15 Scalar-Valued Functions of Two Variables

This chapter is devoted to differential calculus of functions of two variables. In particular we will study geometrical objects such as tangents and tangent planes, maxima and minima, as well as linear and quadratic approximations. The restriction to two variables has been made for simplicity of presentation. All ideas in this and the next chapter can easily be extended (although with slightly more notational effort) to the case of n variables.

We begin by studying the graph of a function with the help of vertical cuts and level sets. As a further tool we introduce partial derivatives, which describe the rate of change of the function in the direction of the coordinate axes. Finally the notion of the Fréchet derivative allows us to define the tangent plane to the graph. As for functions of one variable the Taylor formula plays a central role. We use it, e.g., to determine extrema of functions of two variables.

In the entire chapter D denotes a subset of $\mathbb{R}^2$, and

$$f : D \subset \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto z = f(x, y)$$

denotes a *scalar-valued* function of two variables. Details of vector and matrix algebra used in this chapter can be found in Appendices A and B.

15.1 Graph and Partial Mappings

The *graph*

$$G = \big\{(x, y, z) \in D \times \mathbb{R}\,;\ z = f(x, y)\big\} \subset \mathbb{R}^3$$

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*, Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7_15

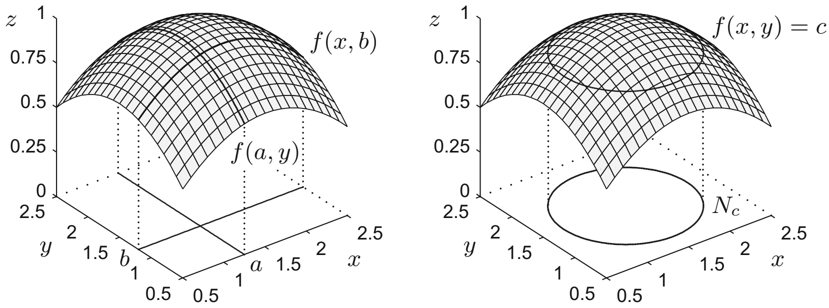


Fig. 15.1 Graph of a function as surface in space with coordinate curves (left) and level curve N_c (right)

of a function of two variables $f : D \to \mathbb{R}$ is a surface in space, if f is sufficiently regular. To describe the properties of this surface we consider particular curves on it.

The *partial mappings*

$$x \mapsto f(x, b), \qquad y \mapsto f(a, y)$$

are obtained by fixing one of the two variables $y = b$ or $x = a$. The partial mappings can be used to introduce the space curves

$$x \mapsto \begin{bmatrix} x \\ b \\ f(x, b) \end{bmatrix}, \qquad y \mapsto \begin{bmatrix} a \\ y \\ f(a, y) \end{bmatrix}.$$

These curves lie on the graph G of the function and are called *coordinate curves*. Geometrically they are obtained as the intersection of G with the vertical planes $y = b$ and $x = a$ respectively, see Fig. 15.1, left.

The *level curves* are the projections of the intersections of the graph G with the horizontal planes $z = c$ to the (x, y)-plane,

$$N_c = \big\{(x, y) \in D\ ;\ f(x, y) = c\big\},$$

see Fig. 15.1, right. The set N_c is called level curve at level c.

Example 15.1 The graph of the quadratic function

$$f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$$

describes a surface in space which is shaped like a saddle and called *hyperbolic paraboloid*. Figure 15.2 shows the graph of $z = x^2/4 - y^2/5$ with coordinate curves (left) as well as some level curves (right).

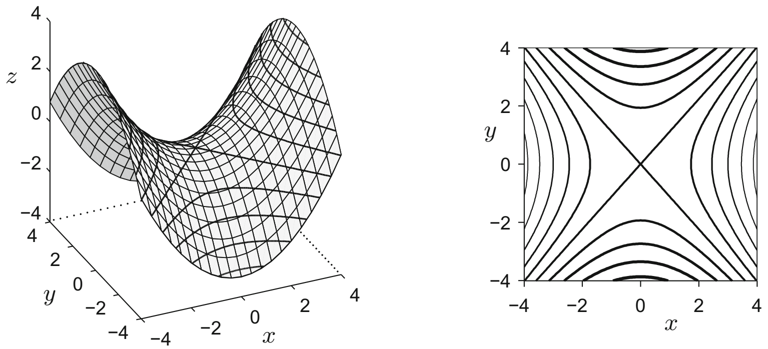


Fig. 15.2 The picture on the left shows the graph of the function $z = x^2/4 - y^2/5$ with coordinate curves. Furthermore, it shows the intersections with the planes $z = c$ for selected values of c. The picture on the right illustrates the level curves of the function for the same values of c (lower levels correspond to thicker lines). The two intersecting straight lines are the level curves at level $c = 0$

Experiment 15.2 With the help of the MATLAB program `mat15_1.m` visualise the elliptic paraboloid $z = x^2 + 2y^2 - 4x + 1$. Choose a suitable domain D and plot the graph and some level curves.

15.2 Continuity

Like for functions in one variable (see Chap. 6) we characterise the continuity of functions of two variables by means of sequences. Thus we need the concept of convergence of vector-valued sequences.

Let $(\mathbf{a}_n)_{n\geq 1} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, ...)$ be a sequence of points in D with terms

$$\mathbf{a}_n = (a_n, b_n) \in D \subset \mathbb{R}^2.$$

The sequence $(\mathbf{a}_n)_{n\geq 1}$ is said to *converge* to $\mathbf{a} = (a, b) \in D$ as $n \to \infty$, if and only if both components of the sequence converge, i.e.

$$\lim_{n\to\infty} a_n = a \qquad \text{and} \qquad \lim_{n\to\infty} b_n = b.$$

This is denoted by

$$(a_n, b_n) = \mathbf{a}_n \to \mathbf{a} = (a, b) \quad \text{as}\, n \to \infty \qquad \text{or} \qquad \lim_{n\to\infty} \mathbf{a}_n = \mathbf{a}.$$

Otherwise the sequence is called *divergent*.

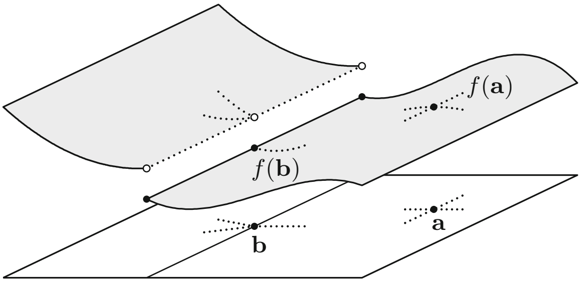


Fig. 15.3 A function which is discontinuous along a straight line. For every sequence $(\mathbf{a}_n)$ which converges to $\mathbf{a}$, the images of the sequence $\big(f(\mathbf{a}_n)\big)$ converge to $f(\mathbf{a})$. For the point $\mathbf{b}$, however, this does not hold; f is discontinuous at that point

An example of a convergent vector-valued sequence is

$$\lim_{n\to\infty}\left(\frac{1}{n}, \frac{2n}{3n+4}\right) = \left(0, \frac{2}{3}\right).$$

Definition 15.3 A function $f : D \to \mathbb{R}$ is called *continuous* at the point $\mathbf{a} \in D$, if

$$\lim_{n\to\infty} f(\mathbf{a}_n) = f(\mathbf{a})$$

for all sequences $(\mathbf{a}_n)_{n\geq 1}$ which converge to $\mathbf{a}$ in D.

For continuous functions, the limit and the function sign can be interchanged. Figure 15.3 shows a function which is discontinuous along a straight line but continuous everywhere else.

15.3 Partial Derivatives

The partial derivatives of a function of two variables are the derivatives of the partial mappings.

Definition 15.4 Let $D \subset \mathbb{R}^2$ be open, $f : D \to \mathbb{R}$ and $\mathbf{a} = (a, b) \in D$. The function f is called *partially differentiable with respect to x* at the point $\mathbf{a}$, if the limit

$$\frac{\partial f}{\partial x}(a, b) = \lim_{x\to a} \frac{f(x, b) - f(a, b)}{x - a}$$

exists. It is called *partially differentiable with respect to y* at the point $\mathbf{a}$, if the limit

$$\frac{\partial f}{\partial y}(a, b) = \lim_{y\to b} \frac{f(a, y) - f(a, b)}{y - b}$$

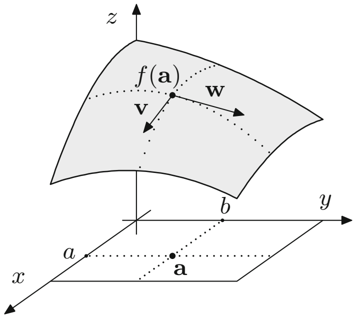


Fig. 15.4 Geometric interpretation of partial derivatives

exists. The expressions

$$\frac{\partial f}{\partial x}(a,b) \quad \text{and} \quad \frac{\partial f}{\partial y}(a,b)$$

are called *partial derivatives* of f with respect to x and y, respectively, at the point (a, b). Further f is called *partially differentiable* at $\mathbf{a}$, if both partial derivatives exist.

Another notation for partial derivatives at the point (x, y) is

$$\frac{\partial f}{\partial x}(x,y) = \frac{\partial}{\partial x} f(x,y) = \partial_1 f(x,y)$$

and likewise

$$\frac{\partial f}{\partial y}(x,y) = \frac{\partial}{\partial y} f(x,y) = \partial_2 f(x,y).$$

Geometrically, partial derivatives can be interpreted as slopes of the tangents to the coordinate curves $x \mapsto [x, b, f(x, b)]^\mathsf{T}$ and $y \mapsto [a, y, f(a, y)]^\mathsf{T}$, see. Fig. 15.4.

The two tangent vectors $\mathbf{v}$ and $\mathbf{w}$ to the coordinate curves at the point $(a, b, f(a, b))$ can therefore be represented as

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ \dfrac{\partial f}{\partial x}(a,b) \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 0 \\ 1 \\ \dfrac{\partial f}{\partial y}(a,b) \end{bmatrix}.$$

Since partial differentiation is nothing else but ordinary differentiation with respect to one variable (while fixing the other one), the usual rules of differentiation apply, e.g. the product rule

$$\frac{\partial}{\partial y}\Big(f(x,y)\cdot g(x,y)\Big) = \frac{\partial f}{\partial y}(x,y)\cdot g(x,y) + f(x,y)\cdot\frac{\partial g}{\partial y}(x,y).$$

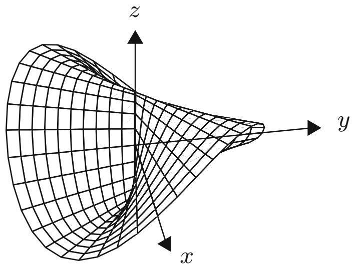


Fig. 15.5 Partially differentiable, discontinuous function

Example 15.5 Let $r : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \sqrt{x^2 + y^2}$. This function is everywhere partially differentiable with the exception of $(x, y) = (0, 0)$. The partial derivatives are

$$\frac{\partial r}{\partial x}(x, y) = \frac{1}{2}\frac{2x}{\sqrt{x^2 + y^2}} = \frac{x}{r(x, y)}, \qquad \frac{\partial r}{\partial y}(x, y) = \frac{1}{2}\frac{2y}{\sqrt{x^2 + y^2}} = \frac{y}{r(x, y)}.$$

In maple one can use the commands `diff` and `Diff` in order to calculate partial derivatives, e.g. in the above example:

```
r:=sqrt(x^2+y^2);
diff(r,x);
```

Remark 15.6 In contrast to functions in one variable (see Application 7.16), partial differentiability does not imply continuity

$$f \text{ partially differentiable } \not\Rightarrow f \text{ continuous.}$$

An example is given by the function (see Fig. 15.5)

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

This function is everywhere partially differentiable. In particular, at the point $(x, y) = (0, 0)$ one obtains

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{x \to 0} \frac{f(x, 0) - f(0, 0)}{x} = 0 = \lim_{y \to 0} \frac{f(0, y) - f(0, 0)}{y} = \frac{\partial f}{\partial y}(0, 0).$$

However, the function is discontinuous at $(0, 0)$. In order to see this, we choose two sequences which converge to $(0, 0)$:

$$\mathbf{a}_n = \left(\tfrac{1}{n}, \tfrac{1}{n}\right) \qquad \text{and} \qquad \mathbf{c}_n = \left(\tfrac{1}{n}, -\tfrac{1}{n}\right).$$

We have

$$\lim_{n\to\infty} f(\mathbf{a}_n) = \lim_{n\to\infty} \frac{1/n^2}{2/n^2} = \frac{1}{2},$$

but also

$$\lim_{n\to\infty} f(\mathbf{c}_n) = \lim_{n\to\infty} \frac{-1/n^2}{2/n^2} = -\frac{1}{2}.$$

The limits do not coincide, in particular, they differ from $f(0,0) = 0$.

Experiment 15.7 Visualise the function given in Remark 15.6 with the help of MATLAB and maple. Using the command

```
plot3d(-x*y/(x^2+y^2), x=-1..1, y=-1..1, shading=zhue);
```

the corresponding plot can be obtained in maple.

Higher-order partial derivatives. Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$ partially differentiable. The assignments

$$\frac{\partial f}{\partial x} : D \to \mathbb{R} \qquad \text{and} \qquad \frac{\partial f}{\partial y} : D \to \mathbb{R}$$

define themselves scalar-valued functions of two variables. If these functions are also partially differentiable, then f is called *twice* partially differentiable. The notation in this case is

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \qquad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right), \quad \text{etc.}$$

Note that there are four partial derivatives of second order.

Definition 15.8 A function $f : D \to \mathbb{R}$ is k-times *continuously (partially) differentiable*, denoted $f \in \mathcal{C}^k(D)$, if f is k-times partially differentiable and all partial derivatives up to order k are continuous.

Example 15.9 The function $f(x, y) = \mathrm{e}^{xy^2}$ is arbitrarily often partially differentiable, $f \in \mathcal{C}^\infty(D)$, and the following holds

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y) &= \mathrm{e}^{xy^2} y^2, \\
\frac{\partial f}{\partial y}(x, y) &= \mathrm{e}^{xy^2} 2xy, \\
\frac{\partial^2 f}{\partial x^2}(x, y) &= \mathrm{e}^{xy^2} y^4,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial y^2}(x,y) &= \mathrm{e}^{xy^2}(4x^2y^2+2x),\\
\frac{\partial^2 f}{\partial y \partial x}(x,y) &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}(x,y)\right) = \mathrm{e}^{xy^2}(2xy^3+2y),\\
\frac{\partial^2 f}{\partial x \partial y}(x,y) &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}(x,y)\right) = \mathrm{e}^{xy^2}(2xy^3+2y).
\end{aligned}$$

The identity

$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y)$$

which is evident in this example is generally valid for twice *continuously* differentiable functions f. This observation is also true for higher derivatives: For k-times continuously differentiable functions the order of differentiation of the kth partial derivatives is irrelevant (Theorem of Schwarz[1]), see [3, Chap. 15, Theorem 1.1].

15.4 The Fréchet Derivative

Our next topic is the study of a *simultaneous* variation of both variables of the function. This leads us to the notion of the Fréchet[2] derivative. For functions of one variable, $\varphi : \mathbb{R} \to \mathbb{R}$, the derivative was defined by the limit

$$\varphi'(a) = \lim_{x \to a} \frac{\varphi(x) - \varphi(a)}{x - a}.$$

For functions of two variables this expression does not make sense anymore as one cannot divide by vectors. We therefore will make use of the equivalent definition of the derivative as a linear approximation

$$\varphi(x) = \varphi(a) + A \cdot (x - a) + R(x, a)$$

with $A = \varphi'(a)$ and the remainder term $R(x, a)$ satisfying

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

This formula can be generalised to functions of two variables.

[1] H.A. Schwarz, 1843–1921.
[2] M. Fréchet, 1878–1973.

Definition 15.10 Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$. The function f is called *Fréchet differentiable* at the point $(a, b) \in D$, if there exists a *linear* mapping $A : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(x, y) = f(a, b) + A(x - a, y - b) + R(x, y; a, b)$$

with a remainder $R(x, y; a, b)$ fulfilling the condition

$$\lim_{(x,y)\to(a,b)} \frac{R(x, y; a, b)}{\sqrt{(x-a)^2 + (y-b)^2}} = 0.$$

The linear mapping A is called *derivative* of f at the point (a, b). Instead of A we also write $\mathrm{D}f(a, b)$. The (1×2)-matrix of the linear mapping is called *Jacobian*[3] of f. We denote it by $f'(a, b)$.

The question whether the derivative of a function is unique and how it can be calculated, is answered in the following proposition.

Proposition 15.11 *Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$. If f is Fréchet differentiable at $(x, y) \in D$, then f is also partially differentiable at (x, y) and*

$$f'(x, y) = \left[\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right].$$

The components of the Jacobian are the partial derivatives. In particular, the Jacobian and consequently the Fréchet derivative are unique.

Proof Exemplarily, we compute the second component and show that

$$\big(f'(x, y)\big)_2 = \frac{\partial f}{\partial y}(x, y).$$

Since f is Fréchet differentiable at (x, y), it holds that

$$f(x, y + h) = f(x, y) + f'(x, y) \begin{bmatrix} 0 \\ h \end{bmatrix} + R(x, y + h; x, y).$$

Therefore

$$\frac{f(x, y + h) - f(x, y)}{h} - \big(f'(x, y)\big)_2 = \frac{R(x, y + h; x, y)}{h} \to 0 \qquad \text{as } h \to 0.$$

Consequently f is partially differentiable with respect to y, and the second component of the Jacobian is the partial derivative of f with respect to y.

[3] C.G.J. Jacobi, 1804–1851.

The next proposition follows immediately from the identity

$$\begin{aligned}\lim_{(x,y)\to(a,b)} f(x,y) &= \lim_{(x,y)\to(a,b)} \Big(f(a,b) + \mathrm{D}f(a,b)(x-a, y-b) + R(x,y;a,b) \Big)\\ &= f(a,b).\end{aligned}$$

Proposition 15.12 *If f is Fréchet differentiable then f is continuous.* □

In particular, the function

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2}, & (x,y) \neq (0,0),\\ 0, & (x,y) = (0,0) \end{cases}$$

is not Fréchet differentiable at the point (0, 0).

Fréchet differentiability follows from partial differentiability under certain regularity assumptions. In fact, one can show that a *continuously* partially differentiable function is Fréchet differentiable, see [4, Chap. 7, Theorem 7.12].

Example 15.13 The function $f : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto x^2 \mathrm{e}^{3y}$ is Fréchet differentiable, its derivative is

$$f'(x,y) = \big[2x\mathrm{e}^{3y}, 3x^2\mathrm{e}^{3y}\big] = x\mathrm{e}^{3y}\,[2, 3x].$$

Example 15.14 The affine function $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$f(x,y) = \alpha x + \beta y + \gamma = [\alpha, \beta] \begin{bmatrix} x \\ y \end{bmatrix} + \gamma$$

is Fréchet differentiable and $f'(x,y) = [\alpha, \beta]$.

Example 15.15 The quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ with

$$\begin{aligned} f(x,y) &= \alpha x^2 + 2\beta xy + \gamma y^2 + \delta x + \varepsilon y + \zeta\\ &= [x, y] \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + [\delta, \varepsilon] \begin{bmatrix} x \\ y \end{bmatrix} + \zeta \end{aligned}$$

is Fréchet differentiable with the Jacobian

$$f'(x,y) = [2\alpha x + 2\beta y + \delta, 2\beta x + 2\gamma y + \varepsilon] = 2[x, y] \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} + [\delta, \varepsilon].$$

The chain rule. Now we are in the position to generalise the chain rule to the case of two variables.

Proposition 15.16 *Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R} : (x, y) \mapsto f(x, y)$ Fréchet differentiable. Furthermore let $I \subset \mathbb{R}$ be an open interval and $\phi, \psi : I \to \mathbb{R}$ differentiable. Then the composition of functions*

$$F : I \to \mathbb{R} : t \mapsto F(t) = f\big(\phi(t), \psi(t)\big)$$

is also differentiable and

$$\frac{\mathrm{d}F}{\mathrm{d}t}(t) = \frac{\partial f}{\partial x}\big(\phi(t), \psi(t)\big)\frac{\mathrm{d}\phi}{\mathrm{d}t}(t) + \frac{\partial f}{\partial y}\big(\phi(t), \psi(t)\big)\frac{\mathrm{d}\psi}{\mathrm{d}t}(t).$$

Proof From the Fréchet differentiability of f it follows that

$$\begin{aligned} F(t+h) - F(t) &= f\big(\phi(t+h), \psi(t+h)\big) - f\big(\phi(t), \psi(t)\big) \\ &= f'\big(\phi(t), \psi(t)\big)\begin{bmatrix}\phi(t+h) - \phi(t) \\ \psi(t+h) - \psi(t)\end{bmatrix} + R\big(\phi(t+h), \psi(t+h); \phi(t), \psi(t)\big). \end{aligned}$$

We divide this expression by h and subsequently examine the limit as $h \to 0$. Let $g(t, h) = \big(\phi(t+h) - \phi(t)\big)^2 + \big(\psi(t+h) - \psi(t)\big)^2$. Then, due to the differentiability of f, ϕ and ψ, we have

$$\lim_{h\to 0} \frac{R\big(\phi(t+h), \psi(t+h); \phi(t), \psi(t)\big)}{\sqrt{g(t,h)}} \cdot \frac{\sqrt{g(t,h)}}{h} = 0.$$

Therefore, the function F is differentiable and the formula stated in the proposition is valid. □

Example 15.17 Let $D \subset \mathbb{R}^2$ be an open set that contains the circle $x^2 + y^2 = 1$ and let $f : D \to \mathbb{R}$ be a differentiable function. Then the restriction F of f to the circle

$$F : \mathbb{R} \to \mathbb{R} : t \mapsto f(\cos t, \sin t)$$

is differentiable as a function of the angle t and

$$\frac{\mathrm{d}F}{\mathrm{d}t}(t) = -\frac{\partial f}{\partial x}\big(\cos t, \sin t\big) \cdot \sin t + \frac{\partial f}{\partial y}\big(\cos t, \sin t\big) \cdot \cos t.$$

For instance, for $f(x, y) = x^2 - y^2$ the derivative is $\frac{\mathrm{d}F}{\mathrm{d}t}(t) = -4\cos t \sin t$.

Interpretation of the Fréchet derivative. Using the Fréchet derivative we obtain, like in the case of one variable, the linear approximation $g(x, y)$ to the graph of the function at (a, b)

$$g(x, y) = f(a, b) + f'(a, b)\begin{bmatrix} x - a \\ y - b\end{bmatrix} \approx f(x, y).$$

Now we want to interpret the plane

$$z = f(a,b) + f'(a,b)\begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

geometrically. For this we use the fact that the components of the Jacobian are the partial derivatives. With that we can write the above equation as

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot(x-a) + \frac{\partial f}{\partial y}(a,b)\cdot(y-b),$$

or alternatively in parametric form ($x-a=\lambda,\ \ y-b=\mu$)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ b \\ f(a,b) \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x}(a,b) \end{bmatrix} + \mu \begin{bmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial y}(a,b) \end{bmatrix}.$$

The plane intersects the graph of f at the point $(a, b, f(a,b))$ and is spanned by the tangent vectors to the coordinate curves. The equation

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot(x-a) + \frac{\partial f}{\partial y}(a,b)\cdot(y-b),$$

consequently describes the *tangent plane* to the graph of f at the point (a,b).

The example shows that the graph of a function which is Fréchet differentiable at the point (x, y) possesses a tangent plane at this point. Note that the existence of tangents to the coordinate curves does *not* imply the existence of a tangent plane, see Remark 15.6.

Example 15.18 We calculate the tangent plane at a point on the northern hemisphere (with radius r)

$$f(x,y) = z = \sqrt{r^2 - x^2 - y^2}.$$

Let $c = f(a,b) = \sqrt{r^2-a^2-b^2}$. The partial derivatives of f at (a,b) are

$$\frac{\partial f}{\partial x}(a,b) = -\frac{a}{\sqrt{r^2-a^2-b^2}} = -\frac{a}{c}, \qquad \frac{\partial f}{\partial y}(a,b) = -\frac{b}{\sqrt{r^2-a^2-b^2}} = -\frac{b}{c}.$$

Therefore, the equation of the tangent plane is

$$z = c - \frac{a}{c}(x-a) - \frac{b}{c}(y-b),$$

or alternatively

$$a(x-a) + b(y-b) + c(z-c) = 0.$$

The last formula actually holds for all points on the surface of the sphere.

15.5 Directional Derivative and Gradient

So far functions $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ were defined on $\mathbb{R}^2$ as a point space. For the purpose of directional derivatives it is useful and customary to write the arguments $(x, y) \in \mathbb{R}^2$ as position vectors $\mathbf{x} = [x, y]^\mathsf{T}$. In this way each function $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ can also be considered as a function of column vectors. We identify these two functions and will not distinguish between $f(x, y)$ and $f(\mathbf{x})$ henceforth.

In Sect. 15.3 we have defined partial derivatives along coordinate axes. Now we want to generalise this concept to differentiation in *any* direction.

Definition 15.19 Let $D \subset \mathbb{R}^2$ be open, $\mathbf{x} = [x, y]^\mathsf{T} \in D$ and $f : D \to \mathbb{R}$. Furthermore let $\mathbf{v} \in \mathbb{R}^2$ with $\|\mathbf{v}\| = 1$. The limit

$$\begin{aligned}\partial_{\mathbf{v}} f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{v}}(\mathbf{x}) &= \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}\\ &= \lim_{h\to 0} \frac{f(x + hv_1, y + hv_2) - f(x, y)}{h}\end{aligned}$$

(in case it exists) is called *directional derivative* of f at $\mathbf{x}$ *in direction* $\mathbf{v}$.

The partial derivatives are special cases of the directional derivative, namely the derivatives in direction of the coordinate axes.

The directional derivative $\partial_{\mathbf{v}} f(\mathbf{x})$ describes the rate of change of the function f at the point $\mathbf{x}$ in the direction of $\mathbf{v}$. Indeed, this can been seen from the following. Consider the straight line $\{\mathbf{x} + t\mathbf{v} \mid t \in \mathbb{R}\} \subset \mathbb{R}^2$ and the function

$$g(t) = f(\mathbf{x} + t\mathbf{v}) \qquad (f \text{ restricted to this straight line})$$

with $g(0) = f(\mathbf{x})$. Then

$$g'(0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = \lim_{h\to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} = \partial_{\mathbf{v}} f(\mathbf{x}).$$

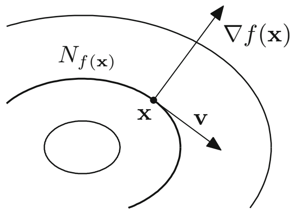
Next we clarify how the directional derivative can be computed. For that we need the following definition.

Definition 15.20 Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$ partially differentiable. The vector

$$\nabla f(x, y) = \begin{bmatrix} \dfrac{\partial f}{\partial x}(x, y) \\ \dfrac{\partial f}{\partial y}(x, y) \end{bmatrix} = f'(x, y)^\mathsf{T}$$

is called *gradient* of f.

Fig. 15.6 Geometric interpretation of ∇f



Proposition 15.21 *Let $D \subset \mathbb{R}^2$ be open, $\mathbf{v} = [v_1, v_2]^\mathsf{T} \in \mathbb{R}^2$, $\|\mathbf{v}\| = 1$ and $f : D \to \mathbb{R}$ Fréchet differentiable at $\mathbf{x} = [x, y]^\mathsf{T}$. Then*

$$\partial_{\mathbf{v}} f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle = f'(x, y)\,\mathbf{v} = \frac{\partial f}{\partial x}(x, y)\, v_1 + \frac{\partial f}{\partial y}(x, y)\, v_2.$$

Proof Since f is Fréchet differentiable at $\mathbf{x}$, it holds that

$$f(\mathbf{x} + h\mathbf{v}) = f(\mathbf{x}) + f'(\mathbf{x}) \cdot h\mathbf{v} + R(x + hv_1, y + hv_2; x, y)$$

and hence

$$\frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} = f'(\mathbf{x}) \cdot \mathbf{v} + \frac{R(x + hv_1, y + hv_2; x, y)}{h}.$$

Letting $h \to 0$ proves the desired assertion. □

Proposition 15.22 (Geometric interpretation of ∇) *Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$ continuously differentiable at $\mathbf{x} = (x, y)$ with $f'(\mathbf{x}) \neq [0, 0]$. Then $\nabla f(\mathbf{x})$ is* perpendicular *to the level curve $N_{f(\mathbf{x})} = \{\widetilde{\mathbf{x}} \in \mathbb{R}^2 \; ; \; f(\widetilde{\mathbf{x}}) = f(\mathbf{x})\}$ and points in direction of the* steepest ascent *of f, see Fig. 15.6.*

Proof Let $\mathbf{v}$ be a tangent vector to the level curve at the point $\mathbf{x}$. From the implicit function theorem (see [4, Chap. 14.1]) it follows that $N_{f(\mathbf{x})}$ can be parametrised as a differentiable curve $\gamma(t) = [x(t), y(t)]^\mathsf{T}$, with

$$\gamma(0) = \mathbf{x} \qquad \text{and} \qquad \dot{\gamma}(0) = \mathbf{v}$$

in a neighbourhood of $\mathbf{x}$. Thus, for all t near $t = 0$,

$$f\big(\gamma(t)\big) = f(\mathbf{x}) = \text{const.}$$

Since f and γ are differentiable, it follows from the chain rule (Proposition 15.16) that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} f\big(\gamma(t)\big)\Big|_{t=0} = f'\big(\gamma(0)\big)\,\dot{\gamma}(0) = \langle \nabla f(\mathbf{x}), \mathbf{v} \rangle$$

because $\gamma(0) = \mathbf{x}$ and $\dot{\gamma}(0) = \mathbf{v}$. Hence $\nabla f(\mathbf{x})$ is perpendicular to $\mathbf{v}$. Let $\mathbf{w} \in \mathbb{R}^2$ be a further unit vector. Then

$$\partial_{\mathbf{w}} f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{w}}(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{w} \rangle = \|\nabla f(\mathbf{x})\| \cdot \|\mathbf{w}\| \cdot \cos \sphericalangle,$$

where $\sphericalangle$ denotes the angle enclosed by $\nabla f(\mathbf{x})$ and $\mathbf{w}$. From this formula one deduces that $\partial_{\mathbf{w}} f(\mathbf{x})$ is maximal if and only if $\cos \sphericalangle = 1$, which means $\nabla f(\mathbf{x}) = \lambda \mathbf{w}$ for some $\lambda > 0$. $\square$

Example 15.23 Let $f(x, y) = x^2 + y^2$. Then $\nabla f(x, y) = 2[x, y]^{\mathsf{T}}$.

15.6 The Taylor Formula in Two Variables

Let $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables. In the following calculation we assume that f is at least three times continuously differentiable. In order to expand $f(x + h, y + k)$ into a Taylor series in a neighbourhood of (x, y), we first fix the second variable and expand with respect to the first:

$$f(x+h, y+k) = f(x, y+k) + \frac{\partial f}{\partial x}(x, y+k) \cdot h + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(x, y+k) \cdot h^2 + \mathcal{O}(h^3).$$

Then we also expand the terms on the right-hand side with respect to the second variable (while fixing the first one):

$$\begin{aligned}
f(x, y + k) &= f(x, y) + \frac{\partial f}{\partial y}(x, y) \cdot k + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(x, y) \cdot k^2 + \mathcal{O}(k^3), \\
\frac{\partial f}{\partial x}(x, y + k) &= \frac{\partial f}{\partial x}(x, y) + \frac{\partial^2 f}{\partial y \partial x}(x, y) \cdot k + \mathcal{O}(k^2), \\
\frac{\partial^2 f}{\partial x^2}(x, y + k) &= \frac{\partial^2 f}{\partial x^2}(x, y) + \mathcal{O}(k).
\end{aligned}$$

Inserting these expressions into the equation above, we obtain

$$\begin{aligned}
f(x + h, y + k) = f(x, y) &+ \frac{\partial f}{\partial x}(x, y) \cdot h + \frac{\partial f}{\partial y}(x, y) \cdot k \\
&+ \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(x, y) \cdot h^2 + \frac{1}{2} \frac{\partial^2 f}{\partial y^2}(x, y) \cdot k^2 + \frac{\partial^2 f}{\partial y \partial x}(x, y) \cdot hk \\
&+ \mathcal{O}(h^3) + \mathcal{O}(h^2 k) + \mathcal{O}(hk^2) + \mathcal{O}(k^3).
\end{aligned}$$

In matrix-vector notation we can also write this equation as

$$f(x + h, y + k) = f(x, y) + f'(x, y) \begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{2} [h, k] \cdot H_f(x, y) \begin{bmatrix} h \\ k \end{bmatrix} + \ldots$$

with the *Hessian matrix*[4]

$$H_f(x,y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2}(x,y) & \dfrac{\partial^2 f}{\partial y \partial x}(x,y) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x,y) & \dfrac{\partial^2 f}{\partial y^2}(x,y) \end{bmatrix}$$

collecting the second-order partial derivatives. By the above assumptions, these derivatives are continuous. Thus the Hessian matrix is symmetric due to Schwarz's theorem.

Example 15.24 We compute the second-order approximation to the function $f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto x^2 \sin y$ at the point $(a, b) = (2, 0)$. The partial derivatives are

	f	$\frac{\partial f}{\partial x}$	$\frac{\partial f}{\partial y}$	$\frac{\partial^2 f}{\partial x^2}$	$\frac{\partial^2 f}{\partial y \partial x}$	$\frac{\partial^2 f}{\partial y^2}$
General	$x^2 \sin y$	$2x \sin y$	$x^2 \cos y$	$2 \sin y$	$2x \cos y$	$-x^2 \sin y$
At (2, 0)	0	0	4	0	4	0

Therefore, the quadratic approximation $g(x, y) \approx f(x, y)$ is given by the formula

$$\begin{aligned} g(x,y) &= f(2,0) + f'(2,0)\begin{bmatrix} x-2 \\ y \end{bmatrix} + \frac{1}{2}[x-2, y] \cdot H_f(2,0)\begin{bmatrix} x-2 \\ y \end{bmatrix} \\ &= 0 + [0,4]\begin{bmatrix} x-2 \\ y \end{bmatrix} + \frac{1}{2}[x-2, y]\begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix}\begin{bmatrix} x-2 \\ y \end{bmatrix} \\ &= 4y + 4y(x-2) \;=\; 4y(x-1). \end{aligned}$$

15.7 Local Maxima and Minima

Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$. In this section we investigate the graph of the function f with respect to maxima and minima.

Definition 15.25 The scalar function f has a *local maximum* (respectively, *local minimum*) at $(a, b) \in D$, if

$$f(x,y) \le f(a,b) \qquad \text{(respectively, } f(x,y) \ge f(a,b)\text{)}.$$

[4]L.O. Hesse, 1811–1874.

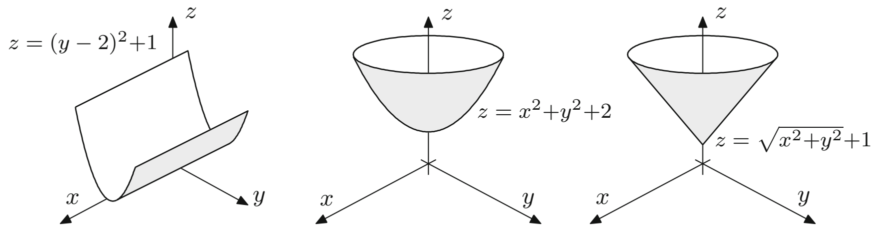


Fig. 15.7 Local and isolated local minima. The function in the picture on the left has local minima along the straight line $y = 2$. The minima are not isolated. The function in the middle picture has an isolated minimum at $(x, y) = (0, 0)$. This minimum is even a global minimum. Finally, the function in the picture on the right-hand side has also an isolated minimum at $(x, y) = (0, 0)$. However, the function is not differentiable at that point

for all (x, y) in a neighbourhood of (a, b). The maximum (minimum) is called *isolated*, if (a, b) is the only point in a neighbourhood with this property.

Figure 15.7 shows a few typical examples. One observes that the existence of a horizontal tangent plane is a necessary condition for *extrema* (i.e. maxima or minima) of *differentiable* functions.

Proposition 15.26 *Let f be partially differentiable. If f has a local maximum or minimum at $(a, b) \in D$, then the partial derivatives vanish at (a, b):*

$$\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0.$$

If, in addition, f is Fréchet differentiable, then $f'(a, b) = [0, 0]$, i.e. f has a horizontal tangent plane at (a, b).

Proof Due to the assumptions, the function $g(h) = f(a + h, b)$ has an extremum at $h = 0$. Thus, Proposition 8.2 implies

$$g'(0) = \frac{\partial f}{\partial x}(a, b) = 0.$$

Likewise one can show that $\frac{\partial f}{\partial y}(a, b) = 0$. □

Definition 15.27 Let f be a Fréchet differentiable function with $f'(a, b) = [0, 0]$. Then (a, b) is called a *stationary point* of f.

Stationary points are consequently candidates for extrema. Conversely, not all stationary points are extrema, they can also be *saddle points*. We call (a, b) a saddle point of f, if there is a vertical cut through the graph which has a local maximum at

(a, b), and a second vertical cut which has a local minimum at (a, b), see, for example, Fig. 15.2. In order to decide what is the case, one resorts to Taylor expansion, similarly as for functions of one variable.

Let $\mathbf{a} = [a, b]^\mathsf{T}$ be a stationary point of f and $\mathbf{v} \in \mathbb{R}^2$ any unit vector. We investigate the behaviour of f, restricted to the straight line $\mathbf{a} + \lambda\mathbf{v},\ \lambda \in \mathbb{R}$. Taylor expansion shows that

$$f(\mathbf{a} + \lambda\mathbf{v}) = f(\mathbf{a}) + f'(\mathbf{a}) \cdot \lambda\mathbf{v} + \tfrac{1}{2}\lambda^2\mathbf{v}^\mathsf{T} H_f(\mathbf{a})\,\mathbf{v} + \mathcal{O}(\lambda^3).$$

Since $\mathbf{a}$ is a stationary point, it follows that $f'(\mathbf{a}) = [0, 0]$ and consequently

$$\frac{f(\mathbf{a} + \lambda\mathbf{v}) - f(\mathbf{a})}{\lambda^2} = \tfrac{1}{2}\,\mathbf{v}^\mathsf{T} H_f(\mathbf{a})\,\mathbf{v} + \mathcal{O}(\lambda).$$

For small λ the sign on the left-hand side is therefore determined by the sign of $\mathbf{v}^\mathsf{T} H_f(\mathbf{a})\,\mathbf{v}$. We ask how this can be expressed by conditions on $H_f(\mathbf{a})$. Writing

$$H_f(\mathbf{a}) = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} v \\ w \end{bmatrix}$$

we get

$$\mathbf{v}^T H_f(\mathbf{a})\,\mathbf{v} = \alpha v^2 + 2\beta vw + \gamma w^2.$$

For an isolated local minimum this expression has to be positive for all $\mathbf{v} \neq 0$. If $w = 0$ and $v \neq 0$, then $\alpha v^2 > 0$ and therefore necessarily

$$\alpha > 0.$$

If $w \neq 0$, we substitute $v = tw$ with $t \in \mathbb{R}$ and obtain

$$\alpha t^2 w^2 + 2\beta t w^2 + \gamma w^2 > 0,$$

or alternatively (multiplying by $\alpha > 0$ and simplifying by w^2)

$$t^2\alpha^2 + 2t\alpha\beta + \alpha\gamma > 0.$$

Therefore,

$$(t\alpha + \beta)^2 + \alpha\gamma - \beta^2 > 0$$

for all $t \in \mathbb{R}$. The left-hand side is smallest for $t = -\beta/\alpha$. Inserting this we obtain the second condition

$$\det H_f(\mathbf{a}) = \alpha\gamma - \beta^2 > 0$$

in terms of the determinant, see Appendix B.1.

We have thus shown the following result.

Proposition 15.28 *The function f has an isolated local minimum at the stationary point* $\mathbf{a}$*, if the conditions*

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) > 0 \quad \textit{and} \quad \det H_f(\mathbf{a}) > 0$$

are fulfilled.

By replacing f by $-f$ one gets the corresponding result for isolated maxima.

Proposition 15.29 *The function f has an isolated local maximum at the stationary point* $\mathbf{a}$*, if the conditions*

$$\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) < 0 \quad \textit{and} \quad \det H_f(\mathbf{a}) > 0$$

are fulfilled.

In a similar way one can prove the following assertion.

Proposition 15.30 *The stationary point* $\mathbf{a}$ *of the function f is a saddle point, if* $\det H_f(\mathbf{a}) < 0$.

If the determinant of the Hessian matrix equals zero, the behaviour of the function needs to be investigated along vertical cuts. One example is given in Exercise 12.

Example 15.31 We determine the maxima, minima and saddle points of the function $f(x, y) = x^6 + y^6 - 3x^2 - 3y^2$. The condition

$$f'(x, y) = [6x^5 - 6x, 6y^5 - 6y] = [0, 0]$$

gives the following nine stationary points

$$x_1 = 0, \quad x_{2,3} = \pm 1, \quad y_1 = 0, \quad y_{2,3} = \pm 1.$$

The Hessian matrix of the function is

$$H_f(x, y) = \begin{bmatrix} 30x^4 - 6 & 0 \\ 0 & 30y^4 - 6 \end{bmatrix}.$$

Applying the criteria of Propositions 15.28 through 15.30, we obtain the following results: The point $(0, 0)$ is an isolated local maximum of f, the points $(-1, -1)$, $(-1, 1)$, $(1, -1)$ and $(1, 1)$ are isolated local minima, and the points $(-1, 0)$, $(1, 0)$, $(0, -1)$ and $(0, 1)$ are saddle points. The reader is advised to visualise this function with maple.

15.8 Exercises

1. Compute the partial derivatives of the functions

$$f(x, y) = \arcsin\left(\frac{y}{x}\right), \qquad g(x, y) = \log\frac{1}{\sqrt{x^2 + y^2}}\,.$$

Verify your results with maple.

2. Show that the function

$$v(x, t) = \frac{1}{\sqrt{t}}\exp\left(\frac{-x^2}{4t}\right)$$

satisfies the *heat equation*

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2}$$

for $t > 0$ and $x \in \mathbb{R}$.

3. Show that the function $w(x, t) = g(x - kt)$ satisfies the *transport equation*

$$\frac{\partial w}{\partial t} + k\,\frac{\partial w}{\partial x} = 0$$

for any differentiable function g.

4. Show that the function $g(x, y) = \log(x^2 + 2y^2)$ satisfies the equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{1}{2}\,\frac{\partial^2 g}{\partial y^2} = 0$$

for $(x, y) \neq (0, 0)$.

5. Represent the ellipsoid $x^2 + 2y^2 + z^2 = 1$ as graph of a function $(x, y) \mapsto f(x, y)$. Distinguish between positive and negative z-coordinates, respectively. Compute the partial derivatives of f and sketch the level curves of f. Find the direction in which ∇f points.

6. Solve Exercise 5 for the hyperboloid $x^2 + 2y^2 - z^2 = 1$.

7. Compute the directional derivative of the function $f(x, y) = xy$ in the direction $\mathbf{v}$ at the four points $\mathbf{a}_1, \ldots \mathbf{a}_4$, where

$$\mathbf{a}_1 = (1, 2),\ \mathbf{a}_2 = (-1, 2),\ \mathbf{a}_3 = (1, -2),\ \mathbf{a}_4 = (-1, -2) \quad\text{and}\quad \mathbf{v} = \frac{1}{\sqrt{5}}\begin{bmatrix}2\\1\end{bmatrix}.$$

At the given points $\mathbf{a}_1, \ldots \mathbf{a}_4$, determine the direction for which the directional derivative is maximal.

8. Consider the function $f(x,y)=4-x^2-y^2$.

(a) Determine and sketch the level curves $f(x,y)=c$ for $c=4,2,0,-2$ and the graphs of the coordinate curves

$$x \mapsto \begin{bmatrix} x \\ b \\ f(x,b) \end{bmatrix}, \qquad y \mapsto \begin{bmatrix} a \\ y \\ f(a,y) \end{bmatrix}$$

for $a,b=-1,0,1$.

(b) Compute the gradient of f at the point $(1,1)$ and determine the equation of the tangent plane at $(1,1,2)$. Verify that the gradient is perpendicular to the level curve through $(1,1,2)$.

(c) Compute the directional derivatives of f at $(1,1)$ in the directions

$$\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix},\ \mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix},\ \mathbf{v}_3=\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ -1 \end{bmatrix},\ \mathbf{v}_4=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Sketch the vectors $\mathbf{v}_1,\dots,\mathbf{v}_4$ in the (x,y)-plane and interpret the value of the directional derivatives.

9. Consider the function $f(x,y)=y\mathrm{e}^{2x-y}$, where $x=x(t)$ and $y=y(t)$ are differentiable functions satisfying

$$x(0)=2, \qquad y(0)=4, \qquad \dot{x}(0)=-1, \qquad \dot{y}(0)=4.$$

From this information compute the derivative of $z(t)=f\big(x(t),y(t)\big)$ at the point $t=0$.

10. Find all stationary points of the function

$$f(x,y)=x^3-3xy^2+6y.$$

Determine whether they are maxima, minima or saddle points.

11. Find the stationary point of the function

$$f(x,y)=\mathrm{e}^x+y\mathrm{e}^y-x$$

and determine whether it is a maximum, minimum or a saddle point.

12. Investigate the function

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

for local extrema and saddle points. Visualise the graph of the function.
Hint. To study the behaviour of the function at $(0,0)$ consider the partial mappings $f(x,0)$ and $f(0,y)$.

13. Determine for the function

$$f(x,y)=x^2\mathrm{e}^{y/3}(y-3)-\tfrac{1}{2}y^2$$

(a) the gradient and the Hessian matrix;
(b) the second-order Taylor approximation at $(0, 0)$;
(c) all stationary points. Find out whether they are maxima, minima or saddle points.

14. Expand the polynomial $f(x, y) = x^2 + xy + 3y^2$ in powers of $x - 1$ and $y - 2$, i.e. in the form

$$\begin{aligned} f(x, y) = \alpha(x-1)^2 + \beta(x-1)(y-2) + \gamma(y-2)^2 \\ + \delta(x-1) + \varepsilon(y-2) + \zeta. \end{aligned}$$

Hint. Use the second-order Taylor expansion at $(1, 2)$.

15. Compute $(0.95)^{2.01}$ numerically by using the second-order Taylor approximation to the function $f(x, y) = x^y$ at $(1, 2)$.