Chapter 9

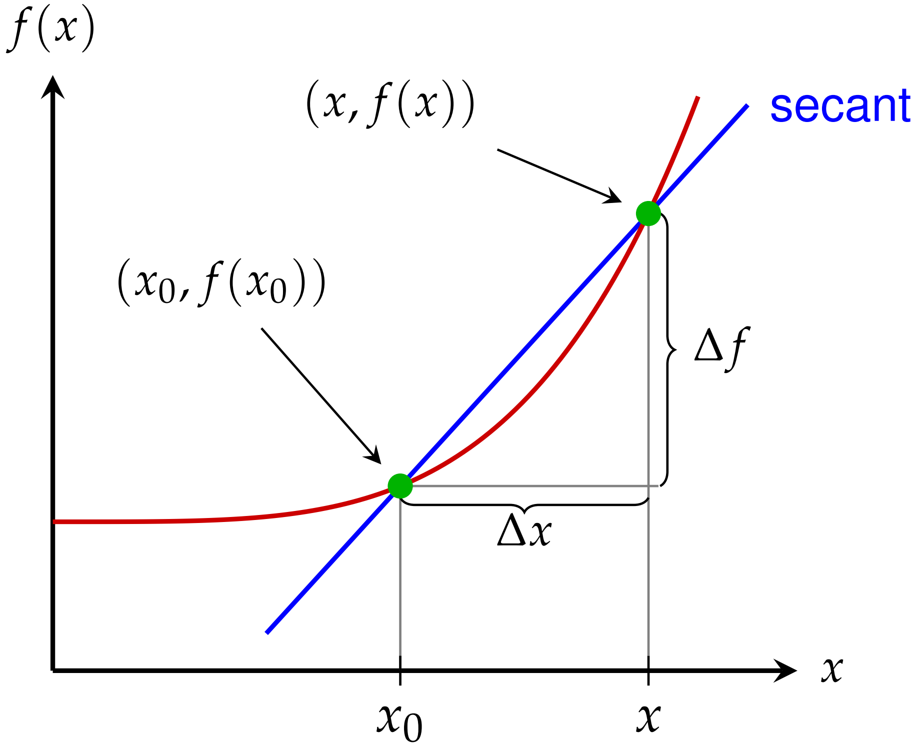
Derivatives

Difference Quotient*

Let $f: \mathbb{R} \to \mathbb{R}$ be some function. Then the ratio

$$\boxed{\frac{\Delta f}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \frac{f(x) - f(x_0)}{x - x_0}}$$

is called **difference quotient**.



Differential Quotient*

If the *limit*

$$\boxed{\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}}$$

exists, then function f is called **differentiable** at x_0. This limit is then called **differential quotient** or **(first) derivative** of function f at x_0.

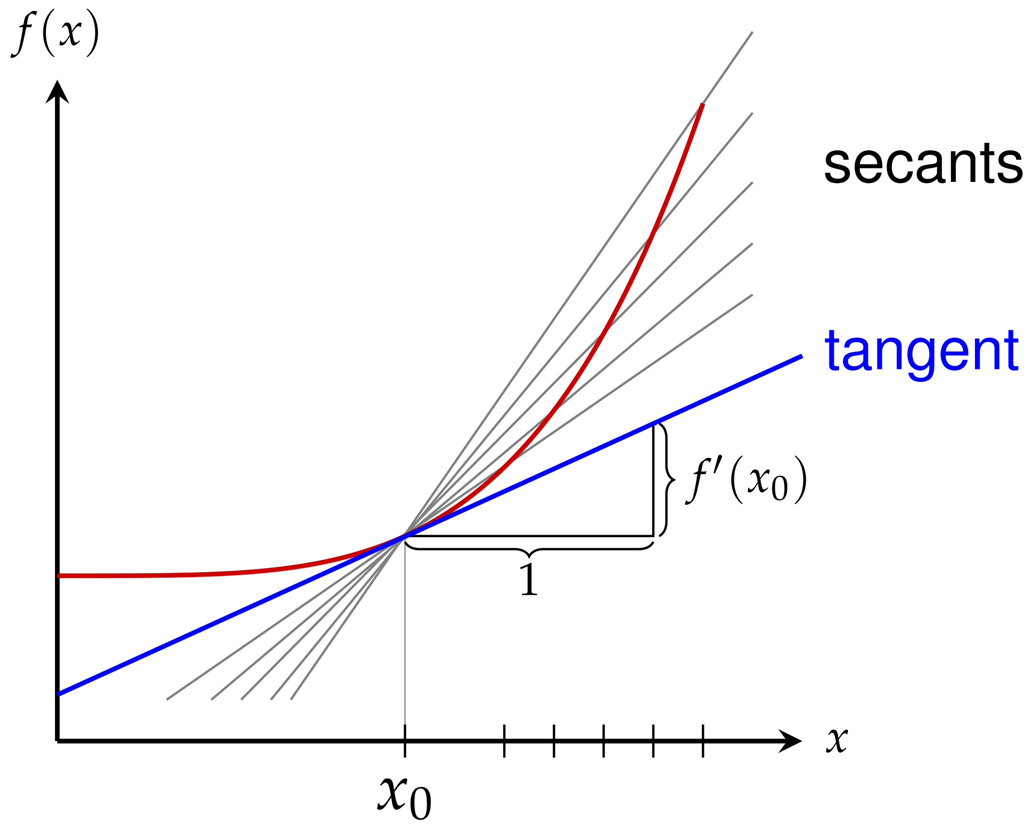
We write

$$\boxed{f'(x_0) \qquad \text{or} \qquad \left.\frac{df}{dx}\right|_{x=x_0}}$$

Function f is called *differentiable*, if it is differentiable at each point of its domain.

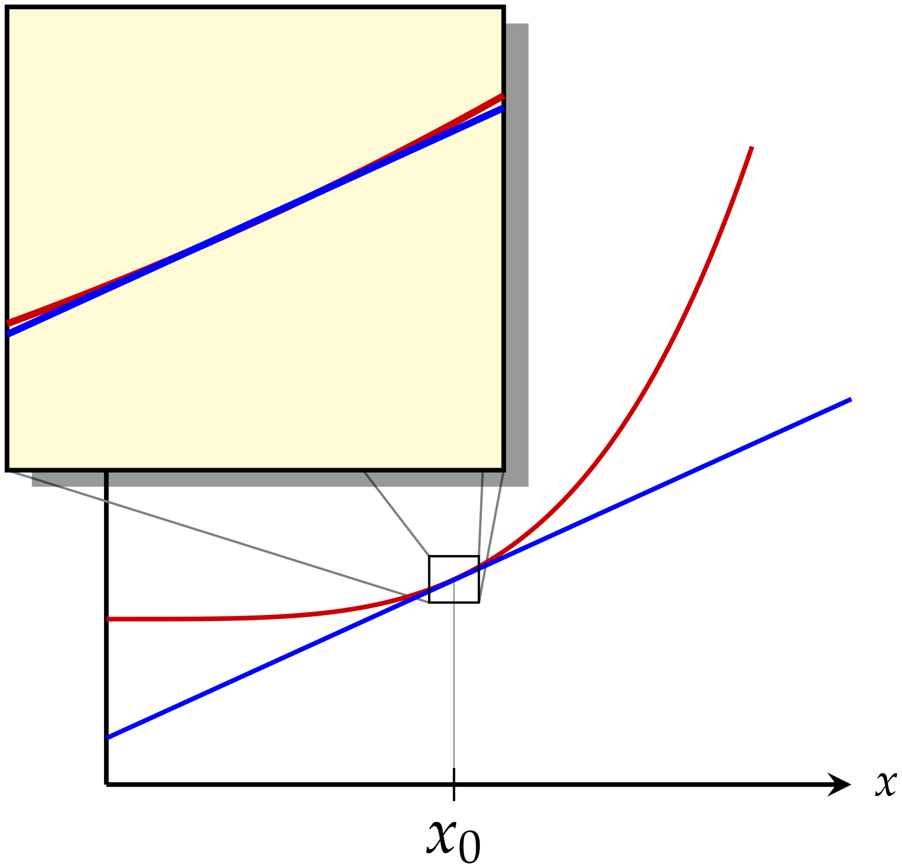
Slope of Tangent*

- The differential quotient gives the *slope of the tangent* to the graph of function $f(x)$ at x_0.



Marginal Function*

- Instantaneous change of function f.
- “Marginal function” (as in *marginal utility*)



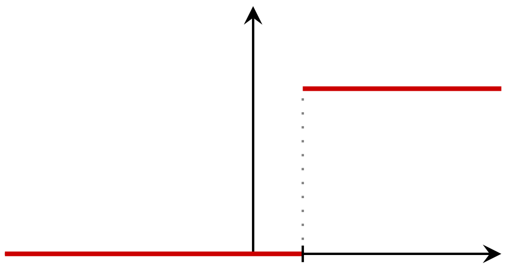
Existence of Differential Quotient*

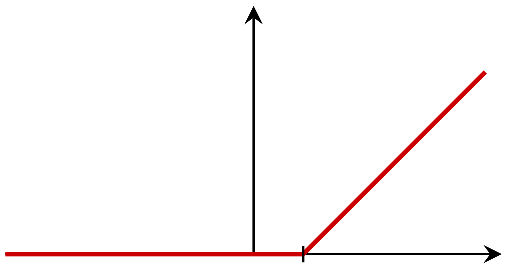
Function f is differentiable at all points, where we can draw the tangent (with finite slope) uniquely to the graph.

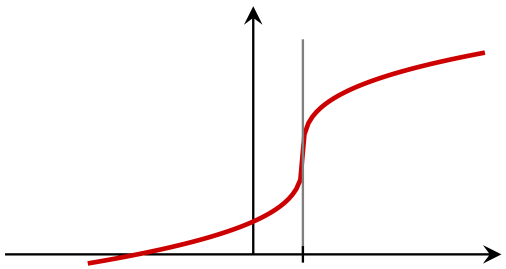
Function f is *not* differentiable at all points where this is not possible.

In particular these are

- jump discontinuities
- “kinks” in the graph of the function
- vertical tangents







Computation of the Differential Quotient*

We can compute a differential quotient by determining the limit of the difference quotient.

Let $f(x) = x^2$. The we find for the first derivative

$$\begin{aligned} f'(x_0) &= \lim_{h\to 0} \frac{(x_0+h)^2 - {x_0}^2}{h} \\ &= \lim_{h\to 0} \frac{{x_0}^2 + 2\,x_0\,h + h^2 - {x_0}^2}{h} \\ &= \lim_{h\to 0} \frac{2\,x_0\,h + h^2}{h} = \lim_{h\to 0} (2\,x_0 + h) \\ &= 2\,x_0 \end{aligned}$$

Derivative of a Function*

Function

$$\boxed{f'\colon D \to \mathbb{R},\ x \mapsto f'(x) = \left.\frac{df}{dx}\right|_x}$$

is called the **first derivative** of function f.
Its domain D is the set of all points where the differential quotient (i.e., the limit of the difference quotient) exists.

Derivatives of Elementary Functions*

$f(x)$	$f'(x)$
c	0
x^α	$\alpha \cdot x^{\alpha-1}$
e^x	e^x
$\ln(x)$	$\dfrac{1}{x}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$

Computation Rules for Derivatives*

- $(c \cdot f(x))' = c \cdot f'(x)$
- $(f(x) + g(x))' = f'(x) + g'(x)$ *Summation rule*
- $(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$ *Product rule*
- $(f(g(x)))' = f'(g(x)) \cdot g'(x)$ *Chain rule*
- $\left(\dfrac{f(x)}{g(x)}\right)' = \dfrac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{(g(x))^2}$ *Quotient rule*

Example – Computation Rules for Derivatives*

$$\big(3\,x^3 + 2\,x - 4\big)' = 3\cdot 3\cdot x^2 + 2\cdot 1 - 0 = 9\,x^2 + 2$$

$$\big(e^x \cdot x^2\big)' = \ (e^x)' \cdot x^2 + e^x \cdot \big(x^2\big)' = e^x \cdot x^2 + e^x \cdot 2\,x$$

$$\big((3\,x^2+1)^2\big)' = 2\,(3\,x^2+1)\cdot 6\,x$$

$$\big(\sqrt{x}\big)' = \ \left(x^{\frac{1}{2}}\right)' = \tfrac{1}{2}\cdot x^{-\frac{1}{2}} = \tfrac{1}{2\sqrt{x}}$$

$$(a^x)' = \ \left(e^{\ln(a)\cdot x}\right)' = e^{\ln(a)\cdot x}\cdot \ln(a) = a^x\ \ln(a)$$

$$\left(\frac{1+x^2}{1-x^3}\right)' = \frac{2x\cdot(1-x^3)-(1+x^2)\cdot 3x^2}{(1-x^3)^2}$$

Higher Order Derivatives*

We can compute derivatives of the derivative of a function.

Thus we obtain the

- **second derivative** $f''(x)$ of function f,
- **third derivative** $f'''(x)$, etc.,
- $\boldsymbol{n}$**-th derivative** $f^{(n)}(x)$.

Other notations:

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

Example – Higher Order Derivatives*

The first five derivatives of function

$$f(x) = x^4 + 2x^2 + 5x - 3$$

are

$$\begin{aligned}
f'(x) &= (x^4 + 2x^2 + 5x - 3)' = 4x^3 + 4x + 5 \\
f''(x) &= (4x^3 + 4x + 5)' = 12x^2 + 4 \\
f'''(x) &= (12x^2 + 4)' = 24x \\
f^{\mathsf{IV}}(x) &= (24x)' = 24 \\
f^{\mathsf{V}}(x) &= 0
\end{aligned}$$

Marginal Change*

We can estimate the derivative $f'(x_0)$ approximately by means of the difference quotient with *small* change Δx:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \approx \frac{\Delta f}{\Delta x}$$

Vice verse we can estimate the change Δf of f for *small* changes Δx approximately by the first derivative of f:

$$\Delta f = f(x_0 + \Delta x) - f(x_0) \approx f'(x_0) \cdot \Delta x$$

Beware:

- $f'(x_0) \cdot \Delta x$ is a *linear function* in Δx.
- It is the *best possible* approximation of f by a linear function *around* x_0.
- This approximation is useful only for *"small"* values of Δx.

Differential*

The approximation

$$\Delta f = f(x_0 + \Delta x) - f(x_0) \approx f'(x_0) \cdot \Delta x$$

becomes exact if Δx (and thus Δf) becomes *infinitesimally small*.
We then write dx and df instead of Δx and Δf, resp.

$$\boxed{df = f'(x_0)\, dx}$$

Symbols df and dx are called the **differentials** of function f and the independent variable x, resp.

Differential*

Differential df can be seen as a linear function in dx.
We can use it to compute f approximately around x_0.

$$f(x_0 + dx) \approx f(x_0) + df$$

Let $f(x) = e^x$.

Differential of f at point $x_0 = 1$:
$df = f'(1)\,dx = e^1\,dx$

Approximation of $f(1.1)$ by means of this differential:
$\Delta x = (x_0 + dx) - x_0 = 1.1 - 1 = 0.1$

$f(1.1) \approx f(1) + df = e + e \cdot 0.1 \approx 2.99$

Exact value: $f(1.1) = 3.004166\ldots$

Elasticity*

The first derivative of a function gives *absolute* rate of change of f at x_0. Hence it depends on the scales used for argument and function values.

However, often *relative* rates of change are more appropriate.

We obtain *scale invariance* and *relative* rate of changes by

$$\frac{\text{change of function value relative to value of function}}{\text{change of argument relative to value of argument}}$$

and thus

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

Elasticity*

Expression

$$\boxed{\varepsilon_f(x) = x \cdot \frac{f'(x)}{f(x)}}$$

is called the **elasticity** of f at point x.

Let $f(x) = 3\,e^{2x}$. Then

$$\varepsilon_f(x) = x \cdot \frac{f'(x)}{f(x)} = x \cdot \frac{6\,e^{2x}}{3\,e^{2x}} = 2\,x$$

Let $f(x) = \beta\,x^{\alpha}$. Then

$$\varepsilon_f(x) = x \cdot \frac{f'(x)}{f(x)} = x \cdot \frac{\beta\,\alpha\,x^{\alpha-1}}{\beta\,x^{\alpha}} = \alpha$$

Elasticity II*

The relative rate of change of f can be expressed as

$$\ln(f(x))' = \frac{f'(x)}{f(x)}$$

What happens if we compute the derivative of $\ln(f(x))$ w.r.t. $\ln(x)$?

Let $\quad v = \ln(x) \quad \Leftrightarrow \quad x = e^v$

Derivation by means of the chain rule yields:

$$\frac{d(\ln(f(x)))}{d(\ln(x))} = \frac{d(\ln(f(e^v)))}{dv} = \frac{f'(e^v)}{f(e^v)}\, e^v = \frac{f'(x)}{f(x)}\, x = \varepsilon_f(x)$$

$$\boxed{\varepsilon_f(x) = \frac{d(\ln(f(x)))}{d(\ln(x))}}$$

Elasticity II*

We can use the chain rule *formally* in the following way:

Let

- $u = \ln(y)$,
- $y = f(x)$,
- $x = e^v \quad \Leftrightarrow \quad v = \ln(x)$

Then we find

$$\frac{d(\ln f)}{d(\ln x)} = \frac{du}{dv} = \frac{du}{dy} \cdot \frac{dy}{dx} \cdot \frac{dx}{dv} = \frac{1}{y} \cdot f'(x) \cdot e^v = \frac{f'(x)}{f(x)}\, x$$

Elastic Functions*

A Function f is called

- **elastic** in x, if $\left| \varepsilon_f(x) \right| > 1$
- **1-elastic** in x, if $\left| \varepsilon_f(x) \right| = 1$
- **inelastic** in x, if $\left| \varepsilon_f(x) \right| < 1$

For elastic functions we then have:
The value of the function changes *relatively* faster than the value of the argument.

Function $f(x) = 3\,e^{2x}$ is $[\, \varepsilon_f(x) = 2\,x \,]$

- 1-elastic, for $x = -\frac{1}{2}$ and $x = \frac{1}{2}$;
- inelastic, for $-\frac{1}{2} < x < \frac{1}{2}$;
- elastic, for $x < -\frac{1}{2}$ or $x > \frac{1}{2}$.

Source of Errors

Beware!
Function f is elastic if the **absolute value** of the *elasticity* is greater than 1.

Elastic Demand*

Let $q(p)$ be an *elastic* demand function, where p is the price.
We have: $p > 0$, $q > 0$, and $q' < 0$ (q is decreasing). Hence

$$\varepsilon_q(p) = p \cdot \frac{q'(p)}{q(p)} < -1$$

What happens to the revenue (= price $\times$ selling)?

$$\begin{aligned} u'(p) &= (p \cdot q(p))' = 1 \cdot q(p) + p \cdot q'(p) \\ &= q(p) \cdot (1 + \underbrace{p \cdot \frac{q'(p)}{q(p)}}_{=\varepsilon_q < -1}) \\ &< 0 \end{aligned}$$

In other words, the revenue decreases if we raise prices.

Partial Derivative*

We investigate the rate of change of function $f(x_1, \dots, x_n)$, when variable x_i changes and the other variables remain fixed.
Limit

$$\boxed{\frac{\partial f}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(\dots, x_i + \Delta x_i, \dots) - f(\dots, x_i, \dots)}{\Delta x_i}}$$

is called the (first) **partial derivative** of f w.r.t. x_i.

Other notations for partial derivative $\frac{\partial f}{\partial x_i}$:

- $f_{x_i}(\mathbf{x})$ (derivative w.r.t. variable x_i)
- $f_i(\mathbf{x})$ (derivative w.r.t. the i-th variable)
- $f'_i(\mathbf{x})$ (i-th component of the gradient)

Computation of Partial Derivatives*

We obtain partial derivatives $\frac{\partial f}{\partial x_i}$ by applying the rules for *univariate* functions for variable x_i while we treat *all other* variables *as constants*.

First partial derivatives of

$$f(x_1, x_2) = \sin(2\,x_1) \cdot \cos(x_2)$$

$$f_{x_1} = 2 \cdot \cos(2\,x_1) \cdot \underbrace{\cos(x_2)}_{\text{treated as constant}}$$

$$f_{x_2} = \underbrace{\sin(2\,x_1)}_{\text{treated as constant}} \cdot (-\sin(x_2))$$

Higher Order Partial Derivatives*

We can compute partial derivatives of partial derivatives analogously to their univariate counterparts and obtain **higher order partial derivatives**:

$$\boxed{\frac{\partial^2 f}{\partial x_k \partial x_i}(\mathbf{x}) \quad \text{and} \quad \frac{\partial^2 f}{\partial x_i^2}(\mathbf{x})}$$

Other notations for partial derivative $\frac{\partial^2 f}{\partial x_k \partial x_i}(\mathbf{x})$:

- $f_{x_i x_k}(\mathbf{x})$ (derivative w.r.t. variables x_i and x_k)
- $f_{ik}(\mathbf{x})$ (derivative w.r.t. the i-th and k-th variable)
- $f''_{ik}(\mathbf{x})$ (component of the Hessian matrix with index ik)

Higher Order Partial Derivatives*

If all second order partial derivatives exists and are *continuous*, then the order of differentiation does not matter (Schwarz's theorem):

$$\boxed{\frac{\partial^2 f}{\partial x_k \partial x_i}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_i \partial x_k}(\mathbf{x})}$$

Remark: Practically all differentiable functions in economic models have this property.

Example – Higher Order Partial Derivatives*

Compute the first and second order partial derivatives of

$$f(x,y) = x^2 + 3\,x\,y$$

First order partial derivatives:

$$f_x = 2\,x + 3\,y \qquad f_y = 0 + 3\,x$$

Second order partial derivatives:

$$\begin{aligned} f_{xx} &= 2 & \qquad f_{xy} &= \boxed{3} \\ f_{yx} &= \boxed{3} & \qquad f_{yy} &= 0 \end{aligned}$$

Gradient

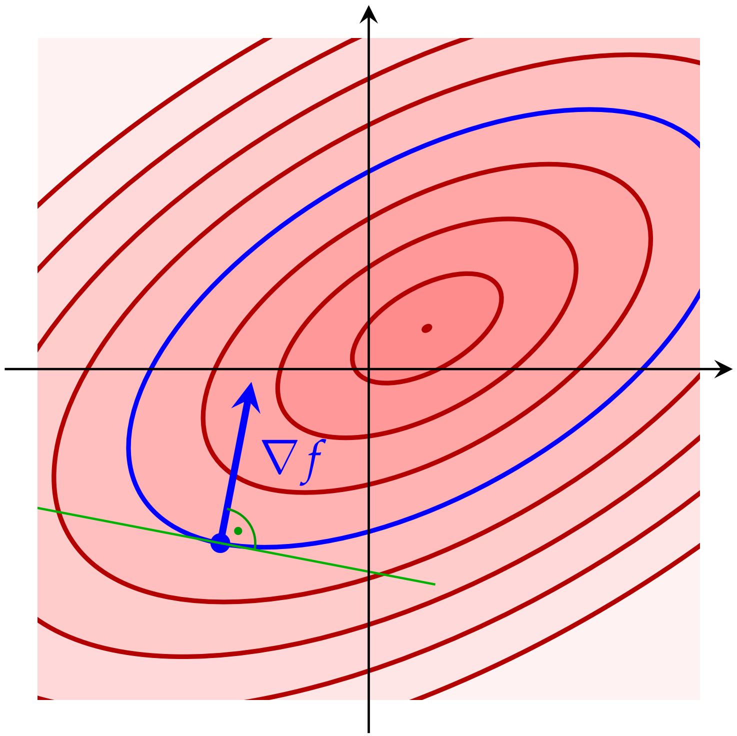
We collect all *first order partial derivatives* into a (row) vector which is called the **gradient** at point $\mathbf{x}$.

$$\boxed{\nabla f(\mathbf{x}) = (f_{x_1}(\mathbf{x}), \dots, f_{x_n}(\mathbf{x}))}$$

- read: “gradient of f” or “nabla f”.
- Other notation: $\quad f'(\mathbf{x})$
- Alternatively the gradient can also be a column vector.
- The gradient is the analog of the first derivative of univariate functions.

Properties of the Gradient

- The gradient of f always points in the direction of *steepest ascent*.
- Its length is equal to the slope at this point.
- The gradient is *normal* (i.e. in right angle) to the corresponding *contour line* (level set).



Example – Gradient

Compute the gradient of

$$f(x,y) = x^2 + 3\,x\,y$$

at point $\mathbf{x} = (3,2)$.

$$\begin{aligned} f_x &= 2\,x + 3\,y \\ f_y &= 0 + 3\,x \end{aligned}$$

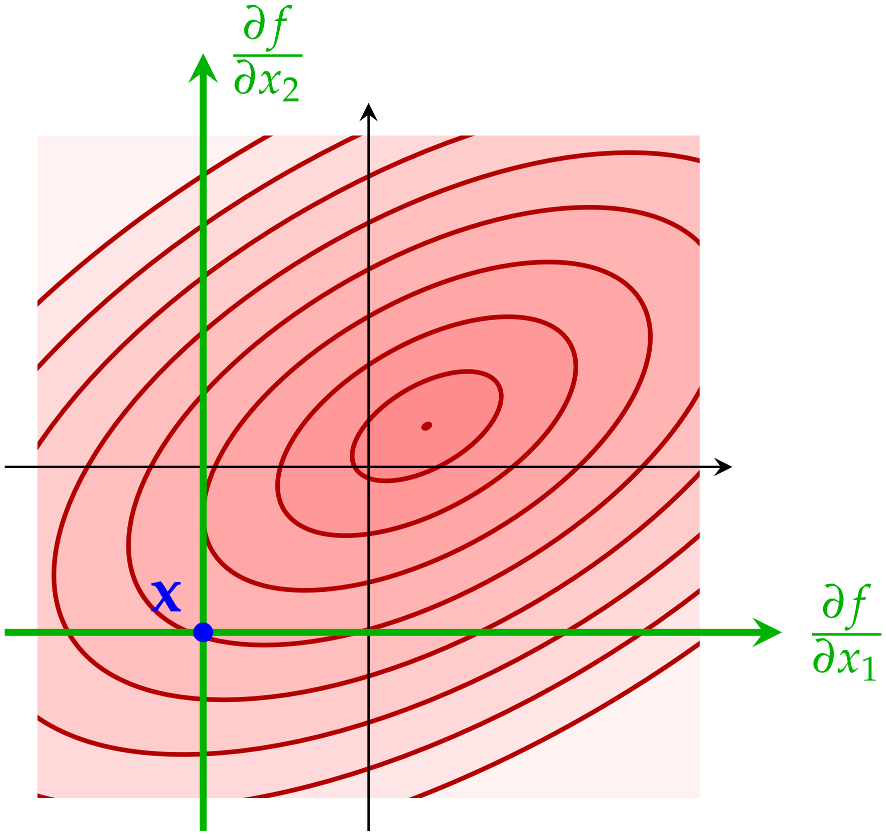
$$\nabla f(\mathbf{x}) = (2x + 3y,\ 3x)$$

$$\nabla f(3,2) = (12, 9)$$

Directional Derivative

We obtain partial derivative $\frac{\partial f}{\partial x_i}$ by differentiating the univariate function $g(t) = f(x_1, \ldots, x_i + t, \ldots, x_n) = f(\mathbf{x} + t \cdot \mathbf{h})$ with $\mathbf{h} = \mathbf{e}_i$ at point $t = 0$:

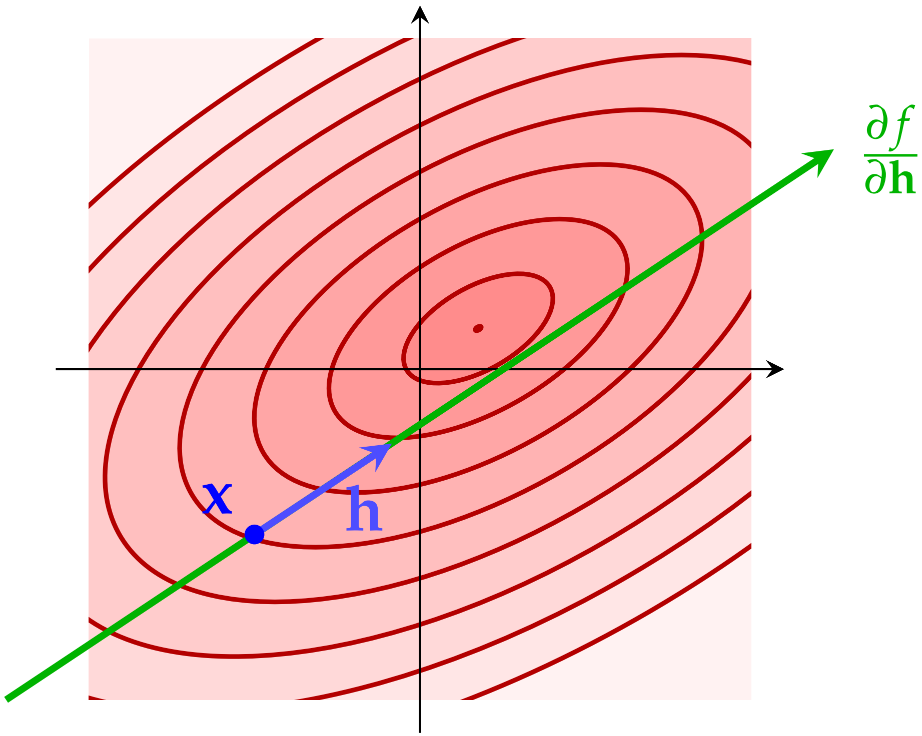
$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \left.\frac{dg}{dt}\right|_{t=0} = \left.\frac{d}{dt} f(\mathbf{x} + t \cdot \mathbf{h})\right|_{t=0}$$



Directional Derivative

Generalization:
We obtain the **directional derivative** $\frac{\partial f}{\partial \mathbf{h}}$ along $\mathbf{h}$ with length 1 by differentiating the univariate function $g(t) = f(\mathbf{x} + t \cdot \mathbf{h})$ at point $t = 0$:

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}) = \left.\frac{dg}{dt}\right|_{t=0} = \left.\frac{d}{dt} f(\mathbf{x} + t \cdot \mathbf{h})\right|_{t=0}$$



The directional derivative describes the change of f, if we move $\mathbf{x}$ in direction $\mathbf{h}$.

Directional Derivative

We have (for $\|\mathbf{h}\| = 1$):

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}) = f_{x_1}(\mathbf{x}) \cdot h_1 + \cdots + f_{x_n}(\mathbf{x}) \cdot h_n = \nabla f(\mathbf{x}) \cdot \mathbf{h}$$

If $\mathbf{h}$ does not have norm 1, we first have to normalize first:

$$\boxed{\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \frac{\mathbf{h}}{\|\mathbf{h}\|}}$$

Example – Directional Derivative

Compute the directional derivative of

$$f(x_1, x_2) = x_1^2 + 3\,x_1\,x_2$$

along $\mathbf{h} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ at $\mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

Norm of $\mathbf{h}$:

$$\|\mathbf{h}\| = \sqrt{\mathbf{h}^\mathsf{T}\,\mathbf{h}} = \sqrt{1^2 + (-2)^2} = \sqrt{5}$$

Directional derivative:

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \frac{\mathbf{h}}{\|\mathbf{h}\|} = \frac{1}{\sqrt{5}}\,(12, 9) \cdot \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -\frac{6}{\sqrt{5}}$$

Total Differential

We want to approximate a function f by some linear function such that the approximation error is as small as possible:

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) \approx f_{x_1}(\mathbf{x})\,h_1 + \ldots + f_{x_n}(\mathbf{x})\,h_n = \nabla f(\mathbf{x}) \cdot \mathbf{h}$$

The approximation becomes exact if $\mathbf{h}$ (and thus Δf) becomes *infinitesimally small*.

The *linear function*

$$\boxed{df = f_{x_1}(\mathbf{x})\,dx_1 + \ldots + f_{x_n}(\mathbf{x})\,dx_n = \sum_{i=1}^{n} f_{x_i}\,dx_i = \nabla f(\mathbf{x}) \cdot d\mathbf{x}}$$

is called the **total Differential** of f at $\mathbf{x}$.

Example – Total Differential

Compute the total differential of

$$f(x_1, x_2) = x_1^2 + 3\,x_1\,x_2$$

at $\mathbf{x} = (3, 2)$.

$$df = f_{x_1}(3, 2)\,dx_1 + f_{x_2}(3, 2)\,dx_2 = 12\,dx_1 + 9\,dx_2$$

Approximation of $f(3.1, 1.8)$ by means of the total differential:

$$\begin{aligned} f(3.1, 1.8) &\approx f(3; 2) + df \\ &= 27 + 12 \cdot 0.1 + 9 \cdot (-0.2) = 26.40 \end{aligned}$$

Exact value: $f(3.1, 1.8) = 26.35$

$$\mathbf{h} = (\mathbf{x} + \mathbf{h}) - \mathbf{x} = \begin{pmatrix} 3.1 \\ 1.8 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ -0.2 \end{pmatrix}$$

Hessian Matrix

Let $f(\mathbf{x}) = f(x_1, \dots, x_n)$ be two times differentiable. Then matrix

$$\boxed{\mathbf{H}_f(\mathbf{x}) = \begin{pmatrix} f_{x_1x_1}(\mathbf{x}) & f_{x_1x_2}(\mathbf{x}) & \dots & f_{x_1x_n}(\mathbf{x}) \\ f_{x_2x_1}(\mathbf{x}) & f_{x_2x_2}(\mathbf{x}) & \dots & f_{x_2x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(\mathbf{x}) & f_{x_nx_2}(\mathbf{x}) & \dots & f_{x_nx_n}(\mathbf{x}) \end{pmatrix}}$$

is called the **Hessian matrix** of f at $\mathbf{x}$.

- The Hessian matrix is symmetric, i.e., $f_{x_ix_k}(\mathbf{x}) = f_{x_kx_i}(\mathbf{x})$.
- Other notation: $f''(\mathbf{x})$
- The Hessian matrix is the analog of the second derivative of univariate functions.

Example – Hessian Matrix

Compute the Hessian matrix of

$$f(x,y) = x^2 + 3\,x\,y$$

at point $\mathbf{x} = (1,2)$.

Second order partial derivatives:

$$\begin{aligned} f_{xx} &= 2 & \qquad f_{xy} &= 3 \\ f_{yx} &= 3 & \qquad f_{yy} &= 0 \end{aligned}$$

Hessian matrix:

$$\mathbf{H}_f(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix} = \mathbf{H}_f(1,2)$$

Differentiability

Theorem:
A function $f\colon \mathbb{R} \to \mathbb{R}$ is **differentiable** at x_0 if and only if there exists a linear map ℓ which approximates f in x_0 in an optimal way:

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

Obviously $\ell(h) = f'(x_0) \cdot h$.

Definition:
A function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $\mathbf{x}_0$ if there exists a linear map ℓ which approximates $\mathbf{f}$ in $\mathbf{x}_0$ in an optimal way:

$$\lim_{\mathbf{h}\to 0} \frac{\|(\mathbf{f}(\mathbf{x}_0+\mathbf{h}) - \mathbf{f}(\mathbf{x}_0)) - \ell(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$$

Function $\ell(\mathbf{h}) = \mathbf{J} \cdot \mathbf{h}$ is called the *total derivative* of $\mathbf{f}$.

Jacobian Matrix

Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m,\ \mathbf{x} \mapsto \mathbf{y} = \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}$

The $m \times n$ matrix

$$\boxed{D\mathbf{f}(\mathbf{x}_0) = \mathbf{f}'(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}}$$

is called the **Jacobian matrix** of $\mathbf{f}$ at point $\mathbf{x}_0$.

It is the generalization of *derivatives* (and gradients) for vector-valued functions.

Jacobian Matrix

For $f\colon \mathbb{R}^n \to \mathbb{R}$ the Jacobian matrix is the gradient of f:

$$Df(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)$$

For vector-valued functions the Jacobian matrix can be written as

$$D\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} \nabla f_1(\mathbf{x}_0) \\ \vdots \\ \nabla f_m(\mathbf{x}_0) \end{pmatrix}$$

Example – Jacobian Matrix

- $f(\mathbf{x}) = f(x_1, x_2) = \exp(-x_1^2 - x_2^2)$

$$D f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = \nabla f(\mathbf{x})$$
$$= \left(-2\,x_1\,\exp(-x_1^2 - x_2^2), -2\,x_2\,\exp(-x_1^2 - x_2^2)\right)$$

- $\mathbf{f}(\mathbf{x}) = \mathbf{f}(x_1, x_2) = \begin{pmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2^2 \\ x_1^2 - x_2^2 \end{pmatrix}$

$$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2\,x_1 & 2\,x_2 \\ 2\,x_1 & -2\,x_2 \end{pmatrix}$$

- $\mathbf{s}(t) = \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}$

$$D\mathbf{s}(t) = \begin{pmatrix} \frac{ds_1}{dt} \\ \frac{ds_2}{dt} \end{pmatrix} = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}$$

Chain Rule

Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g}\colon \mathbb{R}^m \to \mathbb{R}^k$. Then

$$\boxed{(\mathbf{g} \circ \mathbf{f})'(\mathbf{x}) = \mathbf{g}'(\mathbf{f}(\mathbf{x})) \cdot \mathbf{f}'(\mathbf{x})}$$

$$\mathbf{f}(x,y) = \begin{pmatrix} e^x \\ e^y \end{pmatrix} \qquad \mathbf{g}(x,y) = \begin{pmatrix} x^2 + y^2 \\ x^2 - y^2 \end{pmatrix}$$

$$\mathbf{f}'(x,y) = \begin{pmatrix} e^x & 0 \\ 0 & e^y \end{pmatrix} \qquad \mathbf{g}'(x,y) = \begin{pmatrix} 2\,x & 2\,y \\ 2\,x & -2\,y \end{pmatrix}$$

$$(\mathbf{g} \circ \mathbf{f})'(\mathbf{x}) = \mathbf{g}'(\mathbf{f}(\mathbf{x})) \cdot \mathbf{f}'(\mathbf{x}) = \begin{pmatrix} 2\,e^x & 2\,e^y \\ 2\,e^x & -2\,e^y \end{pmatrix} \cdot \begin{pmatrix} e^x & 0 \\ 0 & e^y \end{pmatrix}$$

$$= \begin{pmatrix} 2\,e^{2x} & 2\,e^{2y} \\ 2\,e^{2x} & -2\,e^{2y} \end{pmatrix}$$

Example – Directional Derivative

We can derive the formula for the directional derivative of $f\colon \mathbb{R}^n \to \mathbb{R}$ along $\mathbf{h}$ (with $\|\mathbf{h}\| = 1$) at $\mathbf{x}_0$ by means of the chain rule:

Let $\mathbf{s}(t)$ be a path through $\mathbf{x}_0$ along $\mathbf{h}$, i.e.,

$$\mathbf{s}\colon \mathbb{R} \to \mathbb{R}^n,\ t \mapsto \mathbf{x}_0 + t\mathbf{h}\ .$$

Then

$$\begin{aligned} f'(\mathbf{s}(0)) &= f'(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \\ \mathbf{s}'(0) &= \mathbf{h} \end{aligned}$$

and hence

$$\frac{\partial f}{\partial \mathbf{h}} = (f \circ \mathbf{s})'(0) = f'(\mathbf{s}(0)) \cdot \mathbf{s}'(0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{h}\ .$$

Example – Indirect Dependency

Let $f(x_1, x_2, t)$ where $x_1(t)$ and $x_2(t)$ also depend on t.
What is the total derivative of f w.r.t. t?

Chain rule:

Let $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^3,\ t \mapsto \begin{pmatrix} x_1(t) \\ x_2(t) \\ t \end{pmatrix}$

$$
\begin{aligned}
\frac{df}{dt} &= (f \circ \mathbf{x})'(t) = f'(\mathbf{x}(t)) \cdot \mathbf{x}'(t) \\
&= \nabla f(\mathbf{x}(t)) \cdot \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ 1 \end{pmatrix} = (f_{x_1}(\mathbf{x}(t)), f_{x_2}(\mathbf{x}(t)), f_t(\mathbf{x}(t)) \cdot \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ 1 \end{pmatrix} \\
&= f_{x_1}(\mathbf{x}(t)) \cdot x_1'(t) + f_{x_2}(\mathbf{x}(t)) \cdot x_2'(t) + f_t(\mathbf{x}(t)) \\
&= f_{x_1}(x_1, x_2, t) \cdot x_1'(t) + f_{x_2}(x_1, x_2, t) \cdot x_2'(t) + f_t(x_1, x_2, t)
\end{aligned}
$$

L'Hôpital's Rule

Suppose we want to compute

$$\lim_{x \to x_0} \frac{f(x)}{g(x)}$$

and find

$$\lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0 \qquad \text{(or } = \pm\infty\text{)}$$

However, expressions like $\frac{0}{0}$ or $\frac{\infty}{\infty}$ are not defined.

(You must not reduce the fraction by 0 or ∞!)

L’Hôpital’s Rule

If $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} g(x) = 0$ (or $= \infty$ or $= -\infty$), then

$$\boxed{\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}}$$

Assumption: f and g are differentiable in x_0.

This formula is called **l’Hôpital’s rule** (also spelled as *l’Hospital’s rule*).

Example – L'Hôpital's Rule

$$\lim_{x\to 2} \frac{x^3 - 7\,x + 6}{x^2 - x - 2} = \lim_{x\to 2} \frac{3\,x^2 - 7}{2\,x - 1} = \frac{5}{3}$$

$$\lim_{x\to\infty} \frac{\ln\, x}{x^2} = \lim_{x\to\infty} \frac{\frac{1}{x}}{2\,x} = \lim_{x\to\infty} \frac{1}{2\,x^2} = 0$$

$$\lim_{x\to 0} \frac{x - \ln(1 + x)}{x^2} = \lim_{x\to 0} \frac{1 - (1 + x)^{-1}}{2\,x} = \lim_{x\to 0} \frac{(1 + x)^{-2}}{2} = \frac{1}{2}$$

Example – L'Hôpital's Rule

L'Hôpital's rule can be applied iteratively:

$$\lim_{x \to 0} \frac{e^x - x - 1}{x^2} = \lim_{x \to 0} \frac{e^x - 1}{2x} = \lim_{x \to 0} \frac{e^x}{2} = \frac{1}{2}$$

Summary

- difference quotient and differential quotient
- differential quotient and derivative
- derivatives of elementary functions
- differentiation rules
- higher order derivatives
- total differential
- elasticity
- partial derivatives
- gradient and Hessian matrix
- Jacobian matrix and chain rule
- l’Hôpital’s rule