Functions and Taylor's formula

Keywords: partial derivative, gradient, Hesse matrix, differential, directional derivative, chain rule, implicit function and derivative, Taylor formula, unconstrained optimization

Contents

1 The Taylor formula for a function in one variable

We start with the following important fact and try to approximate functions by polynomials.

Theorem 1.1 *Let I be an open interval, $f : I \to \mathbb{R}$ a $(k+1)$-times continuously differentiable function, $k \in \mathbb{N}$ and $a \in I$. Then for all $t \in I$ we have:*

$$f(t) \; = \; \underbrace{\sum_{j=0}^{k} \frac{f^j(a)}{j!} \, (t-a)^j}_{=:P_k(t,a)} + R_k(a, t-a)$$

with $\lim\limits_{t \to a} \dfrac{f(t) - P_k(t,a)}{(t-a)^k} \; = \; \lim\limits_{t \to a} \dfrac{R_k(a, t-a)}{(t-a)^k} \; = \; 0.$

This means, that $R_k(a, t-a)$ tends faster to 0 as the function $(t-a)^k$ if $t \to a$.

Definition 1.1 *The polynomial (in t) $P_k(t,a)$ is called the k-th Taylor polynomial for f at a.*

Theorem 1.2 *Suppose that*

$$f'(a) \; = \; f^{(2)}(a) \; = \; \ldots \; = \; f^{(k-1)}(a) \; = \; 0$$
$$f^{(k)}(a) \; \neq \; 0$$

1. *If k is even and $f^{(k)}(a) \; > \; 0$, then f has a local minimum at a.*
2. *If k is even and $f^{(k)}(a) \; < \; 0$, then f has a local maximum at a.*
3. *If k is odd, then f has neither a local maximum nor a local minimum at a.*

2 Differentiable functions of several variables

2.1 Partial derivative

Definition 2.1 *Let $y = f(\mathbf{x}) = f(x_1, \ldots, x_i, \ldots, x_n)$ be a function. For $i = 1, 2, \ldots, n$ the i-th partial derivative of f is defined by*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = f_{x_i}(\mathbf{x}) \quad = \quad \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e_i}) - f(\mathbf{x})}{t}$$

The function f is called 2-times (k)-times partially differentiable, if all partial derivatives of second order

$$f_{x_i x_j} \quad = \quad (f_{x_i})_{x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \qquad\qquad (1 \leq i, j \leq n)$$

exist.

The following fact is sometimes important:

Theorem 2.1 *If all partial derivatives of second order exist and are continuous functions, then $f_{x_i x_j} = f_{x_j x_i}$.*

Definition 2.2 *Let $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in D \subset \mathbb{R}^n$ be a point in the domain of f. The vector*

$$\nabla f(\mathbf{a}) \quad = \quad \begin{pmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{pmatrix}$$

is called gradient of f in $\mathbf{a}$. The $n \times n$ matrix

$$\nabla^2 f(\mathbf{a}) \quad = \quad \begin{pmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & \ldots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & \ldots & f_{x_2x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & \vdots \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & \ldots & f_{x_nx_n}(\mathbf{a}) \end{pmatrix}$$

is called Hesse matrix of f in $\mathbf{a}$.

Definition 2.3 *Let* $\mathbf{a} = (a_1, a_2, \dots, a_n) \in D \subset \mathbb{R}^n$ *be a point in the domain of a map* $\mathbf{f}$*:*

$$\mathbf{f} : D \longrightarrow \mathbb{R}^m$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longmapsto \mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_m(x_1, x_2, \dots, x_n) \end{pmatrix}$$

The $m \times n$ *matrix*

$$D\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{a}) & \frac{\partial f_m}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

is called the <u>*Jacobi matrix of* $\mathbf{f}$ *in* $\mathbf{a}$</u>.

2.2 The differential and differentiable functions

Definition 2.4 *The (total) differential df of f is defined by*

$$\boxed{df \;=\; df(\mathbf{x}, d\mathbf{x}) \;=\; f_{x_1}(\mathbf{x}) \cdot dx_1 + \cdots + f_{x_n}(\mathbf{x}) \cdot dx_n}$$

Definition 2.5 *Let $D \subset \mathbb{R}^n$ be an open set, $\mathbf{a}$ and $\mathbf{x} = \mathbf{a} + d\mathbf{x} \in D$. A function $f : D \to \mathbb{R}$ is called (totally) differentiable in $\mathbf{a}$, if*

$$\underbrace{f(\mathbf{x}) \;=\; f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet d\mathbf{x} + R(\mathbf{a}, d\mathbf{x})}_{*} \quad \textit{and} \quad \underbrace{\lim_{\mathbf{x}\to\mathbf{a}} \frac{R(\mathbf{a}, d\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} \;=\; 0}_{\star}$$

- The function $t(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet d\mathbf{x}$ is called tangent hyperplane of f in $\mathbf{a}$:

$$t(\mathbf{x}) \;=\; f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) \;=\; f(\mathbf{a}) + df(\mathbf{a}, d\mathbf{x})$$

- A differentiable function can be approximated (very well) by a linear function and the claim $\star$ is essential.

- If we use the notation $\Delta f(\mathbf{a}, d\mathbf{x}) = f(\mathbf{a} + d\mathbf{x}) - f(\mathbf{a})$ for the real change of f and $\mathbf{x} = \mathbf{a} + d\mathbf{x}$ we get

$$\Delta f(\mathbf{a}, d\mathbf{x}) \;=\; df(\mathbf{a}, d\mathbf{x}) + R(\mathbf{a}, d\mathbf{x})$$

2.3 The directional derivative

Definition 2.6 *Let* $\mathbf{v} \in \mathbb{R}^n$ *be a vector. The limit (if it exists)*

$$\partial_{\mathbf{v}} f(\mathbf{a}) \quad = \quad \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

is called the derivative of f *in* $\mathbf{a}$ *along* $\mathbf{v}$*.*
If $\mathbf{v}$ *is a vector of length* 1 *(unit vector) then* $\partial_{\mathbf{v}} f(\mathbf{a})$ *is called the directional derivative of* f *in* $\mathbf{a}$ *in direction* $\mathbf{v}$*.*

Theorem 2.2 *Let* D *be open,* f *differentiable on* D *and* $\mathbf{v} \in \mathbb{R}^n$ *with* $||\mathbf{v}|| = 1$*. Then*

$$\partial_{\mathbf{v}} f(\mathbf{a}) \quad = \quad \nabla f(\mathbf{a}) \bullet \mathbf{v} \; = \; \sum_{i=1}^{n} f_{x_i}(\mathbf{a})\, v_i$$

Proof: Let f be totally differentiable in $\mathbf{a}$, then

$$f(\mathbf{x}) \; = \; f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + R(\mathbf{a}, d\mathbf{x}) \quad \text{und} \quad \lim_{\mathbf{x} \to \mathbf{a}} \frac{R(\mathbf{a}, d\mathbf{x})}{||\mathbf{x} - \mathbf{a}||} \; = \; 0$$

With $\mathbf{x} = \mathbf{a} + t\mathbf{v}$ we get:

$$f(\mathbf{x}) - f(\mathbf{a}) \; = \; f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a}) \; = \; \nabla f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{a}, d\mathbf{x}).$$

Hence:

$$\begin{aligned}
\partial_{\mathbf{v}} f(\mathbf{a}) &= \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t} \\
&= \lim_{t \to 0} \frac{\nabla f(\mathbf{a}) \bullet t\mathbf{v} + R(\mathbf{a}, d\mathbf{x})}{t} \\
&= \nabla f(\mathbf{a}) \bullet \mathbf{v} + \lim_{t \to 0} \frac{R(\mathbf{a}, d\mathbf{x})}{t} \\
&= \nabla f(\mathbf{a}) \bullet \mathbf{v}.
\end{aligned}$$

□

Theorem 2.3 (Properties of the gradient $\nabla f(\mathbf{a})$)

- *The gradient of f in $\mathbf{a}$ is orthogonal to the level set*

$$L \;=\; L_{f(\mathbf{a})} \;=\; \{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = f(\mathbf{a}) \,\}$$

  (shortly $f(\mathbf{x}) = f(\mathbf{a})$).

- *The gradient of f in $\mathbf{a}$ points in the direction of the greatest rate of increase of the function f in $\mathbf{a}$.*

Proof: For $\mathbf{v} \in \mathbb{R}^n$ with $||\mathbf{v}|| = 1$ we have

$$\begin{aligned} \partial_{\mathbf{v}} f(\mathbf{a}) &= \nabla f(\mathbf{a}) \bullet \mathbf{v} \\ &= ||\nabla f(\mathbf{a})|| \cdot ||\mathbf{v}|| \cdot \cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) \\ &= ||\nabla f(\mathbf{a})|| \cdot \cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) \end{aligned}$$

If $\mathbf{v}$ is a tangent vector to a curve in the level set $f(\mathbf{x}) = f(\mathbf{a})$ then $\partial_{\mathbf{v}} f(\mathbf{a}) = 0$ and $\cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) = 0$ or $\angle(\nabla f(\mathbf{a}), \mathbf{v}) = \pi/2$.

Because $||\nabla f(\mathbf{a})|| > 0$ is constant and $-1 \leq \cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) \leq 1$ we see that

- $\partial_{\mathbf{v}} f(\mathbf{a})$ is maximal if $\cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) = 1$ or $\angle(\nabla f(\mathbf{a}), \mathbf{v}) = 0$ ($\mathbf{v}$ and $\nabla f(\mathbf{a})$ have the same direction),
- $\partial_{\mathbf{v}} f(\mathbf{a})$ is minimal if $\cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) = -1$ or $\angle(\nabla f(\mathbf{a}), \mathbf{v}) = \pi$ ($\mathbf{v}$ and $\nabla f(\mathbf{a})$ have the opposite direction).

$\square$

Sometimes it is useful to understand

$$\partial_{\mathbf{v}} \;=\; v_1 \frac{\partial}{\partial x_1} + \cdots + v_n \frac{\partial}{\partial x_n}$$

as a so called differential operator. We denote by $C^l(D, \mathbb{R})$ the set of all l-times continuously differentiable functions $f : D \to \mathbb{R}$. Then

$$\begin{aligned} \partial_{\mathbf{v}} : C^l(D, \mathbb{R}) &\to C^{l-1}(D, \mathbb{R}) \\ f(\mathbf{x}) &\mapsto v_1 \frac{\partial}{\partial x_1} f(\mathbf{x}) + \cdots + v_n \frac{\partial}{\partial x_n} f(\mathbf{x}) \;=\; \nabla f(\mathbf{x}) \bullet \mathbf{v}. \end{aligned}$$

Then we can recursively define the operators $\partial_{\mathbf{v}}^l$ by

$$\partial_{\mathbf{v}}^l f(\mathbf{x}) \;:=\; \partial_{\mathbf{v}}(\partial_{\mathbf{v}}^{l-1} f(\mathbf{x}))$$

We get:

$$\begin{aligned} \partial_{\mathbf{v}} f(\mathbf{x}) &= \sum_{i=1}^{n} f_{x_i}(\mathbf{x}) \cdot v_i \\ \partial_{\mathbf{v}}^2 f(\mathbf{x}) &= \partial_{\mathbf{v}} \left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x}) \cdot v_i \right) \\ &= \sum_{i=1}^{n} \partial_{\mathbf{v}} \left(f_{x_i}(\mathbf{x}) \right) \cdot v_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}) \cdot v_j \right) \cdot v_i \\ &= \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) \cdot v_i \cdot v_j \\ &= \mathbf{v}^T \, \nabla^2 f(\mathbf{x}) \, \mathbf{v}. \end{aligned}$$

Hence

$$\partial_{\mathbf{v}}^2 f(\mathbf{x}) \;=\; \mathbf{v}^T \begin{pmatrix} f_{x_1 x_1}(\mathbf{x}) & \ldots & f_{x_1 x_n}(\mathbf{x}) \\ \vdots & \vdots & \vdots \\ f_{x_n x_1}(\mathbf{x}) & \ldots & f_{x_n x_n}(\mathbf{x}) \end{pmatrix} \mathbf{v}.$$

2.4 The chain rule

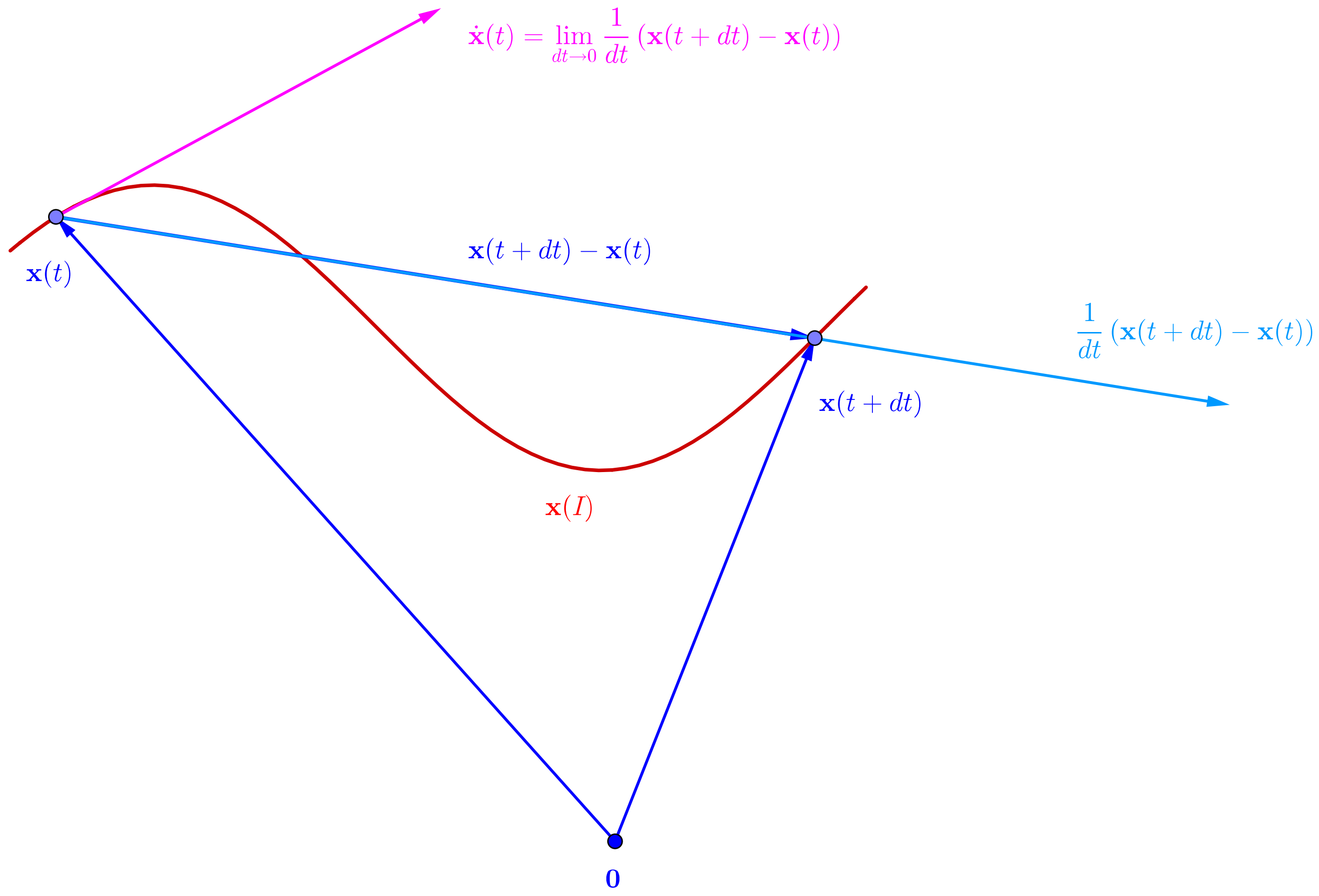
Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ continuously partially differentiable, $I \subset \mathbb{R}$ and

$$\mathbf{x} : I \to D \subset \mathbb{R}^n \quad \text{with} \quad \mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}$$

with differentiable coordinate functions $x_i(t)$ für $1 \leq i \leq n$. The image $\mathbf{x}(I) \subset D \subset \mathbb{R}^n$ is a curve and for all $t \in I$ the vector

$$\dot{\mathbf{x}}(t) = \lim_{dt \to 0} \frac{1}{dt} (\mathbf{x}(t + dt) - \mathbf{x}(t)) = \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix}$$

is the so called tangent vector at the curve $\mathbf{x}(I)$ in the point $\mathbf{x}(t)$.



Theorem 2.4 *The composition* $f \circ \mathbf{x} : I \to \mathbb{R}$ *where* $f \circ \mathbf{x}(t) = f(\mathbf{x}(t))$ *is differentiable with*

$$\boxed{\frac{d}{dt}\, f(\mathbf{x}(t)) \quad = \nabla f(\mathbf{x}(t)) \bullet \frac{d}{dt}\, \mathbf{x}(t)}$$

Expansion:

$$\begin{aligned}
& \frac{d}{dt}\, f(\mathbf{x}(t)) \\
= \quad & \nabla\, f(\mathbf{x}(t)) \bullet \frac{d}{dt}\, \mathbf{x}(t) \\
= \quad & \frac{d}{dt}\, f(x_1(t), x_2(t), \ldots, x_n(t)) \\
= \quad & f_{x_1}(\mathbf{x}(t))\, \frac{d}{dt} x_1(t) + f_{x_2}(\mathbf{x}(t))\, \frac{d}{dt} x_2(t) + \cdots + f_{x_n}(\mathbf{x}(t))\, \frac{d}{dt} x_n(t) \\
= \quad & f_{x_1}(\mathbf{x}(t))\, \dot{x}_1(t) + f_{x_2}(\mathbf{x}(t))\, \dot{x}_2(t) + \cdots + f_{x_n}(\mathbf{x}(t))\, \dot{x}_n(t)
\end{aligned}$$

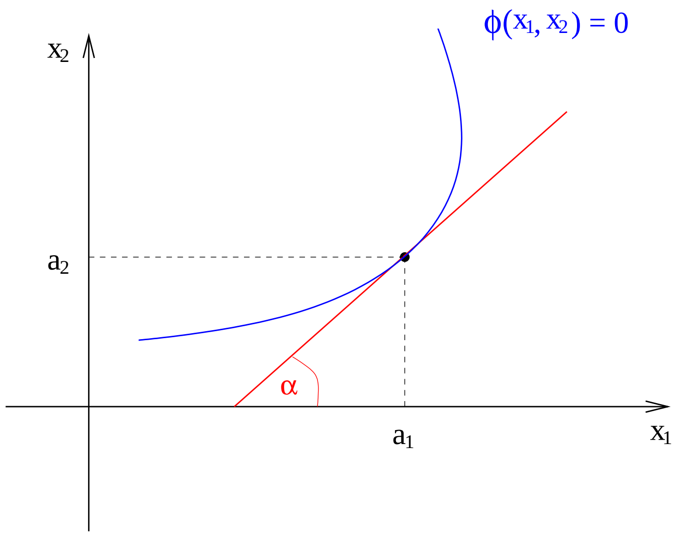
2.5 Implicit function theorem

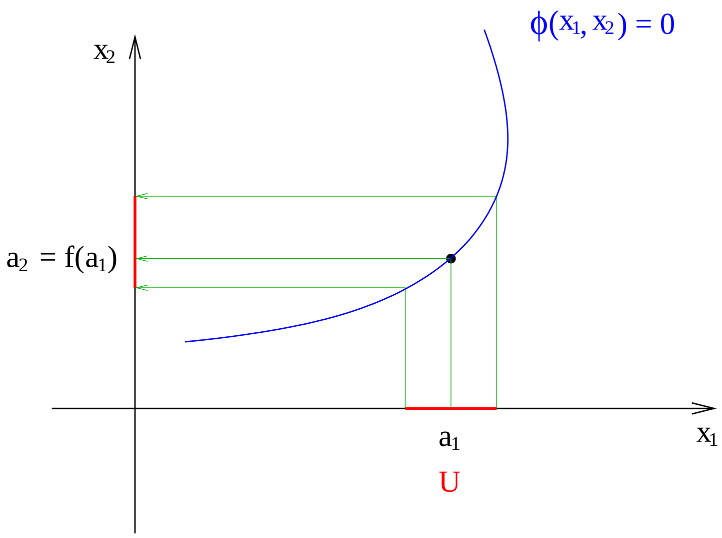
Notation: $(\mathbf{x}, y) = (x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1}$

Theorem 2.5 *Let $M \subset \mathbb{R}^{n+1}$ be open, $\phi : M \to \mathbb{R}$ continuously partially differentiable and $\mathbf{a} = (a_1, \ldots, a_n, a_{n+1}) \in M$ with $\phi(\mathbf{a}) = 0$ and $\phi_y(\mathbf{a}) \neq 0$. Then there is a neighbourhood U of $(a_1, \ldots, a_n)$ and an open interval $I \subset \mathbb{R}$ with $a_{n+1} \in I$ such that:*

1. *$R := \{ (\mathbf{x}, y) \subset \mathbb{R}^{n+1} \mid \mathbf{x} \subset U$ and $y \in I \} \subset M$ and $\phi_y(\mathbf{x}) \neq 0$ for all $(\mathbf{x}, y) \in R$.*
2. *For each $\mathbf{x} \in U$ there exists exactly one $y \in I$ with $\phi(\mathbf{x}, y) = 0$. The function $y := f(\mathbf{x})$ is partially differentiable ($f : U \to I$) and*

$$\phi(\mathbf{x}, y) = \phi(\mathbf{x}, f(\mathbf{x})) = 0 \quad \longrightarrow \quad \frac{\partial}{\partial x_i} f(\mathbf{x}) = -\frac{\frac{\partial}{\partial x_i}\phi(\mathbf{x}, y)}{\frac{\partial}{\partial y}\phi(\mathbf{x}, y)}$$





Let $y := f(\mathbf{x})$ for all $\mathbf{x} \in U$ the function above. Then

$$\phi(\mathbf{x}, y) \;=\; \phi(\mathbf{x}, f(\mathbf{x})) \;=\; 0$$

By the chain rule we get:

$$\begin{aligned}
0 \;=\; \frac{\partial}{\partial x_i}\, 0 \;&=\; \frac{\partial}{\partial x_i}\, \phi(\overbrace{x_1, \ldots, x_n}^{\mathbf{x}}, \overbrace{f(x_1, \ldots, x_n)}^{y}) \\
&=\; \sum_{j=1}^{n} \frac{\partial}{\partial x_j}\, \phi\,(\mathbf{x}, y) \cdot \frac{\partial x_j}{\partial x_i} + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial y}{\partial x_i} \\
&=\; \frac{\partial}{\partial x_i}\, \phi(\mathbf{x}, y) \cdot \frac{\partial x_i}{\partial x_i} + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial y}{\partial x_i} \\
&=\; \frac{\partial}{\partial x_i}\, \phi(\mathbf{x}, y) + \frac{\partial}{\partial y}\, \phi(\mathbf{x}, y) \cdot \frac{\partial}{\partial x_i}\, f(\mathbf{x})
\end{aligned}$$

Solving this equation for $\frac{\partial}{\partial x_i} f(\mathbf{x})$ proves the second part of the Theorem.

□

Example 2.1 *We could prove that if ϕ is twice continuously differentiable and $\phi(x, y)$ defines y as a twice differentiable function of x, then*

$$y'' \;=\; -\frac{\phi_{xx} + 2\phi_{xy} \cdot y' + \phi_{yy} \cdot (y')^2}{\phi_y}.$$

3 The general Taylor formula

Theorem 3.1 *Let $D \subset \mathbb{R}^n$ be an open and convex set and $f : D \to \mathbb{R}$ a $(k+1)$-times continuously differentiable function, $\mathbf{a}, \mathbf{x} \in D$ and $\mathbf{v} = d\mathbf{x} = \mathbf{x} - \mathbf{a}$. Then we have:*

$$f(\mathbf{a}+\mathbf{v}) \;=\; f(\mathbf{x}) \;=\; \underbrace{\sum_{j=0}^{k} \frac{1}{j!}\,\partial_{\mathbf{v}}^{j} f(\mathbf{a})}_{=:P_k(\mathbf{x},\mathbf{a})} + R_k(\mathbf{a},\mathbf{v})$$

with $\lim\limits_{\mathbf{x}\to\mathbf{a}} \dfrac{f(\mathbf{x}) - P_k(\mathbf{x},\mathbf{a})}{(\mathbf{x}-\mathbf{a})^k} \;=\; \lim\limits_{\mathbf{x}\to\mathbf{a}} \dfrac{R_k(\mathbf{a},\mathbf{v})}{(\mathbf{x}-\mathbf{a})^k} \;=\; 0.$

The polynomial (in $\mathbf{x}$) $P_k(\mathbf{x},\mathbf{a})$ is called the <u>*k-th Taylor polynomial for f at $\mathbf{a}$*</u>.

Proof:

We define the function h in one real variable t by

$$h(t) \;:=\; f(\mathbf{a}+t\mathbf{v}) \;=\; f(a_1 + tv_1, \ldots, a_n + tv_n)$$

By using the Taylor formula for h at $t = 0$ we get

$$h(t) \;=\; \sum_{j=0}^{k} \frac{h^j(0)}{j!}\, t^j + R_k(0,t)$$

and

$$f(\mathbf{x}) \;=\; f(\mathbf{a}+\mathbf{v}) \;=\; h(1) \;=\; \sum_{j=0}^{k} \frac{h^j(0)}{j!} + R_k(0,1)$$

By using the chain rule we see that $h^{(j)}(0) \;=\; \partial_{\mathbf{v}}^{j} f(\mathbf{a})$, for instance:

$$h(0) \;=\; f(\mathbf{a})$$

$$h'(t) \;=\; \nabla f(\mathbf{a}+t\mathbf{v}) \bullet \mathbf{v} \qquad h'(0) \;=\; \nabla f(\mathbf{a}) \bullet \mathbf{v} \;=\; \partial_{\mathbf{v}} f(\mathbf{a})$$

$$h''(t) \;=\; \mathbf{v}^T\, \nabla^2 f(\mathbf{a}+t\mathbf{v})\, \mathbf{v} \qquad h''(0) \;=\; \mathbf{v}^T\, \nabla^2 f(\mathbf{a})\, \mathbf{v} \;=\; \partial_{\mathbf{v}}^{2} f(\mathbf{a})$$

□

The Taylor formula can also be given in the following form.

Theorem 3.2 *Let $D \subset \mathbb{R}^n$ be an open and convex set and $f : D \to \mathbb{R}$ an $(k+1)$-times continuously differentiable function, $\mathbf{a}, \mathbf{x} \in D$ and $\mathbf{v} = d\mathbf{x} = \mathbf{x} - \mathbf{a}$. Then we have:*

$$\boxed{f(\mathbf{a}+\mathbf{v}) \;=\; \sum_{j=0}^{k} \frac{1}{j!}\,\partial_{\mathbf{v}}^{j} f(\mathbf{a}) \;+\; \frac{1}{(k+1)!}\,\partial_{\mathbf{v}}^{k+1} f(\mathbf{a}+c\mathbf{v})}$$

for some real number $c \in (0,1)$. This means that the point $\mathbf{a}+c\mathbf{v}$ lies between $\mathbf{a}$ and $\mathbf{a}+\mathbf{v}$ in the convex set D.

Example 3.1 *The 1-st Taylor polynomial of f in $\mathbf{a}$ is the well-known tangent hyperplane:*

$$P_1(\mathbf{x},\mathbf{a}) \;=\; f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x}-\mathbf{a}) \;=\; f(\mathbf{a}) + \sum_{j=1}^{n} f_{x_j}(\mathbf{a})(x_j - a_j)$$

If f is 1-times continuously differentiable on a convex set then

$$f(\mathbf{a}+\mathbf{v}) \;=\; f(\mathbf{a}) + \nabla f(\mathbf{a}+c\mathbf{v}) \bullet \mathbf{v} \;=\; f(\mathbf{a}) + \sum_{j=1}^{n} f_{x_j}(\mathbf{a}+c\mathbf{v})(x_j - a_j)$$

for some $c \in (0,1)$.

Example 3.2 *The 2-nd Taylor polynomial of f in $\mathbf{a}$ is:*

$$P_2(\mathbf{x},\mathbf{a}) \;=\; f(\mathbf{a}) + \nabla\, f(\mathbf{a}) \bullet (\mathbf{x}-\mathbf{a}) + \frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T\, \nabla^2 f(\mathbf{a})\;\, (\mathbf{x}-\mathbf{a})$$

If f is 2-times continuously differentiable on a convex set then

$$f(\mathbf{a}+\mathbf{v}) \;=\; f(\mathbf{a}) + \nabla\, f(\mathbf{a}) \bullet \mathbf{v} + \frac{1}{2}\,\mathbf{v}^T\, \nabla^2 f(\mathbf{a}+c\mathbf{v})\, \mathbf{v}$$

for some $c \in (0,1)$.

4 Local minima in open sets

4.1 Introduction

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Let D be some **open** subset of $\mathbb{R}^n$ and $\mathbf{x}^* \in D$ a local minimum of f over D. This means that there exists an $\epsilon > 0$ such that for all $\mathbf{x} \in D$ satisfying $|\mathbf{x} - \mathbf{x}^*| < \epsilon$ we have $f(\mathbf{x}^*) \leq f(\mathbf{x})$.

The term ,,unconstrained" usually refers to the situation where all points $\mathbf{x}$ sufficiently near $\mathbf{x}^*$ are in D. This is automatically true if D is an open set.

4.2 First-order necessary condition for optimality

Suppose that f is a continuously differentiable function and $\mathbf{x}^* \in D$ is a local minimum.

Pick an arbitrary vector (direction) $\mathbf{v} \in \mathbb{R}^n$. Since we are in the unconstrained case, we have $\mathbf{x}^* + t\mathbf{v} \in D$ for all t with $-t_0 < t < t_0$.

For the fixed $\mathbf{v}$ we can consider $f(\mathbf{x}^* + t\mathbf{v})$ as a function of the real parameter t and we define

$$g(t) \quad := \quad f(\mathbf{x}^* + t\mathbf{v}).$$

Since $\mathbf{x}^*$ is a minimum of f, it is clear that $t = 0$ is a minimum of g, such that $g'(0) = 0$. We will try to re-express this result in terms of the original function f:

$$g(t) \quad = \quad f(\mathbf{x}^* + t\mathbf{v})$$

$$g'(t) \quad = \quad \nabla f(\mathbf{x}^* + t\mathbf{v}) \bullet \mathbf{v}$$

and

$$0 \; = \; g'(0) \quad = \quad \nabla f(\mathbf{x}^*) \bullet \mathbf{v}$$

Since $\mathbf{v}$ was arbitrary, we get the <u>first-order necessary condition for optimality</u>:

$$\boxed{\mathbf{x}^* \text{ is a local minimum } \implies \nabla f(\mathbf{x}^*) \; = \; \mathbf{0}}$$

4.3 Second-order necessary condition for optimality

We assume, as before, that $\mathbf{x}^* \in D$ is a local minimum of f. For an arbitrary vector $\mathbf{v}$ let $g(t) = f(\mathbf{x}^* + t\mathbf{v})$. Then

$$\begin{aligned} g'(t) &= \nabla f(\mathbf{x}^* + t\mathbf{v}) \bullet \mathbf{v} \;=\; \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^* + t\mathbf{v}) \cdot v_i \\ g''(t) &= \sum_{i=1}^{n} \left(\frac{d}{dt} f_{x_i}(\mathbf{x}^* + t\mathbf{v}) \right) \cdot v_i \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} f_{x_i x_j}(\mathbf{x}^* + t\mathbf{v}) \cdot v_j \right) \cdot v_i \\ &= \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}^* + t\mathbf{v}) \cdot v_i \cdot v_j. \end{aligned}$$

and

$$g''(0) \;=\; \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}^*) \cdot v_i \cdot v_j \;=\; \mathbf{v}^T \, \nabla^2 f(\mathbf{x}^*) \, \mathbf{v}.$$

If $\mathbf{x}^*$ is a local minimum of f then $g(t)$ has a local minimum in $t = 0$. Hence

$$0 \;\le\; g''(0) \;=\; \mathbf{v}^T \, \nabla^2 f(\mathbf{x}^*) \, \mathbf{v} \;=\; (v_1 \; v_2 \; \dots \; v_n) \begin{pmatrix} f_{x_1 x_1}(\mathbf{x}^*) & \dots & f_{x_1 x_n}(\mathbf{x}^*) \\ \vdots & \vdots & \vdots \\ f_{x_n x_1}(\mathbf{x}^*) & \dots & f_{x_n x_n}(\mathbf{x}^*) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

for **all** $\mathbf{v} \in \mathbb{R}^n$. We conclude that the matrix $\nabla^2 f(\mathbf{x}^*)$ must be postive semidefinite and this is the <u>second-order necessary condition for optimality</u>:

$\mathbf{x}^*$ is a local minimum $\;\Longrightarrow\;$ $\nabla^2 f(\mathbf{x}^*)$ is positive semidefinite