Differentiability in $\mathbb{R}^n$

In this subsection we consider the total differentiability, a stronger requirement than partial differentiability.

Moreover, we consider expansions of functions of several variables.

Applications:

- Linear approximation of functions, tangent planes
- Taylor series for several variables

Here we consider:

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

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f_1(\mathbf{x}) \\ \dots \\ f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \dots \\ f_m(x_1, \dots, x_n) \end{pmatrix}$$

(Total) Differentiability

Definition (Total differentiability)

$f : \mathbb{R}^n \supseteq D \to \mathbb{R}^m$ is called **totally differentiable** (or linearly approximable) in $\mathbf{x}_0 \in D$, D open (i.e. an open set)

if there exists a matrix $A \in \mathbb{R}^{m \times n}$,
such that for all $\mathbf{x}$ in a neighbourhood $U(\mathbf{x}_0)$, $\mathbf{x}_0 \in D$,

$$\mathbb{R}^m \ni \quad f(\mathbf{x}) = f(\mathbf{x}_0) + \underbrace{A}_{\in \mathbb{R}^{m\times n}} \cdot \underbrace{(\mathbf{x} - \mathbf{x}_0)}_{\in \mathbb{R}^{n\times 1}} + R(\mathbf{x} - \mathbf{x}_0),$$

where

$$R : \mathbb{R}^n \supseteq U(\mathbf{x}_0) \to \mathbb{R}^m \text{ with } \lim_{\|\mathbf{x}-\mathbf{x}_0\| \to 0} \frac{R(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

f is called totally differentiable in E, if this holds for all $\mathbf{x} \in E \subseteq D$.

Jacobi Matrix

If f is totally differentiable, then

- f is continuous
- all f_i, $i = 1, \dots, m$, are continuously partially differentiable (this is even equivalent, but cont. deriv. is important!)
- the matrix A is uniquely determined by the so-called **Jacobi matrix** (Jacobian)

• not the Hessian (a 2nd derivative)

• in general not symmetric

$$J_f(\mathbf{x}) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \frac{\partial f_i}{\partial x_j}(\mathbf{x}) & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

$= \nabla f_1(\mathbf{x})^T$

Special case $m = 1$: transposed gradient

We also write $D_f(\mathbf{x}) = J_f(\mathbf{x})$, indicating this is the general form of the derivative (as a matrix).

Linearization

$$\Delta f = f(x,y) - f(x_0, y_0)$$

Theorem (Tangent plane)

Let $D \subseteq \mathbb{R}^2$, $f : D \to \mathbb{R}$ totally differentiable and $(x_0, y_0) \in D$. D open.

Then the points (x, y, z) of the tangent plane in (x_0, y_0) are described by the equation

$$\underset{f(x,y)}{z} = f(x_0, y_0) + \nabla f(x_0, y_0)^\top \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}$$

$$= f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0) \cdot \underbrace{(x - x_0)}_{=\Delta x} + \frac{\partial f}{\partial y}(x_0, y_0) \cdot \underbrace{(y - y_0)}_{=\Delta y}$$

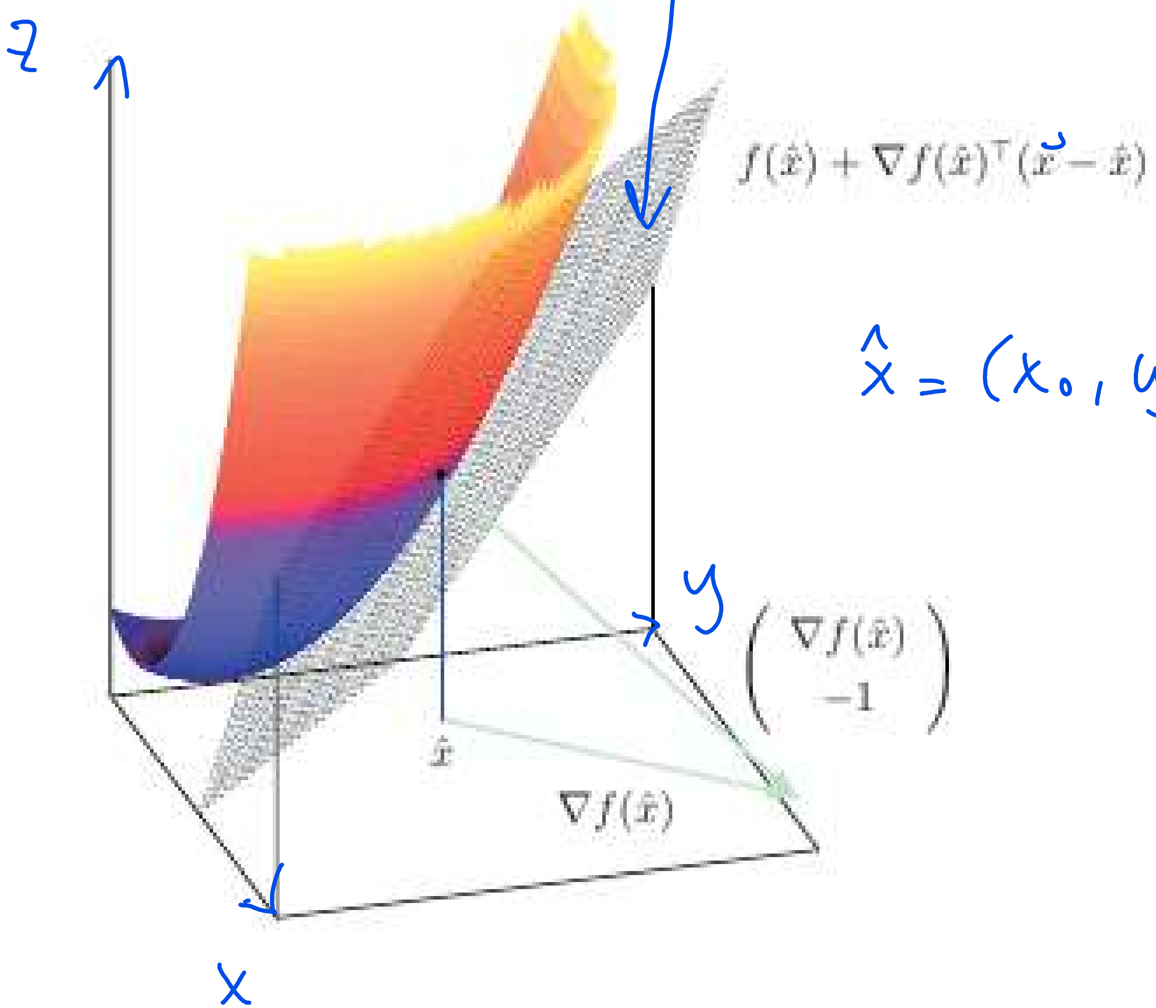
The tangent plane is locally the optimal (affine-)linear approximation for f. The plane consists out of all tangents.

In the general case the linear approximation reads

$$f(\mathbf{x}) \approx f(\hat{\mathbf{x}}) + D_f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}).$$

Tangent Plane as Linear Approximation

In 2d: tangent plane



$\hat{x} = (x_0, y_0)$

Tangent Plane: Example

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x,y) := \begin{pmatrix} xy^2 \\ 2x^2y^2 \end{pmatrix}$$

Linearization in $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$:

$$D_f(x,y) = \begin{pmatrix} y^2, & 2xy \\ 4xy^2, & 4x^2y \end{pmatrix} \overset{\substack{x=x_0 \\ y=y_0}}{=} \begin{pmatrix} 1 & 2 \\ 4, & 4 \end{pmatrix}$$

$$f(x,y) \approx f(1,1) + \begin{pmatrix} 1 & 2 \\ 4 & 4 \end{pmatrix} \cdot \begin{pmatrix} x-1 \\ y-1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} x-1+2y-2 \\ 4x-4+4y-4 \end{pmatrix} = \begin{pmatrix} x+2y-2 \\ 4x+4y-6 \end{pmatrix}$$

Chain Rule (Simple Case)

Let $\mathbf{x} : \mathbb{R} \supseteq (a, b) \to D$ differentiable in $t \in (a, b)$,
$f : \mathbb{R}^n \supseteq D \to \mathbb{R}$ (totally) diff.able in $\mathbf{s} := \mathbf{x}(t) \in D$, D open,

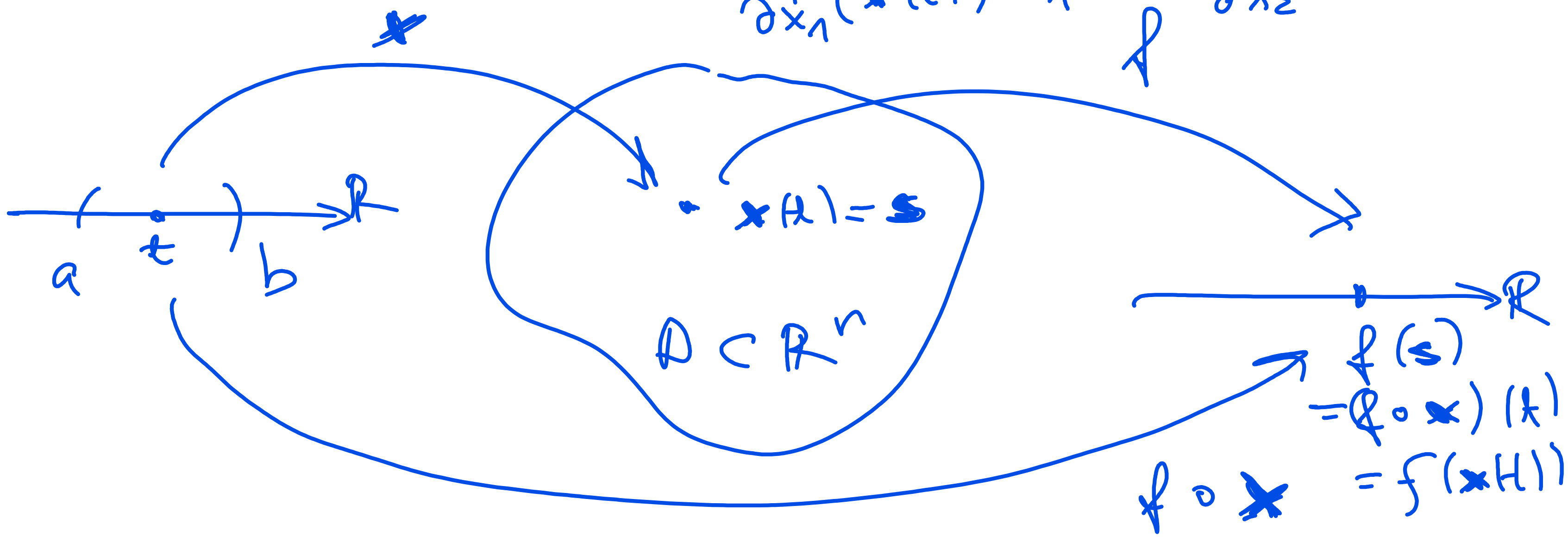
then

$$f \circ \mathbf{x} : (a, b) \to \mathbb{R},\ t \mapsto f(\mathbf{s}) = f(\mathbf{x}(t))$$

is totally differentiable in t and

$$\overset{1\times 1}{\frac{d}{dt} f(\mathbf{x}(t))} = \overset{1\times n}{\nabla f(\mathbf{x}(t))^\top}\ \overset{n\times 1}{\dot{\mathbf{x}}(t)}.$$

$$\overset{n=2}{=} \frac{\partial f}{\partial x_1}(\mathbf{x}(t))\, \dot{x}_1(t) + \frac{\partial f}{\partial x_2}(\mathbf{x}(t))\, \dot{x}_2(t)$$



Introduction
Power series
Differentiation in Higher Dimensions
Limits and Continuity in $\mathbb{R}^n$
Differentiability in $\mathbb{R}^n$
Optimization in $\mathbb{R}^n$
Derivatives
Integration in Higher Dimensions
Further Topics in Calculus
Summary - Outlook and Review

Chain Rule (General Case)

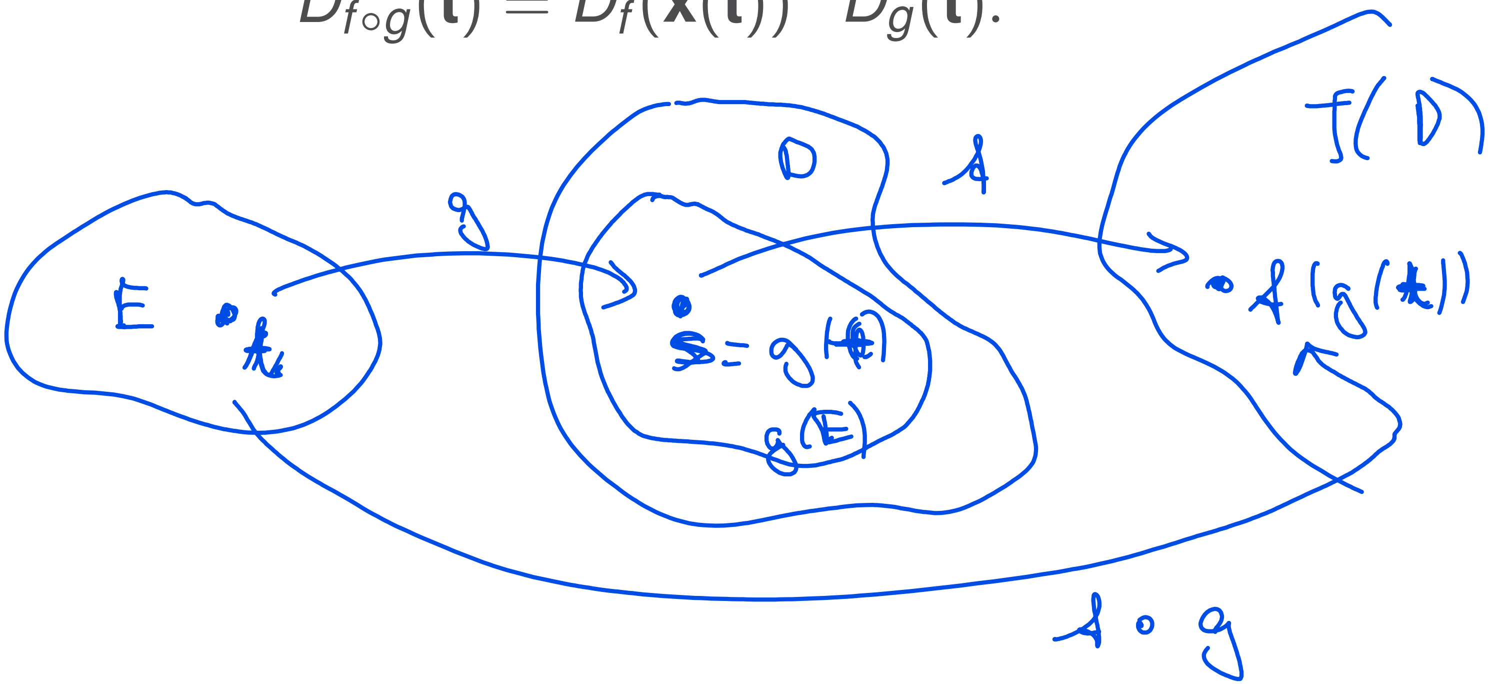
Let $g : \mathbb{R}^p \supseteq E \to D$ (totally) diff.able in $\mathbf{t} \in E$, E open, $f : \mathbb{R}^n \supseteq D \to \mathbb{R}^m$ (totally) diff.able in $\mathbf{s} := g(\mathbf{t}) \in D$, D open,

then

$$f \circ g : E \to \mathbb{R}^m, \mathbf{t} \mapsto f(\mathbf{s}) = f(g(\mathbf{t}))$$

is totally differentiable in $\mathbf{t}$ and

$$D_{f \circ g}(\mathbf{t}) = D_f(\mathbf{x}(\mathbf{t})) D_g(\mathbf{t}).$$



Application: Coordinate Transformation I

Example (Polar Coordinates in $\mathbb{R}^2$)

Cartesian coordinates $(x, y) \in \mathbb{R}^2$
Polar coordinates $(r, \phi) \in [0, \infty) \times (-\pi, \pi]$

Consider $f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto f(x, y)$ and
$F : [0, \infty) \times (-\pi, \pi], (r, \phi) \mapsto F(r, \phi) := f(r\cos(\phi), r\sin(\phi))$.

How do I transform the derivative when changing coordinates?

Let the change of coordinates be described by

$$g : \mathbb{R}^2 \to [0, \infty) \times (-\pi, \pi], \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} r\cos(\phi) \\ r\sin(\phi) \end{pmatrix}.$$

By the chain rule, we compute $D_F = D_f \cdot D_g = \ldots$

Application: Coordinate Transformation II

$$\underset{1\times 2}{D_F} = \underset{1\times 2}{D_f} \cdot \underset{2\times 2}{D_g} =$$

$$= \left(\frac{\partial f}{\partial x}(r\cos\varphi, r\sin\varphi), \frac{\partial f}{\partial y}(r\cos\varphi, r\sin\varphi) \right) \cdot \begin{pmatrix} \cos\varphi, & -r\sin\varphi \\ \sin\varphi, & r\cos\varphi \end{pmatrix}$$

$$= \Big(\frac{\partial f}{\partial x}(r\cos\varphi, r\sin\varphi) \cdot \cos\varphi + \frac{\partial f}{\partial y}(r\cos\varphi, r\sin\varphi) \sin\varphi,$$

$$- \frac{\partial f}{\partial x}(r\cos\varphi, r\sin\varphi) r\sin\varphi + \frac{\partial f}{\partial y}(r\cos\varphi, r\sin\varphi) r\cos\varphi \Big)$$