section 1: Introduction

In studying the behavior of non-linear functions $f : \mathbb{R}^N \to \mathbb{R}^M$ in the vicinity of $\hat{x}$,

- We use derivatives to form the linear aproximation $Df(\hat{x})$
- We use linear theory to study the behavior of the linear mapping $Df(\hat{x}) : \mathbb{R}^N \to \mathbb{R}^M$
- We use calculus theory to translate information about the non-linear function f in a neighborhood of $\hat{x}$.

A vector / point $x \in \mathbb{R}^N$ is represented as:

$$x = (x_1, \ldots, x_N) = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}'_{N \times 1}$$

A function $f : \mathbb{R}^N \to \mathbb{R}^M$ can be represented as:

$$f(x) = (f_1(x), \dots, f_M(x))$$

since $f(x)$ is a point in $\mathbb{R}^M$, it can be represented as an $M \times 1$ matrix. Each of its coordinates is a function $f_m(x) : \mathbb{R}^N \to \mathbb{R}$ for $m = 1, \dots, M$

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_M(x) \end{bmatrix}_{M \times 1}$$

Section 2: Partial and Directional Derivatives

Given a function $f : \mathbb{R}^N \to \mathbb{R}^M$,

Definition 1

The partial derivative of f_m with respect to the n th coordinate, x_n, evaluated at the point $\hat{x}$, is:

$$D_n f_m(\hat{x}) = \frac{\partial f_m(\hat{x})}{\partial x_n} = \lim_{t \to 0} \frac{f_m\left(\hat{x}_1, \ldots, \hat{x}_n + t, \ldots, \hat{x}_N\right) - f_m(\hat{x})}{t}$$

assuming that the limit exists.

The Jacobian

The matrix of partial derivatives of all the coordinate functions f_m with respect to all the x_n evaluated at the point $\hat{x}$ is called **Jacobian** of f at $\hat{x}$.

$$Jf(\hat{x}) = \begin{bmatrix} D_1 f_1(\hat{x}) & \dots & D_N f_1(\hat{x}) \\ \vdots & \ddots & \vdots \\ D_1 f_M(\hat{x}) & \cdots & D_N f_M(\hat{x}) \end{bmatrix}_{M \times N}$$

If the partial derivatives $D_n f_m(\hat{x})$ are defined for all $\hat{x}$ in the domain of f, then one can define functions $D_n f_m$.

- A **direction** in $\mathbb{R}^N$ is a vector $v \in \mathbb{R}^N$ s.t. $\|v\| = 1$.

Definition 2

The directional derivative of the m th coordinate function f_m in the direction v evaluated at the point $\hat{x}$, is:

$$D_v f_m(\hat{x}) = \lim_{t \to 0} \frac{f_m(\hat{x} + tv) - f_m(\hat{x})}{t}$$

assuming that the limit exists.

Directional Derivatives - Example

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = 3x_1 + x_1 x_2, \quad \hat{x} = (1,1), \quad v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$\begin{aligned}
D_v f(\hat{x}) &= \lim_{t\to 0} \frac{f(\hat{x} + tv) - f(\hat{x})}{t} \\
&= \lim_{t\to 0} \frac{f\left(1 + \frac{t}{\sqrt{2}}, 1 + \frac{t}{\sqrt{2}}\right) - f(1,1)}{t} \\
&= \lim_{t\to 0} \frac{3\left(1 + \frac{t}{\sqrt{2}}\right) + \left(1 + \frac{t}{\sqrt{2}}\right)^2 - (3+1)}{t} \\
&= \lim_{t\to 0} \frac{3 + \frac{3t}{\sqrt{2}} + 1 + \frac{2t}{\sqrt{2}} + \frac{1}{2}t^2 - 4}{t} \\
&= \lim_{t\to 0} \frac{5}{\sqrt{2}} + \frac{t}{2} \\
&= \frac{5}{\sqrt{2}}
\end{aligned}$$

Directional Derivatives - Interpretation

Let $D_v f(\hat{x})$ denote the M -dimensional vector containing the directional derivatives $D_v f_m(\hat{x})$ for each coordinate function at $\hat{x}$

$$D_v f(\hat{x}) = (D_v f_1(\hat{x}), \ldots, D_v f_M(\hat{x})) = \begin{bmatrix} D_v f_1(\hat{x}) \\ \vdots \\ D_v f_M(\hat{x}) \end{bmatrix}$$

Then, for t small, the change in the function when $\hat{x}$ changes in direction v can be approximated by $D_v f(\hat{x})$:

$$f(\hat{x} + tv) - f(\hat{x}) \approx D_v f(\hat{x}) t$$

Partial and Directional Derivatives

Partial derivatives are a special case of directional derivatives:

Define $e_n = (0, \ldots, 0, 1, 0, \ldots, 0)$ (1 in the nth position), and let $v = e_n$.

$$\begin{aligned} D_{e_n} f_m(\hat{x}) &= \lim_{t \to 0} \frac{f_m(\hat{x} + te_n) - f_m(\hat{x})}{t} \\ &= \lim_{t \to 0} \frac{f_m(\hat{x}_1 + 0, \ldots, \hat{x}_{n-1} + 0, \hat{x}_n + t, \hat{x}_{n+1} + 0, \ldots, \hat{x}_N) - f_m(\hat{x})}{t} \\ &= \lim_{t \to 0} \frac{f_m(\hat{x}_1, \ldots, \hat{x}_n + t, \ldots, \hat{x}_N) - f_m(\hat{x})}{t} \\ &= D_n f_m(\hat{x}) \end{aligned}$$

Section 3: Differentiability

Given a function $f : \mathbb{R}^N \to \mathbb{R}^M$,

Definition 3

f is differentiable at $\hat{x}$ if there exists an $M \times N$ matrix $Df(\hat{x})$ called the derivative of f at $\hat{x}$, such that for any sequence of vectors $h \in \mathbb{R}^N, h \to 0$, the following limit exists:

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

f is differentiable iff f is differentiable at every x in its domain.

Differentiability - Intuition

- $f(\hat{x}+h)$: how much the function changes when $\hat{x}$ changes in h.
- $f(\hat{x})+Df(\hat{x})h$: linear approximation of the change in the function.
- $f(\hat{x}+h)-f(\hat{x})-Df(\hat{x})h$: approximation error

As $h \to 0$, the approximation error goes to zero faster than h.

Differentiable functions admit good linear approximations.

Given a function $f : \mathbb{R} \to \mathbb{R}$ and an element x_0 of the domain, f is said to be **continuous** at the point x_0 when:

for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all x in the domain of f with $x_0 - \delta < x < x_0 + \delta$, the value of $f(x)$ satisfies

$$f(x_0) - \varepsilon < f(x) < f(x_0) + \varepsilon$$

More formally:

Definition 4

f is **continuous** at $\hat{x}$ if $\forall\, \varepsilon > 0,\ \exists\, \delta > 0$ s.t.

$$\|x - \hat{x}\| < \delta \Rightarrow \|f(x) - f(\hat{x})\| < \varepsilon.$$

We say that f is **continuous** if it is continuous at every point x in its domain.

Definition 5

A function $f : \mathbb{R} \to \mathbb{R}$ is C^r (r continuously differentiable function) iff the r th derivative exists and is continuous.

Theorem 1

$f : \mathbb{R}^N \to \mathbb{R}^M$ is C^1 if and only if $D_n f_m$ is continuous for every n, m.

Proof: Rudin. □

In particular, the function Df exists if all partial derivatives are continuous.

We can think of Df as a function from $\mathbb{R}^N$ to the set of all linear transformations from $\mathbb{R}^N$ to $\mathbb{R}^M$ (remember that $Df(x)$ is an $M \times N$ matrix).

Counterexample

$$f : \mathbb{R}^3 \to \mathbb{R}, \quad f(x) = \begin{cases} \frac{x_1^3}{x_2^2+x_3^2} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Df does not exist even though all partials at 0 exist (because the partial derivatives $D_n f$ are not continuous):

$$Df(0) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

$$Df(\boldsymbol{x}) = \begin{bmatrix} \frac{3x_1^2}{x_2^2+x_3^2} & \frac{-2x_1^3 x_2}{\left(x_2^2+x_3^2\right)^2} & \frac{-2x_1^3 x_3}{\left(x_2^2+x_3^2\right)^2} \end{bmatrix} \text{ for } x \neq 0$$

Henceforth, we will assume that f is at least C^1 .

Theorem 2 (The Chain Rule)

Let $f : \mathbb{R}^N \to \mathbb{R}^M, g : \mathbb{R}^M \to \mathbb{R}^L, \hat{x} \in \mathbb{R}^N$, and define the composite function $h : \mathbb{R}^N \to \mathbb{R}^L$ by $h(x) = \mathrm{g} \circ \mathrm{f}(x) = g(f(x))$. If f is differentiable at $\hat{x}$ and g is differentiable at $\hat{y} = f(\hat{x})$, then h is differentiable at $\hat{x}$ and

$$\underbrace{Dh(\hat{\boldsymbol{x}})}_{L\times N} = \underbrace{Dg(f(\hat{\boldsymbol{x}}))}_{L\times M} \underbrace{Df(\hat{\boldsymbol{x}})}_{M\times N}$$

Proof: Omitted. □

Let $f : \mathbb{R} \to \mathbb{R}^2, f(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$

and $g : \mathbb{R}^2 \to \mathbb{R}, g(\boldsymbol{y}) = y_1^2 + y_2$;

$h(x) = 2x^2 \quad Dh(\hat{x}) = [4\hat{x}]$

$Df(\hat{x}) = \begin{bmatrix} 1 \\ 2\hat{x} \end{bmatrix}, \quad Dg(\hat{y}) = \begin{bmatrix} 2\hat{y}_1 & 1 \end{bmatrix}$

$Dh(\hat{x}) = Dg(f(\hat{x})) \cdot Df(\hat{x}) = \begin{bmatrix} 2\hat{x} & 1 \end{bmatrix} \cdot \begin{bmatrix} \hat{1} \\ 2\hat{x} \end{bmatrix} = 4\hat{x}$

Theorem 3

If $f : \mathbb{R}^N \to \mathbb{R}^M$ is differentiable at $\hat{x}$, then for any $v \in \mathbb{R}^N$ such that $\|v\| = 1$,

$$D_v f(\hat{x}) = Df(\hat{x})v$$

Proof.

Define:
$g : \mathbb{R} \to \mathbb{R}^N$ by $g(t) = \hat{x} + tv$
$h : \mathbb{R} \to \mathbb{R}^M$ by $h = f \circ g, h(t) = f(g(t)) = f(\hat{x} + tv)$
The derivative of h evaluated at 0 is:

$$\frac{h(t) - h(0)}{t} = \frac{f(\hat{x} + tv) - f(\hat{x})}{t}$$

or

$$Dh(0) = D_v f(\hat{x})$$

By the Chain rule:

$$Dh(0) = Df(\hat{x})Dg(0)$$

Using the definition of g:

$$Dg(t) = v$$

So:

$$D_v f(\hat{x}) \ = \ Dh(0) \ = \ Df(\hat{x})v$$

□

$f : \mathbb{R}^2 \to \mathbb{R}$
$f(x) = 3x_1 + x_1x_2, \quad \hat{x} = (1,1), \quad v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$
and we got that $D_v f(\hat{x}) = \frac{5}{\sqrt{2}}$
$Df(\hat{x}) = [3 + x_2 \quad x_1] = [4 \quad 1]$

$$Df(\hat{x})v = \left[\begin{array}{cc} 4 & 1 \end{array}\right] \left[\begin{array}{c} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{array}\right] = \frac{5}{\sqrt{2}} = D_v f(\hat{x})$$

Using the previous theorem, let $v = e_n = (0, \ldots 0, 1, 0, \ldots, 0)$

$$D_{e_\eta} f(\hat{x}) = \left[\begin{array}{c} \frac{\partial f_1}{\partial \partial_n} \\ \vdots \\ \frac{\partial f_m}{\partial x_n} \end{array}\right] = Df(\hat{x})e_n,$$ which is the n th column of $Df(\hat{x})$

So $Df(\hat{x}) = Jf(\hat{x})$ if $Df(\hat{x})$ exists.

Theorem 4

A function $f : \mathbb{R}^N \to \mathbb{R}^M$ is differentiable at $\hat{x}$ iff each of its component functions f_m is differentiable at $\hat{x}$.
Moreover, if f is differentiable at $\hat{x}$, the partial derivatives of the component functions f_m exist at $\hat{x}$, and the derivative of f at $\hat{x}$, $Df(\hat{x})$, is the matrix of first partial derivatives of the component functions evaluated at $\hat{x}$.

$$Df(\hat{x}) = \begin{bmatrix} Df_1(\hat{x}) \\ \vdots \\ Df_M(\hat{x}) \end{bmatrix}_{M\times N}$$

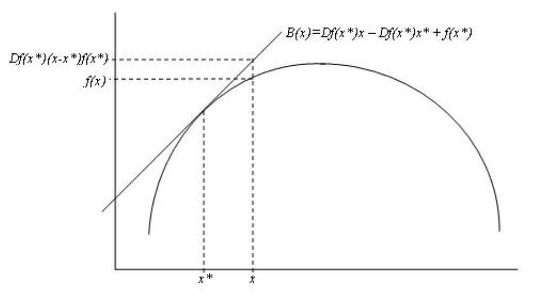
Where each of the $Df_m(\hat{x})$ is a $1 \times N$ row vector,
$Df_m(\hat{x}) = [D_1 f_m(\hat{x}) \cdots D_N f_m(\hat{x})]_{M\times N}$

Proof: Omitted. □

Section 4: Real Valued Functions

If $f : \mathbb{R}^N \to \mathbb{R}$ is differentiable at $\hat{x}$, then the tangent plane is defined as the graph of the function

$$\mathrm{B}(x) = Df(x^*)[x - x^*] + f(x^*)$$



Recall

$$Df(\hat{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}) & \dots & \frac{\partial f}{\partial x_N}(\hat{x}) \end{bmatrix}_{1\times N}$$

The transpose of $Df(\hat{x})$ is called **gradient**

$$\nabla f(\hat{x}) = [Df(\hat{x})]' = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\hat{x}) \end{bmatrix}_{N\times 1} = \left(\frac{\partial f}{\partial x_1}(\hat{x}), \dots, \frac{\partial f}{\partial x_N}(\hat{x}) \right)$$

In particular,

$$\mathrm{D}_v f(\hat{x}) = Df(\hat{x})v = \left[Df(\hat{x})'\right] \cdot v = \nabla f(\hat{x}) \cdot v$$

Directional derivatives can be written as the inner product between the gradient and the direction for a real valued function.

Given a differentiable real valued function $f : \mathbb{R}^N \to \mathbb{R}$ and a point $\hat{x}$ in $\mathbb{R}^N$, $\nabla f(\hat{x})$ points in the direction in which f increases most rapidly.

Proof.

$\max_{v \text{ s.t. } \|v\|=1} D_v f(\hat{x}) = \nabla f(\hat{x}) \cdot v$

Using the geometric definition of the inner product [1],

$$\nabla f(\hat{x}) \cdot v = \|\nabla f(\hat{x})\| \|v\| \cos\theta = \cos\theta \|\nabla f(\hat{x})\|$$

We know that $\cos\theta \in [-1, 1]$ and $\cos\theta = 1$ when $\theta = 0$. So both v and $\nabla f(\hat{x})$ are collinear.

[1] https://en.wikipedia.org/wiki/Dot_product

- Given $f : \mathbb{R}^N \to \mathbb{R}, \quad \hat{x}, \nabla f(\hat{x}) \in \mathbb{R}^N$.
  Given a point $\hat{x}$, $\nabla f(\hat{x})$ points, in the domain, in the direction in which $\hat{x}$ should be increased to obtain the fastest increase in f.

- For $N = 2$, think of a hill: $f(x)$ is altitude, and coordinates are given by $x = (x_1, x_2)$
  The gradient ∇f evaluated at $\hat{x}$ contains all the information we need to know to climbing the hill at the fastest speed possible.

$$\nabla f(\hat{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{x}) \\ \frac{\partial f}{\partial x_2}(\hat{x}) \end{bmatrix}$$

Gradient: Example 1

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

$f(x) = 3 \ln x_1 + \ln x_2, \quad \hat{x} = (2, 2)$

$$\nabla f(\hat{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_2}(\hat{x}) \\ \frac{\partial f}{\partial x_2}(\hat{x}) \end{bmatrix} = \begin{bmatrix} \frac{3}{x_2} \\ \frac{1}{x_2} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$\|\nabla f(\hat{x})\| = \sqrt{\frac{10}{4}} = \frac{\sqrt{10}}{2}$$

$$v = \left(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}} \right)$$

Using the above interpretation, if I travel north a certain small distance, I will ascend 1/2 feet (meters); if I go east, I will go up the hill 3/2 feet (meters). The direction of fastest increase is northeast.

Consider a production function $Q : \mathbb{R}^2 \to \mathbb{R}, Q(K, L) = 4K^{3/4}L^{1/4}$

And the current input bundle is $(\hat{K}, \hat{L}) = (10,000; 625)$

$$\nabla Q(\hat{K}, \hat{L}) = \begin{bmatrix} \frac{\partial Q}{\partial \mathcal{L}}(\hat{K}, \hat{L}) \\ \frac{\partial Q}{\partial L}(\hat{K}, \hat{L}) \end{bmatrix} = \begin{bmatrix} \frac{3\hat{L}^{1/4}}{\hat{K}^{1/4}} \\ \frac{\hat{K}^{3/4}}{\hat{L}^{3/4}} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 8 \end{bmatrix}$$

$$\|\nabla Q(\hat{K}, \hat{L})\| = \frac{\sqrt{41}}{2}$$

$$v = \left(\frac{3}{\sqrt{41}}, \frac{16}{\sqrt{41}} \right)$$

If the firm wants the fastest increase in production, it should add capital and labor at a ratio of 3 to 16 .

Level Sets

Definition 6

$L(c)$ is a level set of the real valued function $f : \mathbb{R}^N \to \mathbb{R}$ iff
$L(c) = \left\{x \mid x \in \mathbb{R}^N \text{ , } f(x) = c\right\}$, where $c \in \mathbb{R}$

We can also define a level set relative to some point $\hat{x}$ in the domain:

$$L(\hat{x}) = \left\{x | x \in \mathbb{R}^N, f(x) = f(\hat{x})\right\}$$

We can completely represent f by its level sets.

They let us reduce by one the number of dimensions needed to represent the function.

Level Sets: Example

$u : \mathbb{R}^2 \to \mathbb{R}, \quad x \in \mathbb{R}^2, \quad u(x) = x_1 x_2$

The level sets of this utility function are indifference curves.

$L(5) = \{x \mid x \in \mathbb{R}^2, u(x) = x_1 x_2 = 5\}$

At $x = (5/2, 2), u(5/2, 2) = 5$

$L(5/2, 2) = \{x \mid x \in \mathbb{R}^2, u(x) = x_1 x_2 = 5\}$

Application of the Chain Rule: Implicit Function Theorem

Consider a continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ and a point $\hat{x} \in R^2$, and let $\hat{y} = f(\hat{x})$

- The level set of f through $\hat{x}$ is the set of points x such that $f(x) = \hat{y}$.
- The Implicit Function theorem states that if f is well behaved at a point $\hat{x}$ then the level set of f through $\hat{x}$ is the graph of a continuously differentiable function, at least near $\hat{x}$.

Implicit Function Theorem continued

In the 2 -dimensional case, $\hat{x}_2 = \psi(\hat{x}_1)$, and

$$\frac{d\psi}{dx_1}(\hat{x}_1) = -\frac{\frac{\partial f}{\partial x_1}(\hat{x})}{\frac{\partial f}{\partial x_2}(\hat{x})}$$

The level set of f through $\hat{x}, L(\hat{x})$, is the graph of ψ.

The behavior close to $\hat{x}$ in $L(\hat{x})$ can be approximated by the tangent line, whose slope is $\frac{d\psi}{dx_1}(\hat{x}_1)$

Implicit Function Theorem: General Case 1

Consider $f : \mathbb{R}^{L+M} \to \mathbb{R}^M$ that is continuously differentiable. Let $O \subseteq \mathbb{R}^L$ be an open subset. Then $f(x) = \hat{y}$ defines implicitly a function $\psi : O \to \mathbb{R}^M$ that defines the last M coordinates of x as a function of the first L coordinates such that $f(x) = \hat{y}$.

The implicit function theorem is used to guarantee that ψ exists and is differentiable.

Even if we don't know ψ, we can compute $D\psi$ using the Chain Rule.

Implicit Function Theorem: General Case 2

Let $Df(\hat{x}) = [D_\lambda f(\hat{x}) \quad D_\mu f(\hat{x})]$. $\quad D_\mu f$ has full rank.

Define $s : O \to R^{L+M}$ such that $s(q) = (q, \psi(q))$, and $h : O \to R^M$ such that $h(x_\lambda) = f(s(x_\lambda))$

Then $h(x_{\boldsymbol{\lambda}}) = f(x_{\boldsymbol{\lambda}}, \psi(x_{\boldsymbol{\lambda}})) = \hat{\boldsymbol{y}}$ for every $x_\lambda \in O$, so $Dh = 0$

By the Chain Rule,

$$\begin{aligned} Dh(\hat{x}_\lambda) &= Df(\hat{x}) Ds(\hat{x}_\lambda) \\ &= [D_\lambda f(\hat{x}) D_\mu f(\hat{x})] \\ &= D_\lambda f(\hat{x}) + D_\mu f(\hat{x}) D\psi(\hat{x}_\lambda) \end{aligned}$$

So $\quad D_\lambda f(\hat{x}) + D_\mu f(\hat{x}) D\psi(\hat{x}_\lambda) = 0$
since $D_Y(\hat{x})$ has full rank and f is continuously differentiable, and since determinant is continuous, $D_\lambda f(\hat{x})$ has full rank for all x in O.
Then:

$$D\psi(\hat{x}_\lambda) = -[D_\mu f(\hat{x})]^{-1} D_\lambda f(\hat{x})$$

Remember we defined the Gradient as

$$\nabla f : \mathbb{R}^N \to \mathbb{R}$$

The derivative of ∇f is called the Hessian of f

$$H(\hat{\boldsymbol{x}}) = D\nabla f(\hat{\boldsymbol{x}}) = \begin{bmatrix} \frac{\partial^2 f}{\partial \hat{x}_1^2} & \cdots & \frac{\partial^2 f}{\partial \hat{x}_1 \partial \hat{x}_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \hat{x}_N \partial \hat{x}_1} & \cdots & \frac{\partial^2 f}{\partial \hat{x}_N^2} \end{bmatrix}_{N\times N}$$

$H(\hat{x})$ is a particular matrix way of displaying $D^2 f(\hat{x})$

The Hessian: Example

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

$$\nabla f(\hat{\boldsymbol{x}}) = \begin{bmatrix} \frac{\ln(\hat{x}_2)}{\hat{x}_1} \\ \frac{\ln(x_1)}{\hat{x}_2} \end{bmatrix}_{2\times 1}$$

$$D\nabla f(\hat{x}) = H(\hat{x}) = \begin{bmatrix} -\ln(\hat{X}_2) & \frac{1}{\tilde{x}_1\hat{x}_2} \\ \frac{1}{\hat{x}_1\hat{x}_2} & \frac{-\ln(\hat{x}_1)}{\hat{x}_2^2} \end{bmatrix}_{2\times 2}$$

For $\hat{x} = (1, 1), H(\hat{x}) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Young's Theorem

Theorem 5 (Young)

If $f : \mathbb{R}^N \to \mathbb{R}$ is C^2 , then the Hessian is symmetric: $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ for all i and j

Proof: Omitted □.

Example

Consider a Cobb Douglas production function $q = kx^a y^b$.

$\frac{\partial Q}{\partial x} = akx^{a-1}y^b$

$\frac{\partial Q}{\partial y} = bkx^a y^{b-1}$

So $\frac{\partial^2 Q}{\partial x \partial y} = abkx^{a-1}y^{b-1} = \frac{\partial^2 Q}{\partial y \partial x}$ as Young's theorem mandates.