Econ 508-A

Multivariate Calculus Review

Carmen Astorne-Figari
Washington University in St. Louis

August 5, 2010

INTRODUCTION

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

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

NOTATION

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

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

Notation (2)

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

$$f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), ..., f_M(\boldsymbol{x}))$$

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

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

Partial Derivatives

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

Definition: the **partial derivative** of f_m with respect to the nth coordinate, x_n, evaluated at the point $\hat{\boldsymbol{x}}$, is:

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

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 $\hat{\boldsymbol{x}}$ is called **Jacobian** of f at $\hat{\boldsymbol{x}}$.

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

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

Directional Derivatives

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

Definition: the **directional derivative** of the mth coordinate function f_m in the direction $\boldsymbol{v}$ evaluated at the point $\hat{\boldsymbol{x}}$, is:

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

assuming that the limit exists.

Directional Derivatives - Example

$$f : \mathbb{R}^2 \to \mathbb{R}$$
$$f(\boldsymbol{x}) = 3x_1 + x_1x_2, \qquad \hat{x} = (1,1)$$
$$v = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

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

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

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

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

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

Partial derivatives are a special case of directional derivatives:

Define $\boldsymbol{e}_n = (0, ..., 0, 1, 0, ..., 0)$ (1 in the nth position), and let $\boldsymbol{v} = \boldsymbol{e}_n$.

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

Definition

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

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

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

(example: the univariate case: $M = N = 1$)

f is **differentiable** iff f is differentiable at every $\boldsymbol{x}$ in its domain.

Intuition

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

[picture for $f : \mathbb{R} \to \mathbb{R}$]

As $\boldsymbol{h} \to \mathbf{0}$, the approximation error goes to zero faster than $\boldsymbol{h}$.

Differentiable functions admit good linear approximations.

CONTINUOUS DIFFERENTIABILITY

Recall the definition of a continuous function:

f is **continuous at** $\hat{\boldsymbol{x}}$ if $\forall \varepsilon > 0, \ \exists \delta > 0$ s.t.
$\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| < \delta \Rightarrow \|f(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})\| < \varepsilon$

We say that f **is continuous** if it is continuous at every point $\boldsymbol{x}$ in its domain.

[picture (two dimensional)]

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

THEOREM: $f : \mathbb{R}^N \to \mathbb{R}^M$ 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 $Df(\boldsymbol{x})$ is an $M \times N$ matrix).

(COUNTER)EXAMPLE

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

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

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

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

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

Differentiability and Continuity

Theorem: differentiabilitiy implies continuity: f is **differentiable at** $\hat{x}$ then f is **continuous at** $\hat{x}$.

The Chain Rule

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

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

$$\underbrace{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}_{L \times N} = \underbrace{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}_{L \times M} \underbrace{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}_{M \times N}$$

Example

Let $f : \mathbb{R} \to \mathbb{R}^2$, $f(x) = (x, x^2)$
and $g : \mathbb{R}^2 \to \mathbb{R}$, $g(\boldsymbol{y}) = y_1^2 + y_2$;
$h(x) = 2x^2$ $Dh(\hat{x}) = [4\hat{x}]$

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

Consider $\hat{x} = 3$, $\hat{\boldsymbol{y}} = (3, 9)$

Then $Dh(3) = 12$, $Df(3) = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$, $Dg(3,9) = \begin{bmatrix} 6 & 1 \end{bmatrix}$ and

$$Dg(3,9)Df(3) = \begin{bmatrix} 6 & 1 \end{bmatrix}_{1\times 2} \begin{bmatrix} 1 \\ 6 \end{bmatrix}_{2\times 1} = 12 = Dh(3)$$

THE DERIVATIVE AND DIRECTIONAL DERIVATIVES

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

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

PROOF:

Define $\begin{array}{lll} g : \mathbb{R} \to \mathbb{R}^N & \text{by} & g(t) = \hat{\boldsymbol{x}} + t\boldsymbol{v} \\ h : \mathbb{R} \to \mathbb{R}^M & \text{by} & h = fog,\ h(t) = f(g(t)) = f(\hat{\boldsymbol{x}} + t\boldsymbol{v}) \end{array}$

$$\begin{array}{lrcll}
\text{substituting} & \dfrac{h(t) - h(0)}{t} & = & \dfrac{f(\boldsymbol{x} + t\boldsymbol{v}) - f(\boldsymbol{x})}{t} & \\
\text{so} & Dh(0) & = & D_{\boldsymbol{v}} f(\hat{\boldsymbol{x}}) & \\
\text{by chain rule} & Dh(0) & = & Df(\hat{\boldsymbol{x}})Dg(0) & \\
\text{and also} & Dg(t) & = & \boldsymbol{v} & \\
\text{hence} & D_{\boldsymbol{v}} f(\hat{\boldsymbol{x}}) & = & Dh(0) & = \quad Df(\hat{\boldsymbol{x}})\boldsymbol{v} \quad \blacksquare
\end{array}$$

Example

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

f($\boldsymbol{x}$)=3x$_1$ $+ x_1 x_2$, $\quad \hat{\boldsymbol{x}} = (1,1), \quad \boldsymbol{v} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

and we got that $D_{\boldsymbol{v}} f(\hat{\boldsymbol{x}}) = \frac{5}{\sqrt{2}}$

$$Df(\hat{\boldsymbol{x}}) = \begin{bmatrix} 3 + x_2 & x_1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \end{bmatrix}$$

$$Df(\hat{\boldsymbol{x}})\boldsymbol{v} = \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{5}{\sqrt{2}} = D_{\boldsymbol{v}} f(\hat{\boldsymbol{x}})$$

Using the previous theorem, let $\boldsymbol{v} = \boldsymbol{e}_n = (0, ...0, 1, 0, ..., 0)$

$$D_{\boldsymbol{e}_n} f(\hat{\boldsymbol{x}}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_n} \\ \vdots \\ \frac{\partial f_m}{\partial x_n} \end{bmatrix} = Df(\hat{\boldsymbol{x}})\boldsymbol{e}_n, \quad \text{which is the } n\text{th column of } Df(\hat{\boldsymbol{x}})$$

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

THEOREM: A function $f : \mathbb{R}^N \to \mathbb{R}^M$ is differentiable at $\hat{\boldsymbol{x}}$ iff each of its component functions f_m is differentiable at $\hat{\boldsymbol{x}}$.

Moreover, if f is differentiable at $\hat{\boldsymbol{x}}$, the partial derivatives of the component functions f_m exist at $\hat{\boldsymbol{x}}$, and the derivative of f at $\hat{\boldsymbol{x}}$, $Df(\hat{\boldsymbol{x}})$, is the matrix of first partial derivatives of the component functions evaluated at $\hat{\boldsymbol{x}}$.

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

Where each of the $Df_m(\hat{\boldsymbol{x}})$ is a $1 \times N$ row vector,

$$Df_m(\hat{\boldsymbol{x}}) = \begin{bmatrix} D_1 f_m(\hat{\boldsymbol{x}}) & \cdots & D_N f_m(\hat{\boldsymbol{x}}) \end{bmatrix}_{M\times N}$$

Chain Rule Hygene

Let $f : \mathbb{R}^{L+M} \to \mathbb{R}^N$ be $f(\boldsymbol{x}, \boldsymbol{y})$,
where $\boldsymbol{y} = g(\boldsymbol{x})$, and $g : \mathbb{R}^L \to \mathbb{R}^M$, $\boldsymbol{x} \in \mathbb{R}^L$

I am interested in calculating $D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))$
(first L columns of $Df(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))$)

$$Df(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) = \left[\; D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) \; \middle| \; D_y f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) \; \right]_{N \times (L+M)}$$

Can we use the chain rule to get $D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))$?

Not yet. f is *not* a composite function of the form specified in the chain rule.

Define: $s : \mathbb{R}^L \to \mathbb{R}^{L+M}$:

$$s(\boldsymbol{x}) = (\boldsymbol{x}, g(\boldsymbol{x})) = \begin{bmatrix} \boldsymbol{x} \\ g(\boldsymbol{x}) \end{bmatrix}_{(L+M)\times 1} \begin{matrix} \leftarrow \text{first } L \text{ rows } (x\text{'s}) \\ \leftarrow \text{last } M \text{ rows } (g(x)\text{'s}) \end{matrix}$$

and $h : \mathbb{R}^L \to \mathbb{R}^N$ by $h(\boldsymbol{x}) = f(s(\boldsymbol{x}))$.

Now we have a composite function, so we can use the chain rule to find $D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))$

By the chain rule,

$$\begin{aligned}
\underbrace{Dh(\hat{\boldsymbol{x}})}_{N\times L} &= \underbrace{Df(s(\hat{\boldsymbol{x}}))}_{N\times(L+M)} \quad \underbrace{Ds(\hat{\boldsymbol{x}})}_{(L+M)\times L} \\
&= Df(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) \quad Ds(\hat{\boldsymbol{x}}) \\
&= \begin{bmatrix} D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) & D_y f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) \end{bmatrix} \begin{bmatrix} I \\ Dg(\hat{\boldsymbol{x}}) \end{bmatrix} \begin{matrix} \leftarrow L\times L \\ \leftarrow M\times L \end{matrix} \\
& \qquad N\times L \qquad\qquad N\times M \qquad\qquad\qquad (L+M)\times L \\
& \qquad\qquad N\times(L+M)
\end{aligned}$$

Solving the matrix multiplication,

$$\begin{matrix} Dh(\hat{\boldsymbol{x}}) & = & D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) & + & D_y f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))Dg(\hat{\boldsymbol{x}}) \\ N\times L & & N\times L & & N\times L \end{matrix}$$

So

$$D_x f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}})) = Dh(\hat{\boldsymbol{x}}) - D_y f(\hat{\boldsymbol{x}}, g(\hat{\boldsymbol{x}}))Dg(\hat{\boldsymbol{x}})$$

AN APPLICATION: INVERSE FUNCTION THEOREM

Intuition:

Suppose that $f : \mathbb{R}^N \to \mathbb{R}^N$ is linear, and $f(\boldsymbol{x}) = \boldsymbol{y}$. Given $\boldsymbol{y}$, we can say that we have a system of N equations witn N unknowns (the components of $\boldsymbol{x}$). We would like to know under what conditions, given $\boldsymbol{y}$, the system can be solved for a value of $\boldsymbol{x}$ that is unique, at least locally.

Since f is linear, we can write the system as $A\boldsymbol{x} = \boldsymbol{y}$. If A is an $N \times N$ invertible matrix $(\Leftrightarrow |A| \neq 0)$, then the system has a unique solution, $\boldsymbol{x}^* = A^{-1}\boldsymbol{y}$.

Put in other words, the system has a unique solution when all the equations in it are linearly independent.

What if f is not linear?

If f is differentiable, look at $Df(\boldsymbol{x}^*)$ in a neighborhood of $\boldsymbol{x}^*$.

Theorem (Inverse Function theorem): Fix $\boldsymbol{x}^* \in \mathbb{R}^N$, let $f : \mathbb{R}^N \to \mathbb{R}^N$ be C^r, where r is a positive integer. Let $\boldsymbol{y}^* = f(\boldsymbol{x}^*)$, and suppose $Df(\boldsymbol{x}^*)$ is invertible. Then there are open sets U; $V \subseteq \mathbb{R}^N$, with $\boldsymbol{x}^* \in U$ and $\boldsymbol{y}^* \in V$, such that $Df(\boldsymbol{x})$ has full rank for all $\boldsymbol{x} \in U$, f maps U 1-1 onto V, and hence has an inverse $f^{-1} : V \to U$. Furthermore, f^{-1} is C^r.

Even if I cannot compute f^{-1}, I can use the Chain Rule to get $Df^{-1}(\boldsymbol{x}^*)$:

Fix $\boldsymbol{x} \in U$ and let $\boldsymbol{y} = f(\boldsymbol{x})$.

Define $h : V \to V$ by $h(\boldsymbol{x}) = f^{-1}(f(\boldsymbol{x}))$.

Then use the chain rule: $Dh = Df^{-1}(\boldsymbol{y})Df(\boldsymbol{x})$

since $h(\boldsymbol{x}) = \boldsymbol{x}$, $Dh(\boldsymbol{x}) = I_N$ ($N \times N$ identity matrix)

Therefore, $I_N = Df^{-1}(\boldsymbol{y})Df(\boldsymbol{x})$

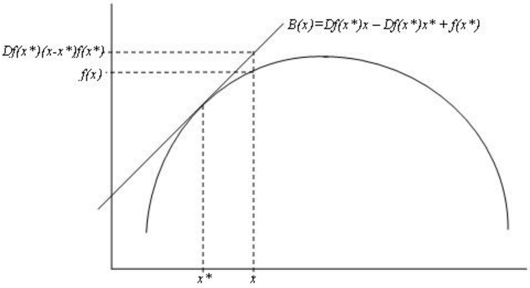
Since $Df(\boldsymbol{x}^*)$ has full rank for all $\boldsymbol{x} \in U$,

$Df^{-1}(\boldsymbol{y}) = [Df(\boldsymbol{x})]^{-1}$

Tangent Plane

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

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



Gradient

Recall

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

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

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

THE GRADIENT AND DIRECTIONAL DERIVATIVES

In particular,

$$\mathrm{D}_{v} f(\hat{\boldsymbol{x}}) = Df(\hat{\boldsymbol{x}})\boldsymbol{v} = [Df(\hat{\boldsymbol{x}})'] \cdot \boldsymbol{v} = \nabla f(\hat{\boldsymbol{x}}) \cdot \boldsymbol{v}$$

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

Key fact about the gradient

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:

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

[picture]

Using the definition of inner product,

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

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.

More gradient intuition

- Given $f : \mathbb{R}^N \to \mathbb{R}$, $\nabla f(\hat{\boldsymbol{x}}) \in \mathbb{R}^N$, just like $\hat{\boldsymbol{x}}$.

  Given a point $\hat{\boldsymbol{x}}$, $\nabla f(\hat{\boldsymbol{x}})$ points, in the *domain*, in the direction in which $\hat{\boldsymbol{x}}$ should be increased to obtain the fastest increase in f.

  [picture]

- For $N = 2$, think of a hill: $f(\boldsymbol{x})$ is altitude, and coordinates are given by $\boldsymbol{x} = (x_1, x_2)$

  The gradient ∇f evaluated at $\hat{\boldsymbol{x}}$ contains all the information we need to know.

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

Example 1

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

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

$$\nabla f(\hat{\boldsymbol{x}}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\hat{\boldsymbol{x}}) \\ \frac{\partial f}{\partial x_2}(\hat{\boldsymbol{x}}) \end{bmatrix} = \begin{bmatrix} \frac{3}{\hat{x}_1} \\ \frac{1}{\hat{x}_2} \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix}$$

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

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

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.

EXAMPLE 2

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 $(\hat{K}, \hat{L}) = (10,000; 625)$

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

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

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

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

[picture]

LEVEL SETS

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

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

$$L(\hat{\boldsymbol{x}}) = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N, \quad f(\boldsymbol{x}) = f(\hat{\boldsymbol{x}})\}$$

We can completely represent f by its level sets.

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

[picture]

Example

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

The level sets of this utility function are indifference curves.

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

At $\boldsymbol{x} = (5/2, 2), \;\; u(5/2, 2) = 5$

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

ANOTHER APPLICATION OF THE CHAIN RULE: IMPLICIT FUNCTION THEOREM

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

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

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

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

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

Behavior close to $\hat{\boldsymbol{x}}$ in $L(\hat{\boldsymbol{x}})$ can be approximated by the tangent line, whose slope is $\dfrac{d\psi}{dx_1}(\hat{x}_1)$.

[picture]

IMPLICIT FUNCTION THEOREM (GENERAL CASE)

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 $f(\boldsymbol{x}) = \hat{\boldsymbol{y}}$ defines implicitly a function $\psi : O \to \mathbb{R}^M$ that defines the last M coordinates of $\boldsymbol{x}$ as a function of the first L coordinates such that $f(\boldsymbol{x}) = \hat{y}$.

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.

Let $Df(\hat{\boldsymbol{x}}) = \begin{bmatrix} D_\lambda f(\hat{\boldsymbol{x}}) & D_\mu f(\hat{\boldsymbol{x}}) \end{bmatrix}$. $D_\mu f$ has full rank.

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

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

Let $\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{\boldsymbol{x}}_{\boldsymbol{\lambda}} & \psi(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) \end{bmatrix}$

By the Chain Rule,
$$\begin{aligned} Dh(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) &= Df(\hat{\boldsymbol{x}}) Ds(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) \\ &= \begin{bmatrix} D_\lambda f(\hat{\boldsymbol{x}}) & D_\mu f(\hat{\boldsymbol{x}}) \end{bmatrix} \begin{bmatrix} I_L \\ D\psi(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) \end{bmatrix} \\ &= D_\lambda f(\hat{\boldsymbol{x}}) + D_\mu f(\hat{\boldsymbol{x}}) D\psi(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) \end{aligned}$$

So $\quad D_\lambda f(\hat{\boldsymbol{x}}) + D_\mu f(\hat{\boldsymbol{x}}) D\psi(\hat{\boldsymbol{x}}_{\boldsymbol{\lambda}}) = 0$

Since $D_\lambda f(\hat{\boldsymbol{x}})$ has full rank and f is continuously differentiable, and since determinant is continuous, $D_\lambda f(\hat{\boldsymbol{x}})$ has full rank for all $\boldsymbol{x}$ in O. Then

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

THE GRADIENT AND THE SLOPE OF THE TANGENT LINE

The gradient of a function $f : \mathbb{R}^2 \to \mathbb{R}$, $\nabla f(\hat{\boldsymbol{x}})$ is perpendicular to the tangent line to the level set $L(\hat{\boldsymbol{x}})$ at $\hat{\boldsymbol{x}}$.

[picture]

This is also true for $N > 2$.

Example

$u : \mathbb{R}^2 \to \mathbb{R}$
$u(x_1, x_2) = x_1 x_2,$
$\hat{\boldsymbol{x}} = (2, 1/2)$
So $u(\hat{\boldsymbol{x}}) = 1$

By direct calculation, we get that the level set of u is the graph of $\psi(x_1) = \frac{1}{x_1}$

Either taking derivatives or using the implicit function theorem, we get $D\psi(x_1) = -\frac{1}{x_1^2}$

So at $\hat{\boldsymbol{x}} = (2, 1/2)$, $D\psi(\hat{x}_1) = -\frac{1}{4}$

Now get the gradient: $\nabla u(\hat{\boldsymbol{x}}) = \begin{bmatrix} \hat{x}_2 \\ \hat{x}_1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}$

The slope of the tangent line is realized by a vector $\begin{bmatrix} 1 \\ -1/4 \end{bmatrix}$

Take the inner product: $\begin{bmatrix} 1/2 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1/4 \end{bmatrix} = 1/2 - 1/2 = 0$

THE HESSIAN

Remember the gradient:

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

The derivative of ∇f is called the **Hessian**.of f.

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

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

Example

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

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

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

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

Young's Theorem

Theorem (**Young**): If $f : \mathbb{R}^N \to \mathbb{R}$ is C^2, then the Hessian is symmetric: $\dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}$ for all i and j.

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