Functions and Taylor's formula

Keywords: partial derivative, gradient, Hesse matrix, differential, directional derivative, chain rule, implicit function and derivative, Taylor formula, concave and convex functions, quasi-concave and quasi-convex functions, 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:*

$$\begin{aligned} f(t) &= \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}\,(t-a)^j + R_{a,k}(t) \\ &= \underbrace{f(a) + \frac{f^{(1)}(a)}{1!}\,(t-a) + \frac{f^{(2)}(a)}{2!}\,(t-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}\,(t-a)^k}_{=:P_{a,k}(t)} + R_{a,k}(t) \end{aligned}$$

with $\lim_{t\to a} \dfrac{f(t) - P_{a,k}(t)}{(t-a)^k} = \lim_{t\to a} \dfrac{R_{a,k}(t)}{(t-a)^k} = 0.$

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

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

Example 1.1 *Let $f(t) = e^t$, $a = 0$ and $k = 2$. Then*

$$\begin{aligned} e^t &= \frac{1}{0!}\,e^0\,t^0 + \frac{1}{1!}\,e^0\,t^1 + \frac{1}{2!}\,e^0\,t^2 + R_{0,2}(t) \\ &= 1 + t + \frac{1}{2}\,t^2 + R_{0,2}(t) \end{aligned}$$

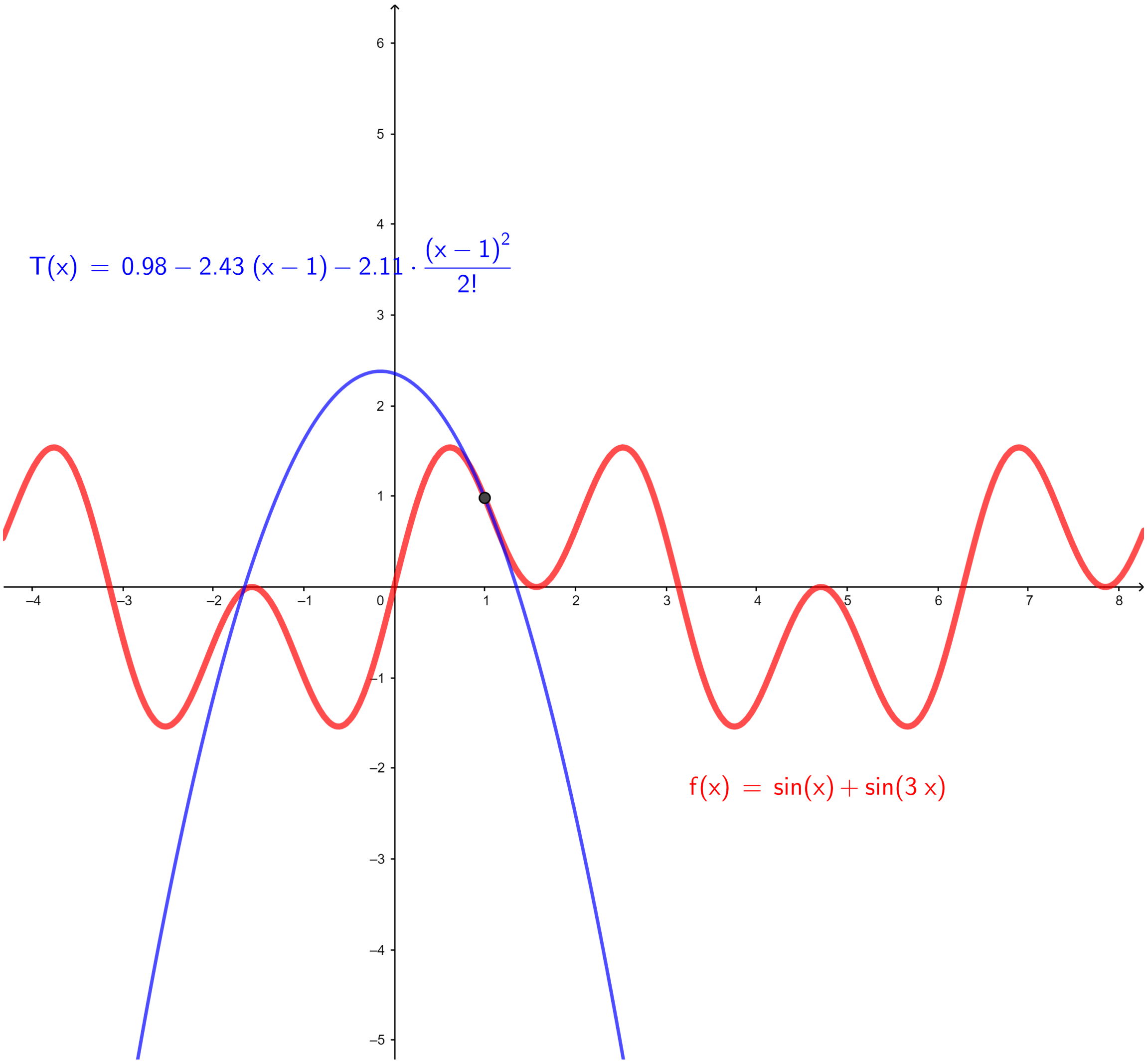
with $\lim_{t\to 0} \dfrac{e^t - 1 - t - \frac{1}{2}\,t^2}{t^2} = 0$ *(Verify this by using l'Hospital's rule!).*
The 2-th Taylor polynomial for f in $a = 0$ is $P_{0,2}(t) = 1 + t + \frac{1}{2}\,t^2$.

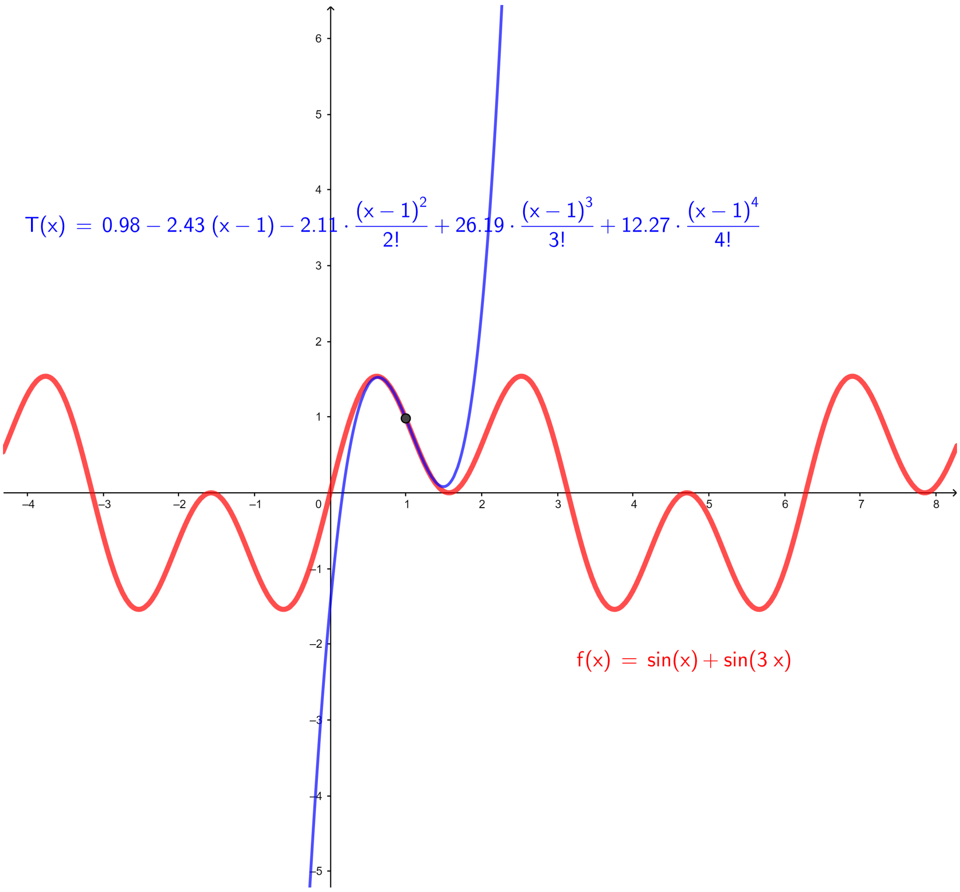
Example 1.2 *Let $f(t) = e^t$, $a = 1$ and $k = 2$. Then*

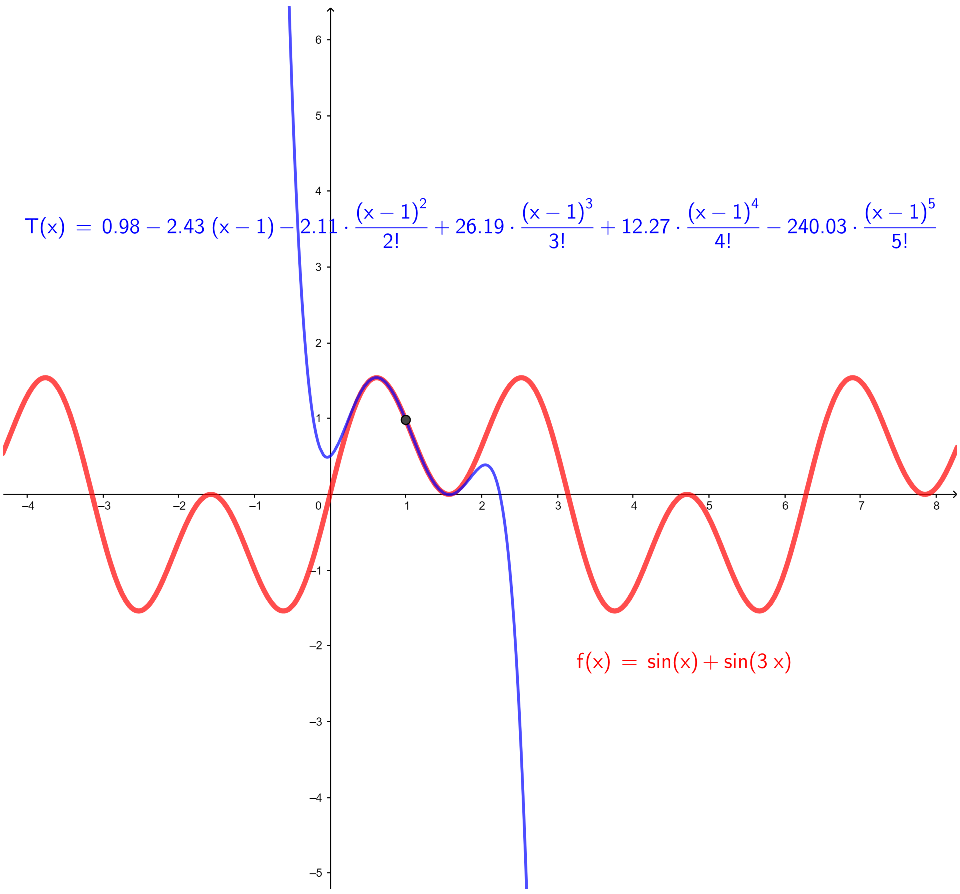
$$\begin{aligned} e^t &= \frac{1}{0!}\,e^1\,(t-1)^0 + \frac{1}{1!}\,e^1\,(t-1)^1 + \frac{1}{2!}\,e^1\,(t-1)^2 + R_{1,2}(t) \\ &= e + e(t-1) + \frac{1}{2}e(t-1)^2 + R_{1,2}(t). \end{aligned}$$

The 2-th Taylor polynomial for f in $a = 1$ is $P_{1,2}(t) = e + e(t-1) + \frac{1}{2}e(t-1)^2$.

Example 1.3 *Here you can see the graphs of some Taylor polynomials, denoted by $T(x)$, for the function $f(x) = \sin(x) + \sin(3x)$ in $a = 1$. We use x instead of t for the independent variable.*




T(x) = 0.98 − 2.43 (x − 1) − 2.11 · (x − 1)^2 / 2! + 26.19 · (x − 1)^3 / 3! + 12.27 · (x − 1)^4 / 4!
f(x) = sin(x) + sin(3 x)


T(x) = 0.98 − 2.43 (x − 1) − 2.11 · (x − 1)^2 / 2! + 26.19 · (x − 1)^3 / 3! + 12.27 · (x − 1)^4 / 4! − 240.03 · (x − 1)^5 / 5!
f(x) = sin(x) + sin(3 x)

We know, that if $f'(a) = 0$ and $f''(a) > (<) 0$, then f has a local minima (maxima) in a. More generally we have:

Theorem 1.2 *Suppose that*

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

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.*

Example 1.4 *Let $f(t) = t^4$. Then $f'(t) = 4t^3$, $f''(t) = 12t^2$, $f^{(3)}(t) = 24t$ and $f^{(4)}(t) = 24$. Hence $f'(0) = f''(0) = f^{(3)}(0) = 0$ and $f^{(4)}(0) = 24 > 0$. We see, that f has a local minimum in $a = 0$.*

2 Differentiable functions of several variables

2.1 Partial derivatives

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*

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

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

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

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*

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

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

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

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

Example 2.1 *Let $f(x_1, x_2) = x_1^{\alpha} x_2^{\beta}$. We see (by direct calculation):*

$$\nabla f(\mathbf{a}) \;=\; \begin{pmatrix} \alpha a_1^{\alpha-1} a_2^{\beta} \\ \beta a_1^{\alpha} a_2^{\beta-1} \end{pmatrix}$$

$$\nabla^2 f(\mathbf{a}) \;=\; \begin{pmatrix} \alpha(\alpha-1) a_1^{\alpha-2} a_2^{\beta} & \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} \\ \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} & \beta(\beta-1) a_1^{\alpha} a_2^{\beta-2} \end{pmatrix}$$

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

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

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

The $m \times n$ *matrix*

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

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

Example 2.2 *Let*

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} e^{x_1} + \cos(x_2) + x_1 + x_2^2 \\ x_1 x_2 \\ 2x_1 - 3x_2 \end{pmatrix}$$

and the Jacobi matrix of $\mathbf{f}$ *in* $\mathbf{a}$ *is the* 3×2 *matrix*

$$\mathbf{Df}(\mathbf{a}) = \begin{pmatrix} e^{a_1} + 1 & -\sin(a_2) + 2a_2 \\ a_2 & a_1 \\ 2 & -3 \end{pmatrix}$$

2.2 The differential and differentiable functions

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

$$df \;=\; df(\mathbf{x}, d\mathbf{x}) \;=\; f_{x_1}(\mathbf{x}) \cdot dx_1 + \cdots + f_{x_n}(\mathbf{x}) \cdot dx_n \;=\; \nabla f(\mathbf{a})^T \cdot d\mathbf{x}$$

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*

$$\underbrace{f(\mathbf{x}) \;=\; f(\mathbf{a}) + \nabla f(\mathbf{a})^T \cdot d\mathbf{x} + R_{\mathbf{a}}(\mathbf{x})}_{*} \quad \text{and} \quad \underbrace{\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_{\mathbf{a}}(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||} \;=\; 0}_{\star}$$

- We have: $\underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}}_{\mathbf{a}} + \underbrace{\begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix}}_{d\mathbf{x}}$

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

$$\begin{aligned} t(\mathbf{x}) &= f(\mathbf{a}) + \nabla f(\mathbf{a})^T \cdot (\mathbf{x} - \mathbf{a}) \\ &= f(\mathbf{a}) + f_{x_1}(\mathbf{a}) \cdot (x_1 - a_1) + \cdots + f_{x_n}(\mathbf{a}) \cdot (x_n - a_n) \\ &= f(\mathbf{a}) + df(\mathbf{a}, d\mathbf{x}) \end{aligned}$$

- 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

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

Example 2.3 *Let $f(x_1, x_2) = x_1^\alpha x_2^\beta$. We know*

$$\nabla f(\mathbf{a}) \;=\; \begin{pmatrix} \alpha a_1^{\alpha-1} a_2^\beta \\ \beta a_1^\alpha a_2^{\beta-1} \end{pmatrix}$$

and hence

$$df(\mathbf{a}, d\mathbf{x}) \;=\; df(a_1, a_2, dx_1, dx_2) \;=\; \alpha a_1^{\alpha-1} a_2^\beta \, dx_1 + \beta a_1^\alpha a_2^{\beta-1} \, dx_2$$

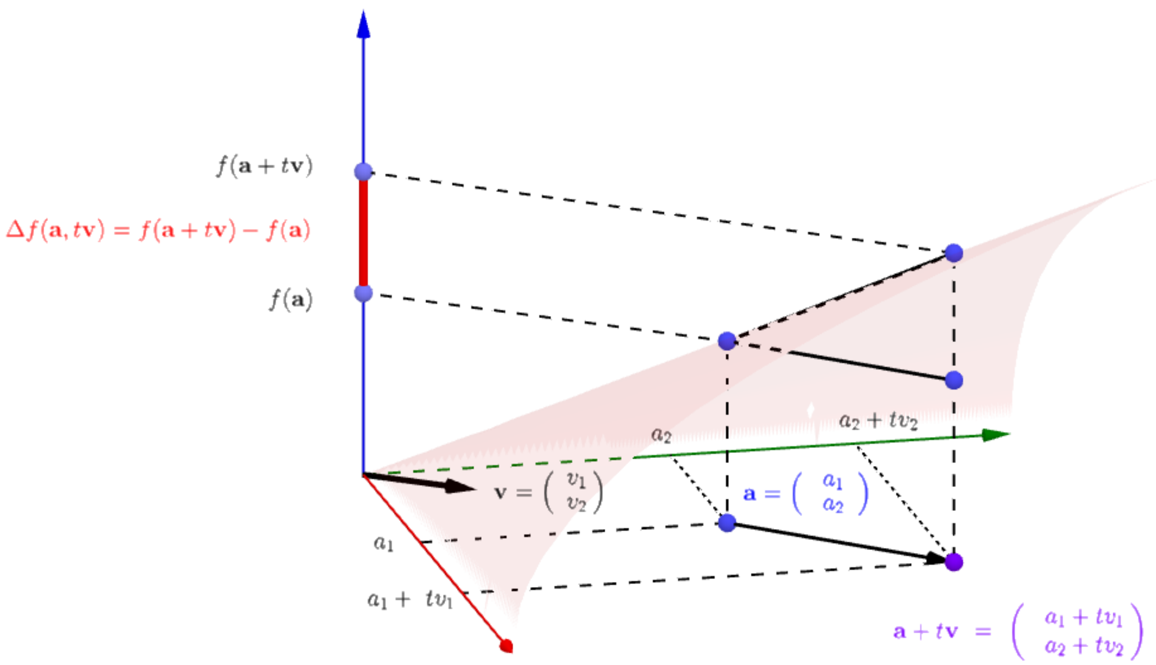
2.3 The directional derivative

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* $\mathbf{v}$.



The directional derivative of f in $\mathbf{a}$ in direction $\mathbf{v}$ is a generalisation of partial derivatives. For all $i = 1, \ldots, n$ we have:

$$\partial_{\mathbf{e}_i} f(\mathbf{a}) \quad = \quad f_{x_i}(\mathbf{a}).$$

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

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

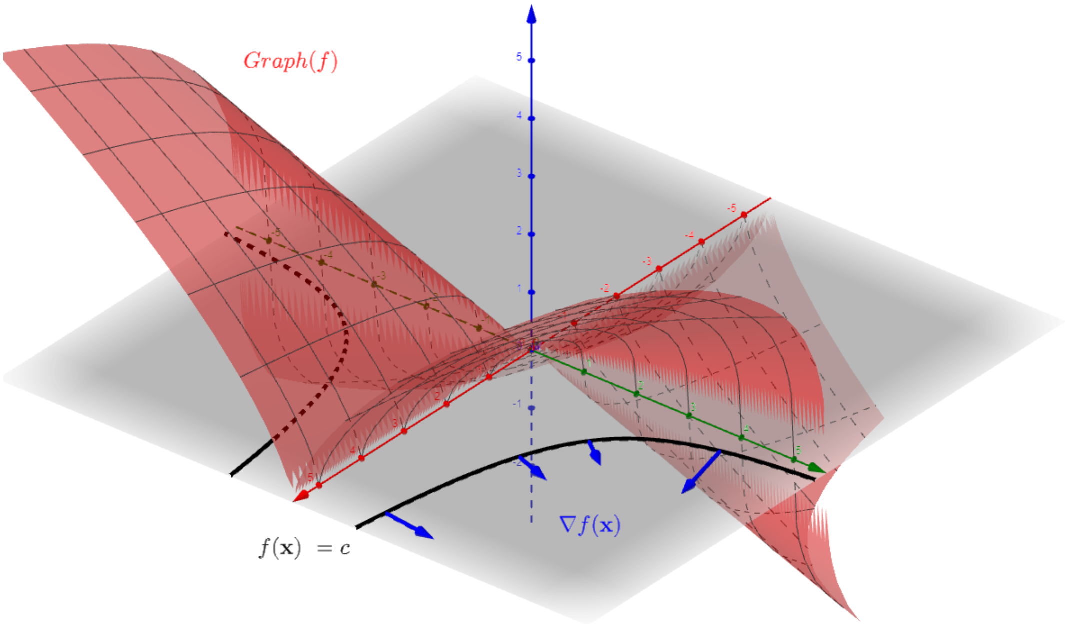
Theorem 2.3 (Properties of the gradient $\nabla f(\mathbf{a})$)

- *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

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

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 $-1 \le \cos \angle(\nabla f(\mathbf{a}), \mathbf{v}) \le 1$ we see that

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

□

2.4 The chain rule

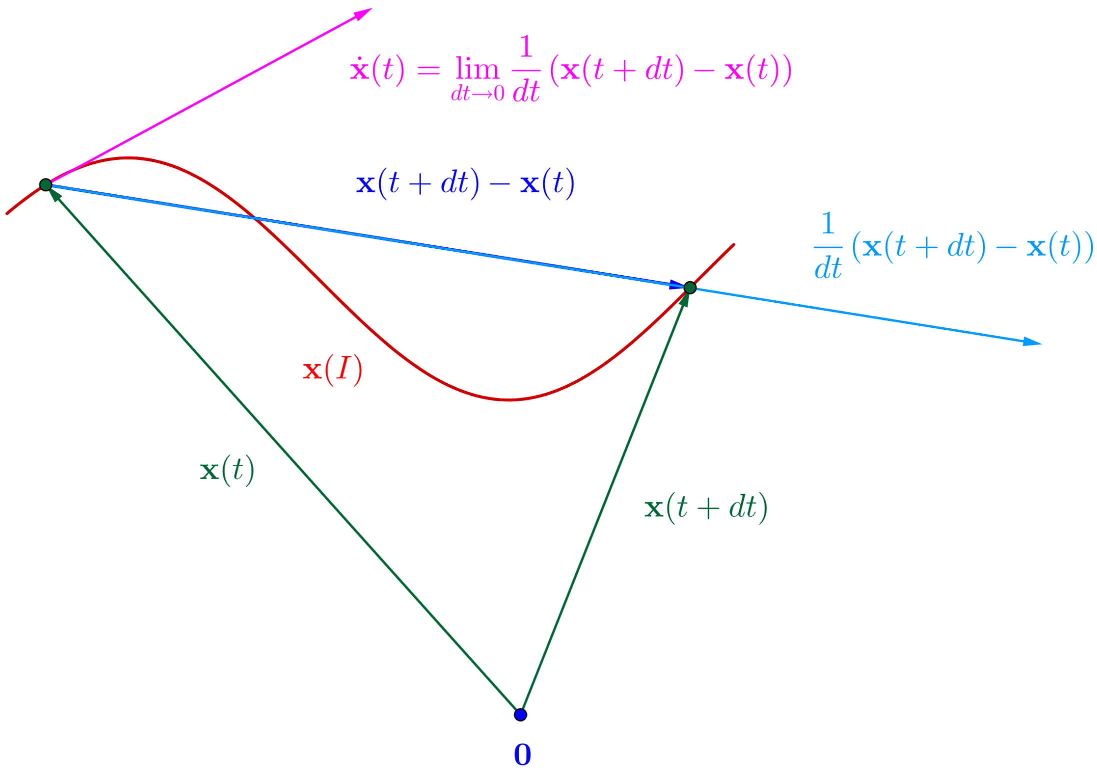
Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ continuously partially differentiable, $I \subset \mathbb{R}$ and

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

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

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

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



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*

$$\frac{d}{dt}\, f(\mathbf{x}(t)) \quad = \nabla f(\mathbf{x}(t))^T \cdot \frac{d}{dt}\mathbf{x}(t) \;=\; \begin{pmatrix} f_{x_1}(\mathbf{x}(t)) \\ f_{x_2}(\mathbf{x}(t)) \\ \vdots \\ f_{x_n}(\mathbf{x}(t)) \end{pmatrix}^T \cdot \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix}$$

Expansion:

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

Example 2.4 *Let $f(x_1, x_2) = x_1^\alpha x_2^\beta$, $x_1(t) = e^{2t}$ and $x_2(t) = t + 1$. By the chain rule we get*

$$\begin{aligned}
\frac{d}{dt}\, f(\mathbf{x}(t)) &= f_{x_1}(\mathbf{x}(t))\, \dot{x}_1(t) + f_{x_2}(\mathbf{x}(t))\, \dot{x}_2(t) \\
&= \alpha(e^{2t})^{\alpha-1}(t+1)^\beta \cdot 2e^{2t} + \beta(e^{2t})^\alpha(t+1)^{\beta-1} \cdot 1 \\
&= e^{2t\alpha}(t+1)^{\beta-1}[2\alpha(t+1) + \beta].
\end{aligned}$$

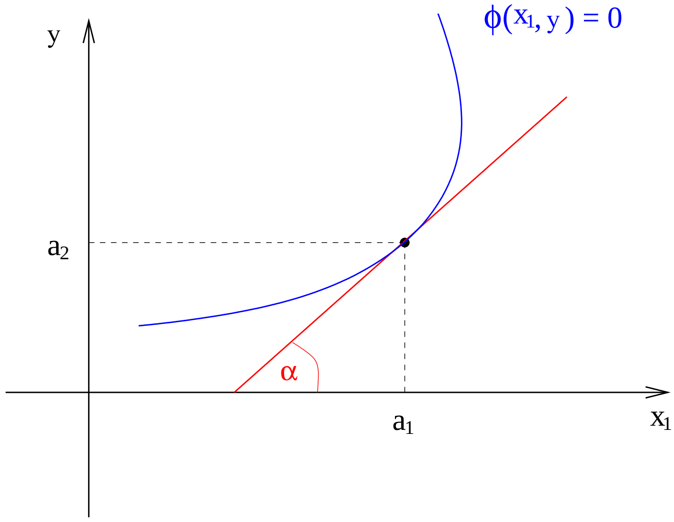
2.5 Implicit function theorem

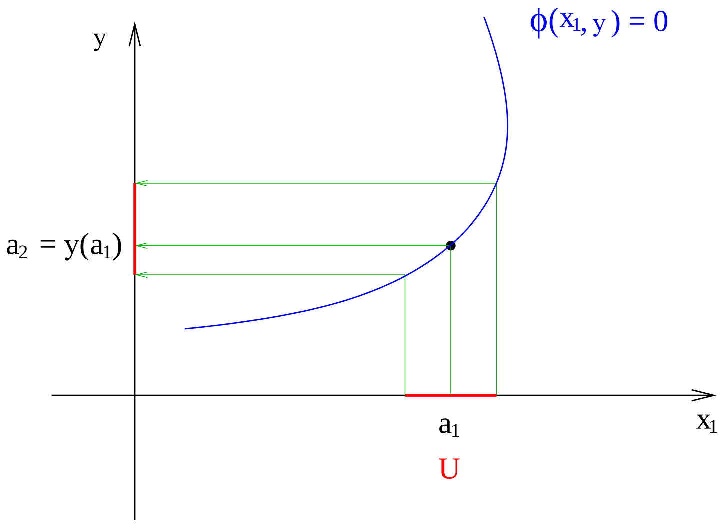
Notation: $(\mathbf{x}, y) = (x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1}$

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 $y := y(\mathbf{x})$ is partially differentiable ($y : U \to I$) and*

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





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

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

By the chain rule we get:

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

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

□

Example 2.5 *We could prove (differentiate the equation $y' = -\phi_x/\phi_y$ with respect to x and solve for y'') that if ϕ is twice continuously differentiable and $\phi(x, y)$ defines y as a twice differentiable function of x, then*

$$\begin{aligned}
y'' \;&=\; -\frac{\phi_{xx} + 2\phi_{xy} \cdot y' + \phi_{yy} \cdot (y')^2}{\phi_y} \\
&=\; \cdots \\
&=\; \frac{1}{(\phi_y)^3} \cdot \det \begin{pmatrix} 0 & \phi_x & \phi_y \\ \phi_x & \phi_{xx} & \phi_{xy} \\ \phi_y & \phi_{yx} & \phi_{yy} \end{pmatrix}
\end{aligned}$$

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 3-times continuously differentiable function, $\mathbf{a}, \mathbf{x} \in D$ and $\mathbf{v} = d\mathbf{x} = \mathbf{x} - \mathbf{a}$. Then we have:*

- $$f(\mathbf{x}) = \underbrace{f(\mathbf{a}) + \nabla f(\mathbf{a})^T \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T\ \nabla^2 f(\mathbf{a})\ \ (\mathbf{x}-\mathbf{a})}_{=:P_{\mathbf{a},2}(\mathbf{x})} + R_{\mathbf{a},k}(\mathbf{x})$$

  with $\lim_{\mathbf{x}\to\mathbf{a}} \frac{f(\mathbf{x}) - P_{\mathbf{a},k}(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^k} = \lim_{\mathbf{x}\to\mathbf{a}} \frac{R_{\mathbf{a},k}(\mathbf{x})}{||\mathbf{x}-\mathbf{a}||^k} = 0.$

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

  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.

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

Example 3.1 *Let $f(x_1,x_2) = x_1^\alpha x_2^\beta$. We already know:*

$$\nabla f(\mathbf{a}) = \begin{pmatrix} \alpha a_1^{\alpha-1} a_2^\beta \\ \beta a_1^\alpha a_2^{\beta-1} \end{pmatrix}$$

$$\nabla^2 f(\mathbf{a}) = \begin{pmatrix} \alpha(\alpha-1) a_1^{\alpha-2} a_2^\beta & \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} \\ \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} & \beta(\beta-1) a_1^\alpha a_2^{\beta-2} \end{pmatrix}$$

and

$$\begin{aligned} P_{\mathbf{a},2}(\mathbf{x}) &= a_1^\alpha a_2^\beta + \begin{pmatrix} \alpha a_1^{\alpha-1} a_2^\beta \\ \beta a_1^\alpha a_2^{\beta-1} \end{pmatrix}^T \cdot (\mathbf{x}-\mathbf{a}) \\ &\quad + \frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T \begin{pmatrix} \alpha(\alpha-1) a_1^{\alpha-2} a_2^\beta & \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} \\ \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} & \beta(\beta-1) a_1^\alpha a_2^{\beta-2} \end{pmatrix} (\mathbf{x}-\mathbf{a}) \\ &= a_1^\alpha a_2^\beta + \begin{pmatrix} \alpha a_1^{\alpha-1} a_2^\beta \\ \beta a_1^\alpha a_2^{\beta-1} \end{pmatrix}^T \cdot \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} \\ &\quad + \frac{1}{2} \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix}^T \begin{pmatrix} \alpha(\alpha-1) a_1^{\alpha-2} a_2^\beta & \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} \\ \alpha\beta a_1^{\alpha-1} a_2^{\beta-1} & \beta(\beta-1) a_1^\alpha a_2^{\beta-2} \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} \end{aligned}$$

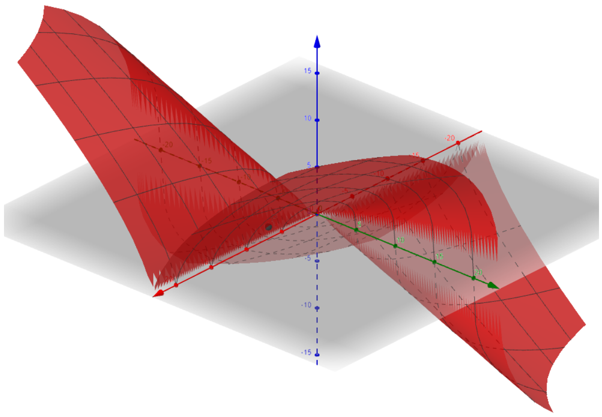
Example 3.2 *Let* $f(x_1, x_2) = x_1^{1/3} x_2^{2/3}$ *and* $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 8 \end{pmatrix}$. *We have* $f(\mathbf{a}) = f(1,8) = 4$,

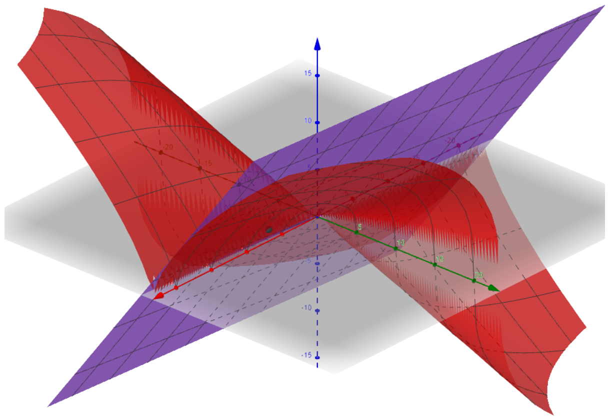
$$\nabla f(\mathbf{a}) = \begin{pmatrix} 4/3 \\ 1/3 \end{pmatrix}, \qquad \nabla^2 f(\mathbf{a}) = \begin{pmatrix} -8/9 & 1/9 \\ 1/9 & -1/72 \end{pmatrix}$$

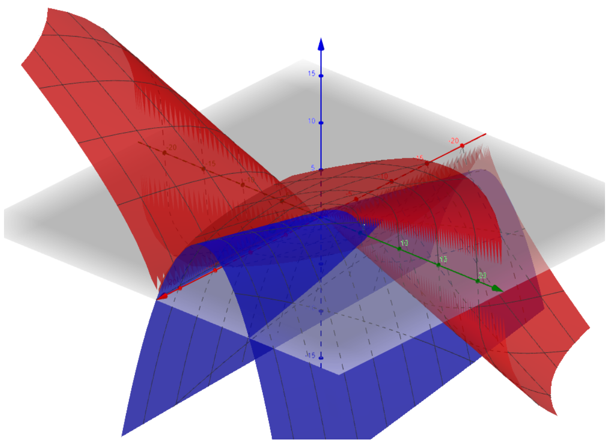
and

$$P_{\mathbf{a},k}(\mathbf{x}) = 4 + \begin{pmatrix} 4/3 \\ 1/3 \end{pmatrix}^T \cdot \begin{pmatrix} x_1 - 1 \\ x_2 - 8 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x_1 - 1 \\ x_2 - 8 \end{pmatrix}^T \begin{pmatrix} -8/9 & 1/9 \\ 1/9 & -1/72 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 - 8 \end{pmatrix}$$

Here you can see the graph of f *(red), the graph of* f *and of the 1-st Taylor polynomial (violet) and the graph of* f *and of the 2-nd Taylor polynomial (blue).*

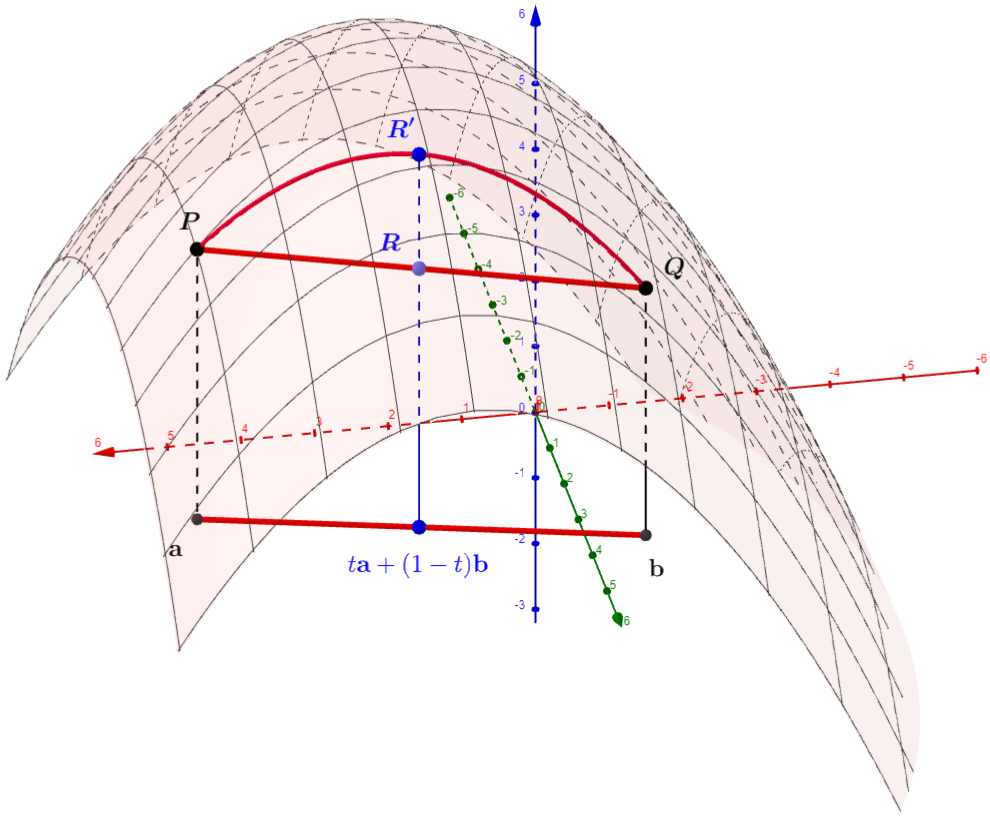






4 Concave and convex functions

The function $y = f(\mathbf{x})$ is called concave (convex) if it is defined on a convex set and the line segment joining any two points on the graph is never above (below) the graph. This definition is easy to understand but difficult to use. How could we check this criterion for a concrete function given by a complicated formula? We need an algebraic definition for concavity/convexity.



The two points P and Q correspond to the points $\mathbf{a}$ and $\mathbf{b}$ in the convex domain of f:

$$P = (\mathbf{a}, f(\mathbf{a})) \text{ and } Q = (\mathbf{b}, f(\mathbf{b})).$$

An arbitrary point R on the line segment PQ has the coordinates

$$\begin{aligned} R &= tQ + (1-t)P \\ &= P + t(Q-P) \\ &= (\, t\mathbf{a} + (1-t)\mathbf{b}, tf(\mathbf{a}) + (1-t)f(\mathbf{b}) \,) \end{aligned}$$

for a suitable $t \in [0, 1]$. This point R lies directly above the point $t\mathbf{a} + (1-t)\mathbf{b}$ and the point $t\mathbf{a}+(1-t)\mathbf{b}$ lies on the line segment between $\mathbf{a}$ and $\mathbf{b}$ in the (convex) domain of f. The corresponding point R' on the graph of f can be expressed as

$$R' = (\, t\mathbf{a} + (1-t)\mathbf{b}, f(t\mathbf{a} + (1-t)\mathbf{b}) \,)\,.$$

The fact that R does not lie above R' can be described by the following inequality:

$$f(t\mathbf{a} + (1-t)\mathbf{b}) \;\geq\; tf(\mathbf{a}) + (1-t)f(\mathbf{b}).$$

Definition 4.1 *A function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ defined on a convex set S is* <u>*concave (convex)*</u> *on S, if*

$$f(t\mathbf{a} + (1-t)\mathbf{b}) \ \geq (\leq)\ tf(\mathbf{a}) + (1-t)f(\mathbf{b})$$

or

$$f(t\mathbf{a} + (1-t)\mathbf{b}) - tf(\mathbf{a}) - (1-t)f(\mathbf{b}) \ \geq (\leq)\ 0$$

for **all** $\mathbf{a},\ \mathbf{b} \in S$ *and* **all** $t \in [0,1]$.

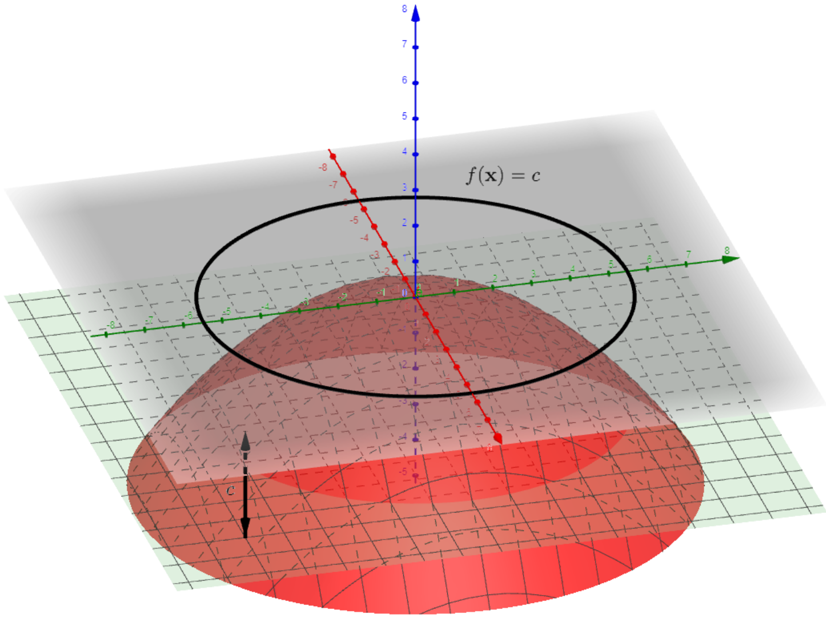
We may ask: What is the relationship between concavity/convexity and partial derivatives? The second partial derivatives $f_{x_1x_1}, f_{x_2x_2}, \ldots$ measure curvature along sections through the function, holding one variable constant. By Taylor's formula we know, that near $\mathbf{a} \in \mathbb{R}^n$ we have:

$$f(\mathbf{x}) \ \approx \ P_{\mathbf{a},2}(\mathbf{x}) \ = \ f(\mathbf{a}) + \nabla f(\mathbf{a})^T \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}\,(\mathbf{x}-\mathbf{a})^T\ \nabla^2 f(\mathbf{a})\ \ (\mathbf{x}-\mathbf{a})$$

The tangent hyperplane $f(\mathbf{a}) + \nabla f(\mathbf{a})^T \cdot (\mathbf{x}-\mathbf{a})$ has no curvature (this means: is concave **and** convex) and we may believe, that the Hessian $\nabla^2 f(\mathbf{a})$ of the function f embodies all the informations needed to determine concavity/convexity.

Theorem 4.1 *The function $y = f(\mathbf{x})$ is concave (convex) on the convex set S if and only if $\nabla^2 f(\mathbf{a})$ is negative semi-definite (positive semi-definite) for all $\mathbf{a} \in S$.*

Example 4.1



The function $f(\mathbf{x}) = f(x_1, x_2) = -x_1^2 - x_2^2$ *(in the picture you can see the graph of* f *and a typical level set) is concave on the convex set* $S = \mathbb{R}^2$*, because*

$$\nabla^2 f(\mathbf{a}) \;=\; \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

is obviously negative definite (and negative semi-definite) on $\mathbb{R}^2$.

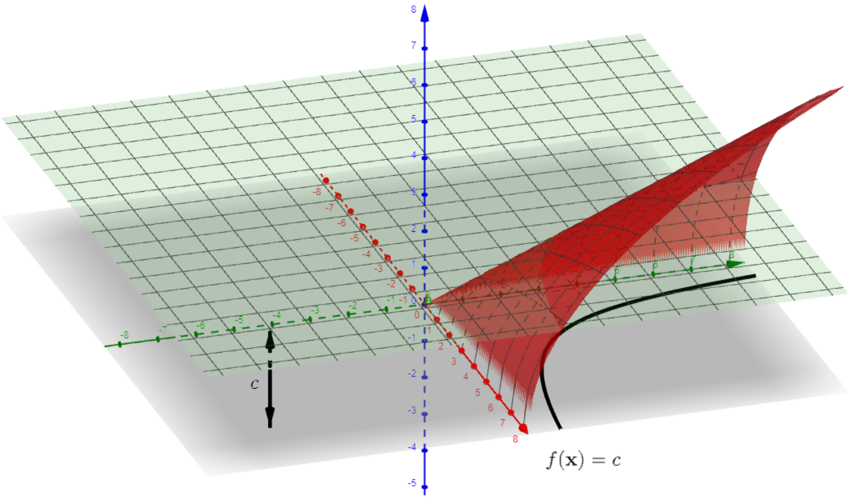
For a direct proof, we have to show that

$$f(t\mathbf{a} + (1-t)\mathbf{b}) - tf(\mathbf{a}) - (1-t)f(\mathbf{b}) \;\geq\; 0$$

for **all** $\mathbf{a} = (a_1, a_2),\ \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ *and* **all** $t \in [0,1]$. *Let us do it.*

$$\begin{aligned} & f(t\mathbf{a} + (1-t)\mathbf{b}) - tf(\mathbf{a}) - (1-t)f(\mathbf{b}) \\ = \;& -[ta_1 + (1-t)b_1]^2 - [ta_2 + (1-t)b_2]^2 - t(-a_1^2 - a_2^2) - (1-t)(-b_1^2 - b_2^2) \\ = \;& \ldots \\ = \;& t(1-t)[\,(a_1 - b_1)^2 + (a_2 - b_2)^2\,] \\ \geq \;& 0. \end{aligned}$$

Example 4.2



The function $f(\mathbf{x}) = f(x_1, x_2) = x_1^{1/2} x_2^{1/2}$ *(in the picture you can see the graph of* f *and a typical level set) is concave on the convex set* $S = \mathbb{R}^2_{++}$*. The Hessian*

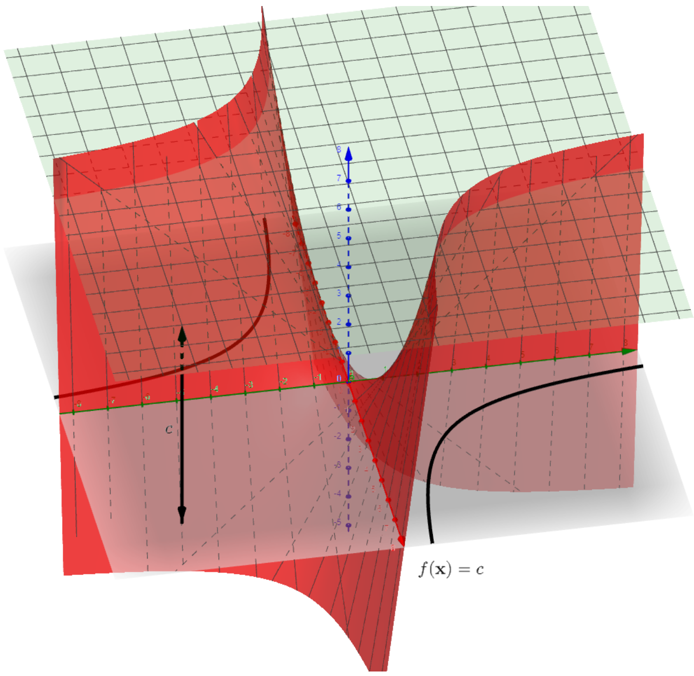
$$\nabla^2 f(\mathbf{a}) \;=\; \begin{pmatrix} -\frac{1}{4}a_1^{-3/2}a_2^{1/2} & \frac{1}{4}a_1^{-1/2}a_2^{-1/2} \\ \frac{1}{4}a_1^{-1/2}a_2^{-1/2} & -\frac{1}{4}a_1^{1/2}a_2^{-3/2} \end{pmatrix}$$

is negative semi-definite for all $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2_{++}$ $(a_1, a_2 > 0)$*, because (Hurwitz criterion)*

$$\begin{aligned} -\frac{1}{4}a_1^{-3/2}a_2^{1/2} &\leq 0 \quad \textit{(first leading subdeterminant)} \\ -\frac{1}{4}a_1^{1/2}a_2^{-3/2} &\leq 0 \\ \det \nabla^2 f(\mathbf{a}) \;=\; \frac{1}{16}a_1^{-1}a_2^{-1} - \frac{1}{16}a_1^{-1}a_2^{-1} &= 0 \end{aligned}$$

$a_1, a_2 > 0$

Example 4.3



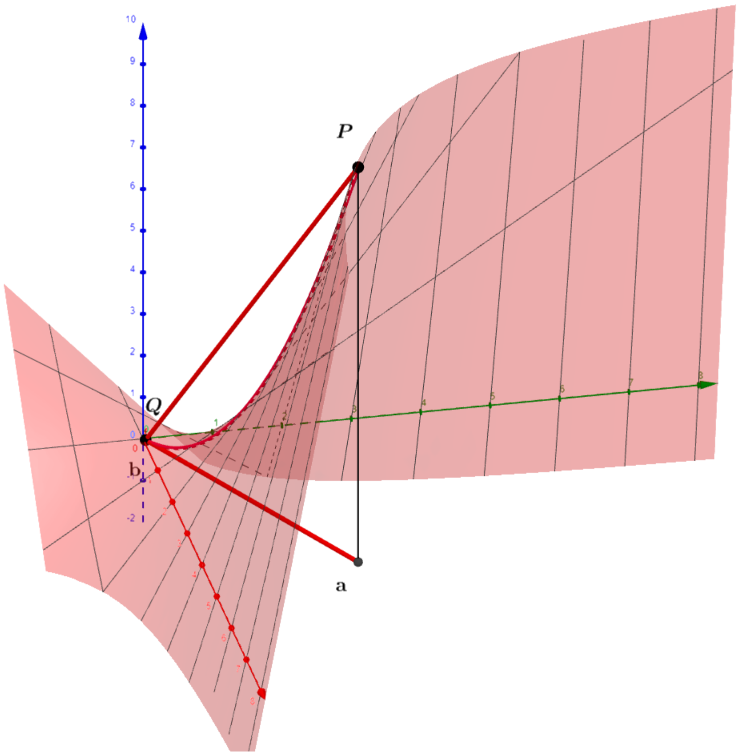
The function $f(\mathbf{x}) = f(x_1, x_2) = x_1 x_2$ *(in the picture you can see the graph of* f *and a typical level set) is neither concave nor convex on the convex set* $S = \mathbb{R}^2_{++}$*, because*

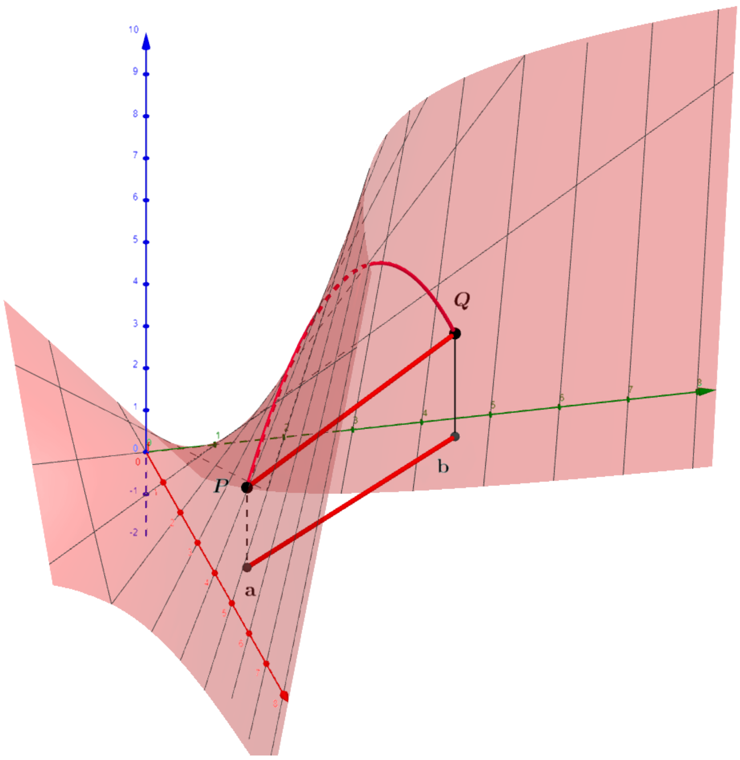
$$\nabla^2 f(\mathbf{a}) \;=\; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is obviously indefinite on $\mathbb{R}^2$*. By a direct calculation we see:*

$$\begin{aligned} & f(t\mathbf{a} + (1-t)\mathbf{b}) - tf(\mathbf{a}) - (1-t)f(\mathbf{b}) \\ = \; & [ta_1 + (1-t)a_2] \cdot [tb_1 + (1-t)b_2] - ta_1a_2 - (1-t)b_1b_2 \\ = \; & \ldots \\ = \; & t(1-t)[\,(-a_1 + b_1)(a_2 - b_2)\,] \end{aligned}$$

and there are points in $S = \mathbb{R}^2_{++}$ *where this term is positive (for example* $(0,3)$ *and* $(3,0)$*) or negative (for example* $(0,0)$ *and* $(3,3)$*) for all* $t \in [0,1]$.



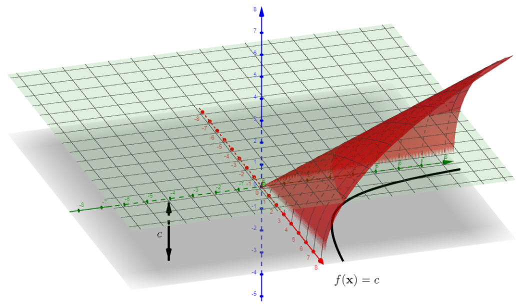


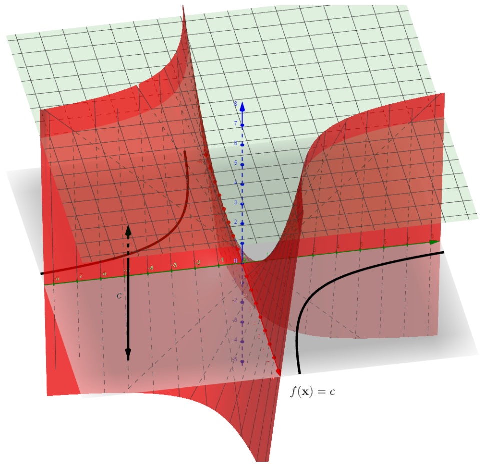
5 *Quasi-concave and quasi-convex functions*

We have seen that the level sets $f(\mathbf{x}) = c$ of the functions $f(x_1, x_2) = x_1^{1/2} x_2^{1/2}$ and $f(x_1, x_2) = x_1 x_2$ look similiar for $(x_1, x_2) \in \mathbb{R}^2_{++}$. Actually, we have

$$\begin{aligned} x_1^{1/2} x_2^{1/2} &= c \;\Rightarrow\; x_2(x_1) = \frac{c^2}{x_1} \\ x_1 x_2 &= c \;\Rightarrow\; x_2(x_1) = \frac{c}{x_1} \end{aligned}$$

and both level sets (curves) are decreasing and **convex** curves in the plane!





Both functions are increasing away from the origin ($f_{x_1}, f_{x_2} > 0$ in $\mathbb{R}^2_{++}$) **but only one** is concave in $\mathbb{R}^2_{++}$. Such level sets represent ,,mound-like" graphs, in a significant sense; and they are pervasive and important in economics. In fact, they are functions which we call quasi-concave (quasi-convex).

Definition 5.1 *A function $y = f(\mathbf{x})$ is called quasi-concave (quasi-convex) on a convex set S if and only if, whenever $\mathbf{a}, \mathbf{b} \in S$ with $f(\mathbf{a}) \geq c$ $(\leq c)$ and $f(\mathbf{b}) \geq c$ $(\leq c)$ then also*

$$f(t\mathbf{a} + (1-t)f(\mathbf{b})) \;\geq\; c \;\; (\leq c)$$

for all $t \in [0, 1]$.

Theorem 5.1

$$\begin{array}{lcl} f \text{ concave on } S & \Rightarrow & f \text{ quasi-concave on } S & \not\Rightarrow & f \text{ concave on } S \\ f \text{ convex on } S & \Rightarrow & f \text{ quasi-convex on } S & \not\Rightarrow & f \text{ convex on } S \end{array}$$

Proof:

We prove only the part ,,f concave on $S \Rightarrow f$ quasi-concave on S".

For ,,f quasi-concave on $S \not\Rightarrow f$ concave on S" look at the function $f(x_1, x_2) = x_1 x_2$ on $S = \mathbb{R}^2_{++}$, which is quasi-concave but not concave.

Let f be concave on S and $\mathbf{a}, \mathbf{b} \in S$.

If $f(\mathbf{a}) \geq c$ and $f(\mathbf{b}) \geq c$ then we have by the definition of concavity:

$$f(t\mathbf{a} + (1-t)\mathbf{b}) \;\geq\; t\underbrace{f(\mathbf{a})}_{\geq c} + (1-t)\underbrace{f(\mathbf{b})}_{\geq c} \;\geq\; tc + (1-t)c \;=\; c$$

□

Question 5.1 *Prove that the function $f(x_1, x_2) = x_1 x_2$ is quasi-concave on $S = \mathbb{R}^2_{++}$.*

Question 5.2 *Let $\lambda_1, \lambda_2 \in \mathbb{R}$. Verify that the function $f(x_1, x_2) = \lambda_1 x_1 + \lambda_2 x_2$ is concave, convex, quasi-concave and quasi-convex. Is it strictly any of these? Sketch its level sets.*

Level sets of concave functions Let $y = f(\mathbf{x}) = f(x_1, x_2)$ be a concave function on the convex set $S = \mathbb{R}^2_{++}$. What can we say about the properties of the level sets $f(x_1, x_2) = c$? Generally (almost) nothing!

With the implicit function theorem we know:

$$x_2'(x_1) \;=\; -\frac{f_{x_1}(x_1, x_2)}{f_{x_2}(x_1, x_2)}$$

and **if** $f_{x_1}, f_{x_2} > 0$ on S, then $x_2'(x_1)$ is negative and $x_2(x_1)$ is a decreasing function for $x_1 \in \mathbb{R}_+$.

The second derivative of $x_2(x_1)$ is

$$x_2''(x_1) \;=\; -\frac{f_{x_1x_1} + 2f_{x_1x_2} \cdot x_2' + f_{x_2x_2} \cdot (x_2')^2}{f_{x_2}}.$$

- **If** $f_{x_1}, f_{x_2} > 0$ and f is concave on S (which means, that $\nabla^2 f$ is negative semi-definite or by Hurwitz $f_{x_1x_1} \leq 0$, $f_{x_2x_2} \leq 0$ and $f_{x_1x_1} f_{x_2x_2} - f^2_{x_1x_2} \geq 0$) then

$$x_2''(x_1) \;=\; -\frac{\overbrace{f_{x_1x_1}}^{\leq 0} + 2f_{x_1x_2} \cdot \overbrace{x_2'}^{<0} + \overbrace{f_{x_2x_2}}^{\leq 0} \cdot \overbrace{(x_2')^2}^{>0}}{\underbrace{f_{x_2}}_{>0}}.$$

  Hence $x_2''(x_1)$ seems to be positive and $x_2(x_1)$ should be a convex function in one variable. But we do not know the sign of $f_{x_1x_2}$!

- **If in addition** f is a homogeneous function of degree g, then we know (Euler's relation) that

$$\begin{aligned} g \cdot f(x_1, x_2) &= x_1\, f_{x_1}(x_1, x_2) + x_2\, f_{x_2}(x_1, x_2) \\ g \cdot (g-1) \cdot f(x_1, x_2) &= x_1^2\, f_{x_1x_1}(x_1, x_2) + 2\, x_1\, x_2\, f_{x_1x_2}(x_1, x_2) + x_2^2\, f_{x_2x_2}(x_1, x_2). \end{aligned}$$

  For positive functions on $S = \mathbb{R}^2_{++}$ (this means $f > 0$) of degree $d \geq 1$ the derivative $f_{x_1x_2}$ should be positive!?

6 Local minima in open sets

6.1 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.

6.2 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 local minimizer of f, it is clear that $t = 0$ is a minimizer of g, such that $g'(0) = 0$. We will try to re-express this result in terms of the original function f:

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

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

and

$$0 \; = \; g'(0) \quad = \quad \nabla f(\mathbf{x}^*)^T \cdot \mathbf{v}$$

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

$$\boxed{\mathbf{x}^* \text{ is a local minimizer (maximizer)} \quad \Longrightarrow \quad \nabla f(\mathbf{x}^*) \; = \; \mathbf{0}}$$

6.3 Second-order necessary condition for optimality

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

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

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 minimizer of f then $g(t)$ has a local minimum in $t = 0$. Hence

$$0 \;\le\; g''(0) \;=\; \mathbf{v}^T \, \nabla^2 f(\mathbf{x}^*) \, \mathbf{v} \;=\; (v_1 \; v_2 \; \dots \; v_n) \begin{pmatrix} f_{x_1x_1}(\mathbf{x}^*) & \dots & f_{x_1x_n}(\mathbf{x}^*) \\ \vdots & \vdots & \vdots \\ f_{x_nx_1}(\mathbf{x}^*) & \dots & f_{x_nx_n}(\mathbf{x}^*) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

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

$\mathbf{x}^*$ is a local minimizer (maximizer) $\Longrightarrow$ $\nabla^2 f(\mathbf{x}^*)$ is positive (negative) semi-definite