Differential Calculus

Lecture 4

Maria Titova

August 7, 2026

Econ 8000 — Math Camp

Functions of one variable

Average rate of change

Let $I \subseteq \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$.

If the input moves from x_0 to $x_0 + h$, then the value changes by $f(x_0 + h) - f(x_0)$.

For $h \neq 0$, the ratio

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

is the *average rate of change* between the two points.

Secants and the tangent line

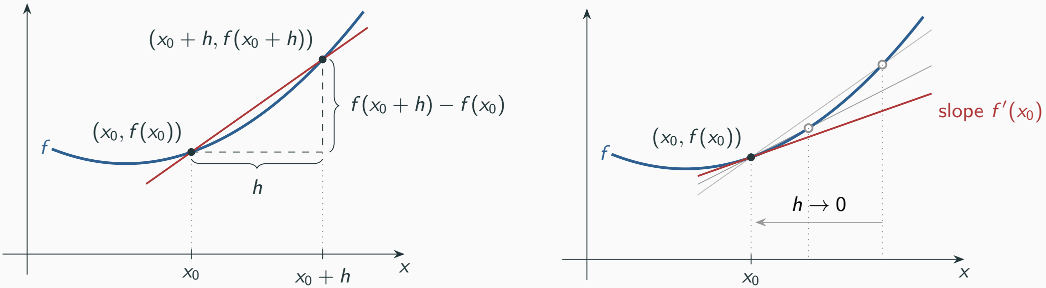

Figure 1.1

Definition 1.1 (Derivative)

Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$, and let $x_0 \in I$.

The function f is *differentiable at* x_0 if there is $a \in \mathbb{R}$ such that

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

The number a is the *derivative* of f at x_0, written $f'(x_0) = a$.

If f is differentiable for all $x_0 \in I$, then $f' : I \to \mathbb{R}$ denotes the *derivative function*.

Three derivatives from the definition

Example 1.2 (Constant, affine, and quadratic functions)

Fix $x_0 \in \mathbb{R}$, and let $a, b, c \in \mathbb{R}$.

Find $f'(x_0)$ for $f(x) = c$, for $f(x) = ax + b$, and for $f(x) = x^2$.

Three derivatives from the definition

Example 1.2 (Constant, affine, and quadratic functions)

Fix $x_0 \in \mathbb{R}$, and let $a, b, c \in \mathbb{R}$.

Find $f'(x_0)$ for $f(x) = c$, for $f(x) = ax + b$, and for $f(x) = x^2$.

Solution.

If $f(x) = c$, then $[f(x_0 + h) - f(x_0)]/h = 0$ for every $h \neq 0$, so $f'(x_0) = 0$.

If $f(x) = ax + b$, then $[a(x_0 + h) + b - (ax_0 + b)]/h = a$, so $f'(x_0) = a$.

If $f(x) = x^2$, then $[(x_0 + h)^2 - x_0^2]/h = 2x_0 + h \longrightarrow 2x_0$, so $f'(x) = 2x$. $\square$

Derivative as a linear approximation

Theorem 1.3 (Derivative as a linear approximation)

Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$, and let $x_0 \in I$.

Then f is differentiable at x_0 if and only if there exists $a \in \mathbb{R}$ such that

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

In that case a is unique, and $a = f'(x_0)$.

Derivative as a linear approximation

Theorem 1.3 (Derivative as a linear approximation)

Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$, and let $x_0 \in I$.

Then f is differentiable at x_0 if and only if there exists $a \in \mathbb{R}$ such that

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

In that case a is unique, and $a = f'(x_0)$.

Proof sketch.

Subtracting a and using that $|u(h)| \to 0$ if and only if $u(h) \to 0$,

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = a \iff \lim_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - ah|}{|h|} = 0.$$

$\square$

Fixed point and moving input

In the definition of differentiability above, x_0 changes to $x_0 + h$ as $h \to 0$.

Sometimes it is useful to have x_0 change to x as $x \to x_0$ instead.

Equivalently, f is differentiable at x_0 if and only if there exists $a \in \mathbb{R}$ such that

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

Thus the two forms use the same change:

$$h = x - x_0, \qquad x = x_0 + h.$$

Linear approximation and remainder

Let $r(h) = f(x_0 + h) - f(x_0) - ah$.

Then f is differentiable at x_0 if and only if there exists $a \in \mathbb{R}$ such that

$$f(x_0 + h) = f(x_0) + ah + r(h), \text{ where } \lim_{h \to 0} \frac{|r(h)|}{|h|} = 0.$$

- Here $f(x_0) + ah$ is the *linear approximation* of f evaluated at $x_0 + h$.
- The term $r(h)$ is the *remainder*.
- The condition $|r(h)|/|h| \to 0$ says that the remainder is negligible compared to h.

Reading the derivative from an expansion

Example 1.4 (Reading the derivative from an expansion)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$, and let $x_0 = 3$.

Example 1.4 (Reading the derivative from an expansion)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$, and let $x_0 = 3$.

Solution.

The function evaluated at $x_0 + h$ is

$$f(3+h) = (3+h)^2 = \underbrace{9}_{f(3)} + \underbrace{6h}_{\text{linear term}} + \underbrace{h^2}_{r(h)}.$$

The remainder satisfies

$$\lim_{h\to 0} \frac{|r(h)|}{|h|} = \lim_{h\to 0} \frac{|h^2|}{|h|} = \lim_{h\to 0} |h| = 0.$$

Therefore the linear coefficient is the derivative: $f'(3) = 6$. □

Two ways of computing a derivative

We now have two ways of calculating derivatives for functions from $\mathbb{R}$ to $\mathbb{R}$.

- The difference-quotient definition.
- The linear-expansion method of the theorem above.

The latter method is generalizable to functions of several variables.

Differentiability implies continuity

Theorem 1.5 (Differentiability implies continuity)

Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$, and let $x_0 \in I$.

If f is differentiable at x_0, then f is continuous at x_0.

Continuity without differentiability

The converse fails: a continuous function need not be differentiable.

Example 1.6 (Continuity without differentiability)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = |x|$.

The converse fails: a continuous function need not be differentiable.

Example 1.6 (Continuity without differentiability)

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = |x|$.

Solution.

At the origin, $|f(x) - f(0)| = |x| \to 0$, so f is continuous.

But for $h \neq 0$, $[f(h) - f(0)]/h = |h|/h$ equals 1 when $h > 0$ and -1 when $h < 0$.

It has no limit as $h \to 0$, so f is not differentiable at 0. $\square$

Rules of differentiation

Theorem 1.7 (Rules of differentiation)

Let f and g be differentiable at x_0, and let $\alpha, \beta \in \mathbb{R}$. Then

1. $(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0)$;
2. $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$;
3. *if* $g(x_0) \neq 0$, *then*
$$(f/g)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}.$$

Chain rule

Theorem 1.8 (Chain rule)

Let $I, J \subseteq \mathbb{R}$ be open intervals, let $g : I \to \mathbb{R}$ satisfy $g(I) \subseteq J$, and let $f : J \to \mathbb{R}$.

If g is differentiable at x_0 and f at $g(x_0)$, then $f \circ g$ is differentiable at x_0 and

$$(f \circ g)'(x_0) = f'(g(x_0))\, g'(x_0).$$

Interior extremum

Theorem 1.9 (Interior extremum)

Let $a < b$*, let* $f : [a, b] \to \mathbb{R}$*, and let* $x^* \in (a, b)$*.*

Suppose f *is differentiable at* x^* *and that for some* $\varepsilon > 0$*,*

$$f(x^*) \geq f(x) \qquad \text{for every } x \in [a, b] \text{ with } |x - x^*| < \varepsilon.$$

Then $f'(x^*) = 0$*, and the same holds if instead* $f(x^*) \leq f(x)$ *for every such* x*.*

Theorem 1.9 (Interior extremum)

Let $a < b$, let $f : [a, b] \to \mathbb{R}$, and let $x^* \in (a, b)$.

Suppose f is differentiable at x^* and that for some $\varepsilon > 0$,

$$f(x^*) \geq f(x) \qquad \text{for every } x \in [a, b] \text{ with } |x - x^*| < \varepsilon.$$

Then $f'(x^*) = 0$, and the same holds if instead $f(x^*) \leq f(x)$ for every such x.

Proof sketch.

The hypothesis makes the difference quotient ≤ 0 for $h > 0$ and ≥ 0 for $h < 0$.

Both one-sided limits equal $f'(x^*)$, so $f'(x^*) \leq 0$ and $f'(x^*) \geq 0$. □

Rolle's theorem

Theorem 1.10 (Rolle's theorem)

Let $a < b$ and let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).

If $f(a) = f(b)$, then there is $c \in (a, b)$ such that $f'(c) = 0$.

Rolle's theorem

Theorem 1.10 (Rolle's theorem)

Let $a < b$ and let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).

If $f(a) = f(b)$, then there is $c \in (a, b)$ such that $f'(c) = 0$.

Proof sketch.

If f is constant, then $f'(c) = 0$ for every $c \in (a, b)$.

Otherwise $f(t) > f(a)$ for some t, and f has a maximizer $c \in [a, b]$ by Weierstrass.

Then $f(c) \geq f(t) > f(a) = f(b)$, so $c \in (a, b)$ and $f'(c) = 0$ by interior extremum.

If instead $f(t) < f(a)$, use a minimizer. $\square$

Mean value theorem

Theorem 1.11 (Mean value theorem)

Let $a < b$ *and let* $f : [a, b] \to \mathbb{R}$ *be continuous on* $[a, b]$ *and differentiable on* (a, b).

Then there is $c \in (a, b)$ *such that*

$$f(b) - f(a) = (b - a)f'(c).$$

Mean value theorem

Theorem 1.11 (Mean value theorem)

Let $a < b$ and let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).

Then there is $c \in (a, b)$ such that

$$f(b) - f(a) = (b - a)f'(c).$$

Proof sketch.

Let $g(t) = f(t) - f(a) - \frac{f(b)-f(a)}{b-a}(t - a)$ for $t \in [a, b]$.

Then g is continuous on $[a, b]$, differentiable on (a, b), and $g(a) = g(b) = 0$.

By Rolle's theorem there is $c \in (a, b)$ with $0 = g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}$. $\square$

Real-valued functions of several variables

Partial derivatives

Now we consider functions from $\mathbb{R}^n$ to $\mathbb{R}$, with column-vector inputs.

Recall from Lecture 3 that e_i is the ith standard basis vector of $\mathbb{R}^n$.

- It has 1 in coordinate i and 0 in every other coordinate.
- So $x_0 + te_i$ adds t to coordinate i of x_0 and leaves the others unchanged.

Definition 2.1 (Partial derivative)

Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$, and let $x_0 \in U$.

The *ith partial derivative* of f at x_0, where $i \in \{1, \ldots, n\}$, is

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{t \to 0} \frac{f(x_0 + te_i) - f(x_0)}{t},$$

provided the limit exists.

A function of two variables

This is an ordinary one-variable derivative.

- Hold all variables except x_i fixed and differentiate with respect to x_i.

Example 2.2 (A function of two variables)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

This is an ordinary one-variable derivative.

- Hold all variables except x_i fixed and differentiate with respect to x_i.

Example 2.2 (A function of two variables)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

Solution.

Holding x_2 fixed and differentiating in x_1, and then reversing their roles, gives

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1 x_2, \qquad \frac{\partial f}{\partial x_2}(x_1, x_2) = x_1^2.$$

$\square$

Partial derivatives describe a function only along the coordinate lines through a point.

Differentiability requires one linear approximation to work in all directions at once.

Definition 2.3 (Differentiability)

Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$, and let $x_0 \in U$.

The function f is *differentiable at* x_0 if there is $a \in \mathbb{R}^n$ such that
$$\lim_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - a \cdot h|}{\|h\|} = 0,$$
where the limit is over $h \in \mathbb{R}^n$ with $h \neq 0$, and $a \cdot h$ is the inner product.

It is *differentiable on* U if it is differentiable at every point of U.

Fixed base point and moving input

In the definition, x_0 is fixed and $h \to 0$ through vectors in $\mathbb{R}^n$.

Set $x = x_0 + h$, or equivalently $h = x - x_0$. Then the same definition is

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

The one-variable case

When $n = 1$, the vector h and the coefficient a are scalars.

- Then $a \cdot h = ah$ and $\|h\| = |h|$.
- So the definition above is exactly the linear-approximation form of the derivative.

Equivalently,

$$f(x_0 + h) = f(x_0) + a \cdot h + r(h), \qquad \frac{|r(h)|}{\|h\|} \longrightarrow 0 \quad \text{as } h \to 0.$$

So the linear approximation at $x_0 + h$ is now $f(x_0) + a \cdot h$.

Gradient

Definition 2.4 (Gradient)

If all partial derivatives of f exist at x_0, the *gradient* of f at x_0 is the column vector

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

Consequences of differentiability

When f is differentiable, the vector in its linear approximation is exactly the gradient.

Theorem 2.5 (Consequences of differentiability)

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable at $x_0 \in U$.

Let $a \in \mathbb{R}^n$ be as in the definition of differentiability.

Then every partial derivative of f exists at x_0 and $a = \nabla f(x_0)$.

Moreover, f is continuous at x_0.

Example 2.6 (A linear approximation in two variables)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x_1, x_2) = x_1^2 x_2$, and let $x_0 = (1, 2)^\top$.

Example 2.6 (A linear approximation in two variables)

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x_1, x_2) = x_1^2 x_2$, and let $x_0 = (1, 2)^\top$.

Solution.

Expanding around x_0 gives

$$\begin{aligned} f(x_0 + h) &= (1 + h_1)^2(2 + h_2) \\ &= \underbrace{2}_{f(x_0)} + \underbrace{4h_1 + h_2}_{(4,1)^\top \cdot h} + \underbrace{2h_1^2 + 2h_1 h_2 + h_1^2 h_2}_{r(h)}. \end{aligned}$$

For $\|h\| \le 1$ we have $|r(h)| \le 5\|h\|^2$, so $|r(h)|/\|h\| \le 5\|h\| \to 0$.

Thus f is differentiable at x_0 and $\nabla f(x_0) = (4, 1)^\top$. $\square$

Partial derivatives can exist without a valid linear approximation.

Definition 2.7 (C^1 function)

Let $U \subseteq \mathbb{R}^n$ be open.

A function $f : U \to \mathbb{R}$ is C^1 *on* U if every $\partial f/\partial x_i$ exists on U and is continuous.

We also say that f is *continuously differentiable on* U.

Theorem 2.8 (C^1 implies differentiable)

If $f : U \to \mathbb{R}$ is C^1 on the open set $U \subseteq \mathbb{R}^n$, then f is differentiable on U.

The gradient $\nabla f : U \to \mathbb{R}^n$ is then continuous.

Verifying differentiability

This provides the usual way to verify differentiability.

- Compute the partial derivatives and check that they are continuous.
  - In particular, polynomials are C^1 on $\mathbb{R}^n$.
  - Exponential functions are C^1 on $\mathbb{R}^n$.

The implications established in this section are

$$C^1 \implies \text{differentiable} \implies \text{continuous},$$

$$\text{differentiable} \implies \text{all partial derivatives exist.}$$

None of the converses holds in general.

Differential notation

The linear approximation can also be written using differential notation.

Remark 2.9 (Differential notation)

If f is differentiable at x_0, its *differential at* x_0 is the linear map

$$\mathrm{d}f_{x_0}(h) = \nabla f(x_0) \cdot h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) h_i.$$

Thus it is exactly the linear term in

$$f(x_0 + h) = f(x_0) + \boldsymbol{\nabla f(x_0) \cdot h} + r(h).$$

Remark 2.9 (Differential notation)

Economics texts often write the same expression as

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0)\,\mathrm{d}x_i.$$

Here $\mathrm{d}x_i$ labels a coordinate change.

- It is not the product of a number d and x_i.

It is just different notation for the linear map $\mathrm{d}f_{x_0}$ from previous slide.

Directional derivatives

A directional derivative allows to approximate the change in f along a given direction.

- Let $v = (v_1, \dots, v_n)^\top$ be a direction vector in $\mathbb{R}^n$.
- Then $x_0 + tv$ for $t \in \mathbb{R}$ changes x_0 in the direction of v with "weight" t.

Definition 2.10 (Directional derivative)

Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$, and let $x_0 \in U$.

The *directional derivative* of f at x_0 in the direction $v \in \mathbb{R}^n$ is

$$\partial_v f(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t},$$

provided the limit exists.

Partial derivatives are special cases of directional derivatives, with $v = e_i$.

Directional derivatives from the gradient

When f is differentiable, its gradient gives every directional derivative.

Theorem 2.11 (Directional derivatives from the gradient)

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable at $x_0 \in U$.

Then, for every $v \in \mathbb{R}^n$, the directional derivative exists and

$$\partial_v f(x_0) = \nabla f(x_0) \cdot v = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_0) v_i.$$

Steepest increase

To compare directions independently of scale, we restrict v to unit vectors ($\|v\| = 1$).

- We ask which one gives the largest directional derivative.

Theorem 2.12 (Steepest increase)

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable at $x_0 \in U$.

Suppose $\nabla f(x_0) \neq 0$ and define the unit vector $v^* = \nabla f(x_0)/\|\nabla f(x_0)\|$.

Among all unit vectors $v \in \mathbb{R}^n$, the unique maximizer of $\partial_v f(x_0)$ is $v = v^*$.

The unique minimizer is $v = -v^*$.

The corresponding values are

$$\partial_{v^*} f(x_0) = \|\nabla f(x_0)\|, \qquad \partial_{-v^*} f(x_0) = -\|\nabla f(x_0)\|.$$

Gradient descent

This result underlies *gradient descent*.

- It is an algorithm used to train many machine-learning and deep learning models.

Let f be the loss function and let x_k be the vector of model weights at iteration k.

Gradient descent chooses a step size $\alpha_k > 0$ and updates

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

- α_k here is the learning rate chosen by the user.

The negative gradient gives the direction of steepest local decrease.

Vector-valued functions

Vector-valued functions

A map $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is described by m component functions $F_i : U \to \mathbb{R}$:

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_m(x) \end{pmatrix}, \qquad x \in U.$$

Each F_i is a real-valued function of n variables.

- We already know how to differentiate each $F_i : U \to \mathbb{R}$.
- It remains to package all m of those derivatives together.

Jacobian

A linear approximation of a map $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$.

- It is represented by an $m \times n$ matrix.

Definition 3.1 (Jacobian)

Let $U \subseteq \mathbb{R}^n$ be open, let $F : U \to \mathbb{R}^m$, and let $x_0 \in U$.

The map F is *differentiable at* x_0 if there is an $m \times n$ matrix A such that

$$\lim_{h \to 0} \frac{\|F(x_0 + h) - F(x_0) - Ah\|}{\|h\|} = 0.$$

The matrix A is the *derivative*, or *Jacobian*, of F at x_0, written $A = \mathrm{D}F(x_0)$.

Jacobian approximation

Equivalently,

$$F(x_0 + h) = F(x_0) + \mathrm{D}F(x_0)h + r(h), \qquad \frac{\|r(h)\|}{\|h\|} \longrightarrow 0 \quad \text{as } h \to 0.$$

So the linear approximation of F at $x_0 + h$ is $F(x_0) + \mathrm{D}F(x_0)h$.

Reading the Jacobian approximation

With the moving input $x = x_0 + h$, the same definition is

$$\lim_{x \to x_0} \frac{\|F(x) - F(x_0) - \mathrm{D}F(x_0)(x - x_0)\|}{\|x - x_0\|} = 0.$$

The linear term has dimensions

$$\mathrm{D}F(x_0)h : \qquad (m \times n)(n \times 1) = m \times 1.$$

so each row corresponds to an output and each column to an input.

C^1 functions

Definition 2.7 (C^1 function)

The map F is C^1 *on* U if every component F_i is C^1 on U.

- It follows that every C^1 map is differentiable.

Entries of the Jacobian

Theorem 3.2 (Entries of the Jacobian)

F is differentiable at x_0 if and only if every component F_i is differentiable at x_0.

In that case the Jacobian is unique and

$$\mathrm{D}F(x_0) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x_0) & \cdots & \frac{\partial F_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(x_0) & \cdots & \frac{\partial F_m}{\partial x_n}(x_0) \end{pmatrix}.$$

Thus the ith row of $\mathrm{D}F(x_0)$ is $\nabla F_i(x_0)^\top$.

Example 3.3 (A 2×3 Jacobian)

Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$F(x, y, z) = \begin{pmatrix} x^2 y \\ y + 3z \end{pmatrix}.$$

Example 3.3 (A 2×3 Jacobian)

Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$F(x, y, z) = \begin{pmatrix} x^2 y \\ y + 3z \end{pmatrix}.$$

Solution.

Then

$$\mathrm{D}F(x, y, z) = \begin{pmatrix} 2xy & x^2 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \qquad \mathrm{D}F(x_0) = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix} \quad \text{at } x_0 = (1, 2, 0)^\top.$$

The two rows correspond to the two outputs.

The three columns correspond to the inputs x, y, z. $\square$

Gradient and Jacobian

For $f : U \to \mathbb{R}$, the Jacobian is a $1 \times n$ row, whereas the gradient is an $n \times 1$ column:

$$\mathrm{D}f(x_0) = \nabla f(x_0)^\top.$$

Consequently, $\mathrm{D}f(x_0)h = \nabla f(x_0) \cdot h$.

The gradient is convenient for geometric statements.

The Jacobian is convenient for function composition (chain rule on the next slide).

When $n = 1$, both contain the ordinary derivative $f'(x_0)$.

Chain rule

The chain rule for vector-valued functions is a straightforward generalization of the one-variable chain rule.

- Matrix multiplication takes the place of scalar multiplication.

Theorem 3.4 (Chain rule)

Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open.

Let $G : U \to \mathbb{R}^m$ satisfy $G(U) \subseteq V$, and let $F : V \to \mathbb{R}^p$.

If G is differentiable at $x_0 \in U$ and F at $G(x_0)$, then $F \circ G$ is differentiable at x_0 and

$$\mathrm{D}(F \circ G)(x_0) = \mathrm{D}F(G(x_0))\,\mathrm{D}G(x_0).$$

Reading the chain rule

Using the row description of the Jacobian, the product is

$$\mathrm{D}(F \circ G)(x_0) = \underbrace{\begin{pmatrix} \nabla F_1(G(x_0))^\top \\ \vdots \\ \nabla F_p(G(x_0))^\top \end{pmatrix}}_{p \times m} \underbrace{\begin{pmatrix} \nabla G_1(x_0)^\top \\ \vdots \\ \nabla G_m(x_0)^\top \end{pmatrix}}_{m \times n}.$$

The product is $p \times n$, the required size for the derivative of $F \circ G : U \to \mathbb{R}^p$.

The outer rows are evaluated at $G(x_0)$, while the inner rows are evaluated at x_0.

When $n = m = p = 1$, the formula reduces to the one-variable chain rule.

Derivative along a curve

A useful special case follows a curve through the domain of a real-valued function.

Corollary 3.5 (Derivative along a curve)

Let $V \subseteq \mathbb{R}^m$ be open and let $I \subseteq \mathbb{R}$ be an open interval.

Let $\gamma : I \to V$ be differentiable at $t_0 \in I$, and let $f : V \to \mathbb{R}$ be differentiable at $\gamma(t_0)$.

Then

$$(f \circ \gamma)'(t_0) = \nabla f(\gamma(t_0)) \cdot \gamma'(t_0) = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(\gamma(t_0))\gamma_i'(t_0).$$

- Indeed, the chain rule gives $\mathrm{D}f(\gamma(t_0))\mathrm{D}\gamma(t_0) = \nabla f(\gamma(t_0))^\top \gamma'(t_0)$.

A vector-valued composition

Example 3.6 (A vector-valued composition)

Define

$$G : \mathbb{R} \to \mathbb{R}^3, \quad G(t) = \begin{pmatrix} 1+t \\ 2 \\ 3+2t \end{pmatrix}, \qquad F : \mathbb{R}^3 \to \mathbb{R}^2, \quad F(u,v,w) = \begin{pmatrix} uw \\ v+w^2 \end{pmatrix}.$$

A vector-valued composition

Example 3.6 (A vector-valued composition)

Define

$$G : \mathbb{R} \to \mathbb{R}^3, \quad G(t) = \begin{pmatrix} 1+t \\ 2 \\ 3+2t \end{pmatrix}, \qquad F : \mathbb{R}^3 \to \mathbb{R}^2, \quad F(u, v, w) = \begin{pmatrix} uw \\ v + w^2 \end{pmatrix}.$$

Solution.

To find the derivative of $F \circ G$ at $t = 0$, compute

$$\mathrm{D}G(t) = G'(t) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \qquad \mathrm{D}F(u, v, w) = \begin{pmatrix} w & 0 & u \\ 0 & 1 & 2w \end{pmatrix}.$$

□

Example 3.6 (A vector-valued composition)

$G(t) = (1 + t,\ 2,\ 3 + 2t)^\top$ and $F(u, v, w) = (uw,\ v + w^2)^\top$.

The chain rule product

Example 3.6 (A vector-valued composition)

$G(t) = (1 + t,\ 2,\ 3 + 2t)^\top$ and $F(u, v, w) = (uw,\ v + w^2)^\top$.

Solution.

Because $G(0) = (1, 2, 3)^\top$, the chain rule gives

$$\begin{aligned} \mathrm{D}(F \circ G)(0) &= \mathrm{D}F(G(0))\,\mathrm{D}G(0) \\ &= \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 12 \end{pmatrix}. \end{aligned}$$

The dimensions are $(2 \times 3)(3 \times 1) = 2 \times 1$, as required for a map from $\mathbb{R}$ to $\mathbb{R}^2$. $\square$

A direct check

Example 3.6 (A vector-valued composition)

$G(t) = (1 + t,\ 2,\ 3 + 2t)^\top$ and $F(u, v, w) = (uw,\ v + w^2)^\top$.

A direct check

Example 3.6 (A vector-valued composition)

$G(t) = (1 + t,\ 2,\ 3 + 2t)^\top$ and $F(u, v, w) = (uw,\ v + w^2)^\top$.

Solution.

For a direct check,

$$(F \circ G)(t) = \begin{pmatrix} (1+t)(3+2t) \\ 2 + (3+2t)^2 \end{pmatrix} = \begin{pmatrix} 3 + 5t + 2t^2 \\ 11 + 12t + 4t^2 \end{pmatrix}.$$

Differentiating the components at 0 again gives $(5, 12)^\top$. □