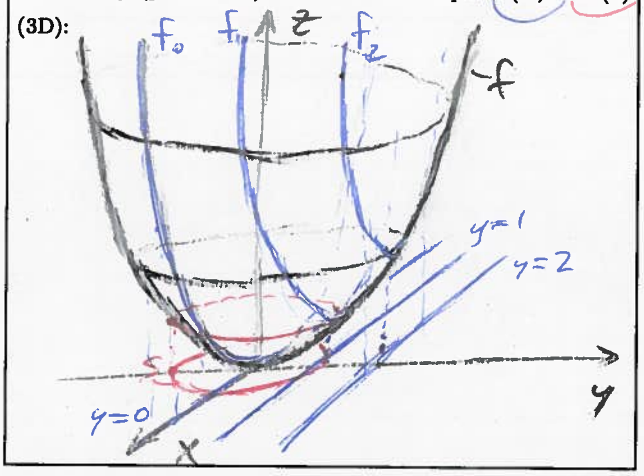
(iv) A very important tool for working with multivariate functions is restriction to lines in the domain, cf. example (ii) below.

Example 2.3. (i) Draw the graph of $f(x, y) = x^2 + y^2$.
Sol.:

Graph of f (paraboloid) with curves from parts (iv) and (v) (3D):



(ii) Find the domain and range of $g(x, y) = 1 + \sqrt{y - x^2}$.
Sol.: The function g is defined on the set

$$D(g) = \{(x, y) \in \mathbb{R}^2 \mid y \geq x^2\} = \{y \geq x^2\},$$

since we need $y - x^2 \geq 0$ to be able to evaluate the square root. On this domain, the argument of the square root takes all non-negative real values $t \in [0, \infty)$, producing non-negative real numbers $\sqrt{t} \in [0, \infty)$. Therefore, the range is $R(g) = [1, \infty)$.

(iii) Find the domain and range of $h(x_1, x_2, x_3) = \sin(x_1 x_2 + x_3)$.
Sol.: The domain is $D = \mathbb{R}^3$ since h can be evaluated at any (x_1, x_2, x_3), and the range of h is $R = [-1, 1]$.

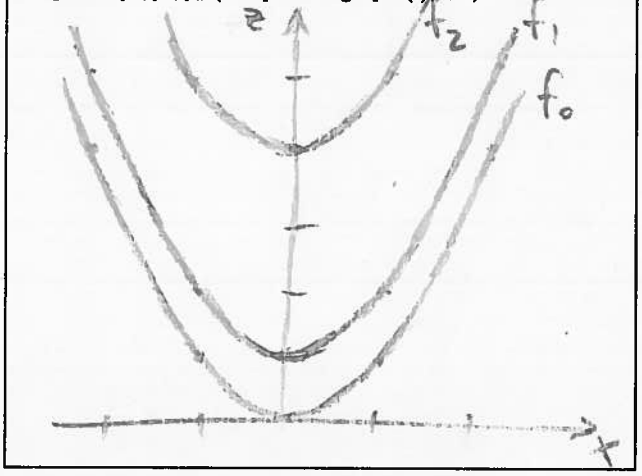
(iv) Draw the graphs of the following one-variable functions

$$f_0(x) = x^2,$$
$$f_1(x) = x^2 + 1,$$
$$f_2(x) = x^2 + 4.$$

Where do these curves appear in the graph in (i)?
Sol.: The functions f_0, f_1, f_2 are restrictions of $f(x, y)$ in (i) to the lines $y = 0, y = 1, y = 2$. This can also be expressed as

$$f_0(x) = f(x, 0) = x^2 + 0^2,$$
$$f_1(x) = f(x, 1) = x^2 + 1^2,$$
$$f_2(x) = f(x, 2) = x^2 + 2^2.$$

Graphs of f_0, f_1, f_2 (compare to graph (i); 2D):

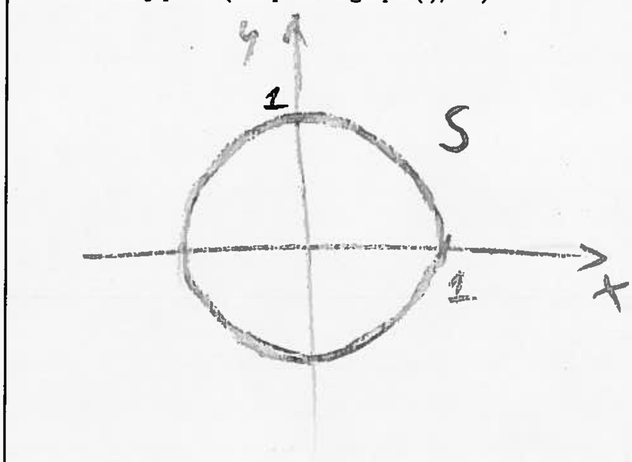


(v) The domain of the function $f(x, y) = x^2 + y^2$ from (i) is $D(f) = \mathbb{R}^2$. Describe the subset S of points in D with $f(x, y) = 1$.
Sol.: The equation

$$x^2 + y^2 = 1$$

describes a circle in the xy-plane:

Curve S in xy-plane (compare to graph (i); 2D):

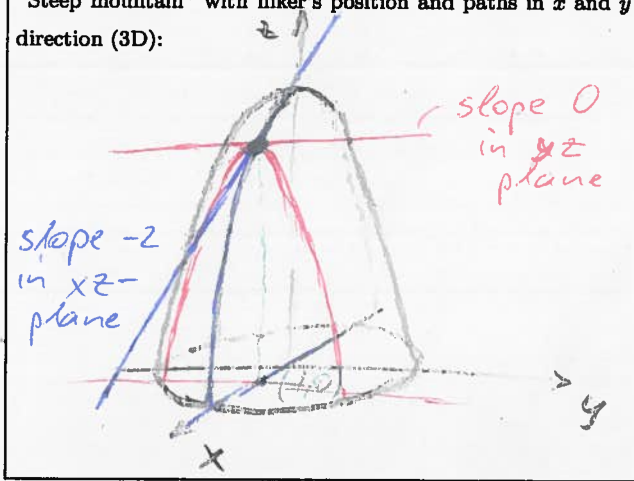


Definition 2.4 (Partial Derivatives). For a function f of two variables, the *partial derivatives* are the functions

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

We also write f_x for the partial derivative $\dfrac{\partial f}{\partial x}$ – that is, the partial derivative of f with

"Steep mountain" with hiker's position and paths in x and y direction (3D):



Suppose she checks her GPS and finds that her coordinates are $(x, y) = (1, 0)$. The slopes at this point in both directions are

$$z_x(1,0) = \frac{\partial}{\partial x}\left(4 - (x^2 + y^2)\right)_{|(x,y)=(1,0)} = (-2x)_{|(x,y)=(1,0)} = -2,$$
$$z_y(1,0) = \frac{\partial}{\partial y}\left(4 - (x^2 + y^2)\right)_{|(x,y)=(1,0)} = (-2y)_{|(x,y)=(1,0)} = 0.$$

That is, there is no slope in the y direction. The slope in the x direction is negative, because the height decreases as x increases (from where she stands, increasing x would be moving away from the peak).

(ii) In order to be able to work with a larger set of functions – not just sums of powers of variables – we extend the single-variable differentiation rules from Term 1 to the multivariate setting: For example, the power rule becomes

$$(fg)_x = f_x g + f g_x$$

for an x-derivative and similarly for the y-derivative and the quotient rule.

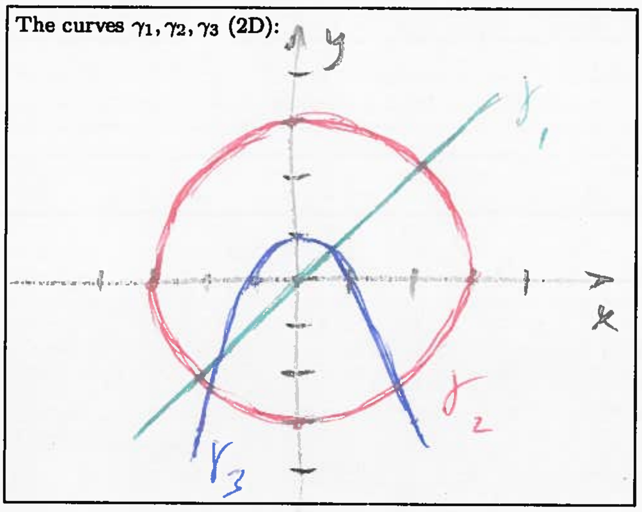
(iii) In order to find partial derivatives of functions like

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

we also need to combine the chain rule for single-variable functions – note that the outer function, $h(t) = \sin(t)$, is a single-variable function – with partial derivatives. Let c be a constant and review the following applications of the single-variable chain rule from Term 1.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\sin(t+c)\right) &= \cos(t+c)\frac{\mathrm{d}}{\mathrm{d}t}(t+c) = \cos(t+c), \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\sin(c+t^2)\right) &= \cos(c+t^2)\frac{\mathrm{d}}{\mathrm{d}t}(c+t^2) = 2t\cos(c+t^2). \end{aligned} \tag{2.1}$$

and highlight the point corresponding to $t = 0$ in each of them.
Sol.:



(ii) Consider the functions

$$\begin{aligned} f_1 &: \quad f_1(x,y) = x - y, \\ f_2 &: \quad f_2(x,y) = x^2 + y^2, \\ f_3 &: \quad f_3(x,y) = \ln(3 + x^2 - y), \end{aligned}$$

and find the three single-variable functions $F_i = f_i \circ \gamma_i$.
Sol.:

$$\begin{aligned} F_1(t) &= (f_1 \circ \gamma_1)(t) = f_1(t,t) = t - t = 0, \\ F_2(t) &= (f_2 \circ \gamma_2)(t) = f_2(3\cos t, 3\sin t) = (3\cos t)^2 + (3\sin t)^2 = 9, \\ F_3(t) &= \ln\left(3 + t^2 - (1 - t^2)\right) = \ln\left(2 + 2t^2\right) = \ln 2 + \ln\left(1 + t^2\right). \end{aligned} \tag{2.2}$$

Theorem 2.16 (The Chain Rule I). Let f and F be as in remark 2.14. Then

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}.$$

Example 2.17. (i) The derivatives of the functions F_i in 2.15 are

$$\begin{aligned} \frac{\mathrm{d}F_1}{\mathrm{d}t} &= \frac{\partial f_1}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f_1}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\partial(x-y)}{\partial x}\frac{\mathrm{d}t}{\mathrm{d}t} + \frac{\partial(x-y)}{\partial y}\frac{\mathrm{d}t}{\mathrm{d}t} = 1 \cdot 1 + (-1) \cdot 1 = 0, \\ \frac{\mathrm{d}F_2}{\mathrm{d}t} &= 2x(t) \cdot (-3\sin t) + 2y(t) \cdot 3\cos t = -6\cos t \sin t + 6 \sin t \cos t = 0, \\ \frac{\mathrm{d}F_3}{\mathrm{d}t} &= \frac{2x}{3 + x^2 - y} \cdot 1 + \frac{-1}{3 + x^2 - y} \cdot (-2t) = \frac{4t}{2 + 2t^2} = \frac{2t}{1 + t^2}. \end{aligned}$$

Alternatively, one could have found these derivatives by directly differentiating the functions F_i in (2.2). However, note that the chain rule allowed us to find them without using the explicit expressions of t on the right-hand sides of (2.2).

$$\begin{aligned}
T^{(2)}_{(1,1)}f(x,y) &= 3 + \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x-1 & y-1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix} \\
&= 3 + 4x - 4 + 2y - 2 + \frac{1}{2} \begin{bmatrix} x-1 & y-1 \end{bmatrix} \begin{bmatrix} 4x-4 \\ 2y-2 \end{bmatrix} \\
&= -3 + 4x + 2y + \begin{bmatrix} x-1 & y-1 \end{bmatrix} \begin{bmatrix} 2x-2 \\ y-1 \end{bmatrix} \\
&= -3 + 4x + 2y + 2x^2 - 2x - 2x + 2 + y^2 - y - y + 1 \\
&= 2x^2 + y^2.
\end{aligned}$$

Note that here, we have obtained $T^{(2)}f = f$, because $f(x,y)$ is already a very simple function – an expression of x and y containing terms of order at most 2. The analogous situation in the single-variable theory from last term is: The m-th-order Taylor approximation of a polynomial of order m is the original polynomial itself, and any higher-order approximation will be the same.

It is important not to forget about one of the parts we were asked about – the first-order approximation! It is contained in the above computation:

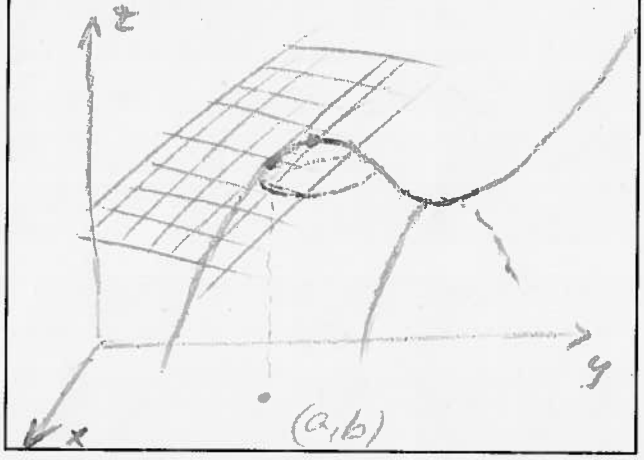
$$T^{(1)}_{(1,1)}f(x,y) = -3 + 4x + 2y.$$

$T^{(1)}_{(1,1)}f$ is again a function of two variables, and we plot such functions in $\mathbb{R}^3$ by drawing its function values as the z-coordinate. Hence we could write

$$z = -3 + 4x + 2y,$$

which is the equation of a plane in $\mathbb{R}^3$, namely the tangent plane of f at the point $(x_0, y_0, z_0) = (1, 1, f(1,1))$! This is similar to the corresponding 1D theory: the first-order Taylor approximation gives the equation of the tangent line. The following figure is a sketch of a generic two-variable function (some function, not the one above) and its first-order Taylor approximation:

Graphs of f and $T^{(1)}_{(a,b)}f$ (3D):

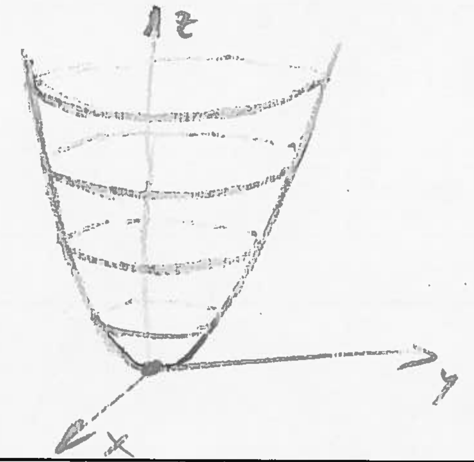


where we rescaled the x axis, $x \to \widetilde{x} = 2x$. The expression on the right is a paraboloid, for which $(0,0)$ is a minimum. The rescaling causes a deformation – imagine a paraboloid in front of you, then squeeze it, so that its level sets are ellipses rather than circles – but, again, this does not affect the property of having a minimum, maximum, or neither at the point $(0,0)$. We have therefore found out that h with $\lambda = 4$ and $\mu = 1$ has a local minimum at $(0,0)$. By this argument, we find that all the cases with $\lambda > 0$ and $\mu > 1$ are qualitatively the same and have a minimum. Let us choose the identity matrix as representing those cases,

$$M_1 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

for which $h(x,y) = x^2 + y^2$ and the graph of h is a paraboloid:

Example for case (i) (h induced by M_1, paraboloid; 3D):

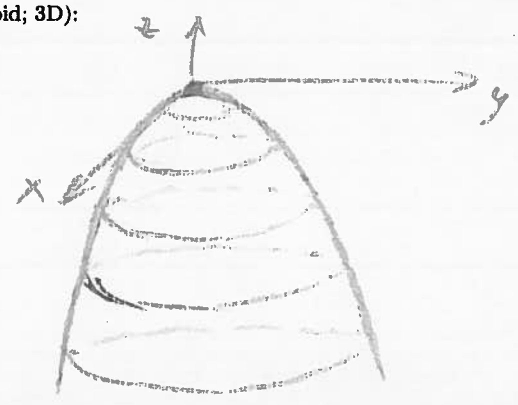


This is case (i) of the second-derivative test, when we have a minimum. Note that the representative $M_1 = I$ we have chosen for this class satisfies the assumptions of (i): $\det M_1 > 0$ and its upper left entry is positive.

Continuing this line of reasoning, we find that the other cases to consider are:

$$M_2 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, M_3 = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}, M_4 = \begin{bmatrix} +1 & 0 \\ 0 & 0 \end{bmatrix}, M_5 = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}, M_6 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

M_2 leads to the upside-down paraboloid $h(x,y) = -(x^2 + y^2)$, Example for case (ii) (h induced by M_2, upside-down paraboloid; 3D):

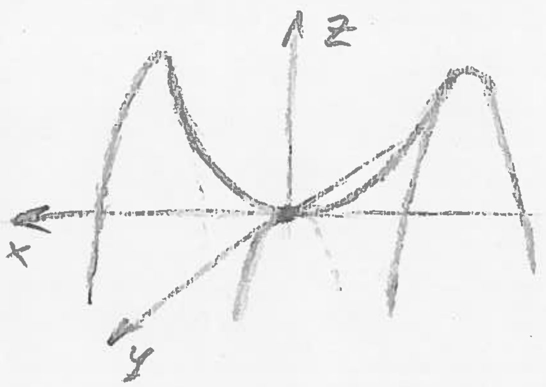


which has a maximum at $(0,0)$. Again, note that M_2 satisfies the assumptions of (ii) of the second-derivative test. Choosing M_3 gives

$$h(x,y) = x^2 - y^2,$$

which does not have an extremum at $(0,0)$: we have $h(0,0) = 0$, but both positive and negative function values arbitrarily close (e.g. $h(\varepsilon, 0) > 0$ and $h(0,\varepsilon) < 0$). This is case (iii) of the second-derivative test, and both the graph of $h(x,y) = x^2 - y^2$,

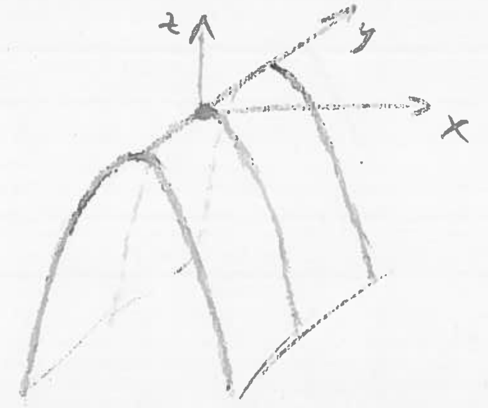
Example for case (iii), (h induced by M_3, saddle; 3D):



and the form of M_3 agree with the statement (iii) above.

Finally, we argue that the cases represented by M_4, M_5, M_6 do not allow us to draw a conclusion, i.e. they comprise case (iv) of the second-derivative test. They all do satisfy the assumption $\det M = 0$. The choice M_5 gives $h(x,y) = -x^2$, which has the following graph:

Example for case (iv) (h induced by M_5; 3D):



M_4 gives the same graph, but upside-down. M_6 gives the function $h(x, y) = 0$. In each of these cases, there is at least one straight line passing through the critical point $(0, 0)$ in question. On this line, small contributions of the original function f, that are not reflected by the second-order Taylor *approximation* of f we are using for the second-derivative test, could tip the balance to different conclusions. Some guidance for understanding this will be provided in the exercises below.

The purpose of this remark was to explain the workings behind the second derivative test, to sketch its proof (borrowing some second-year material from MTH2002), and, perhaps most importantly, to help you avoid confusion of the cases (i) and (ii): For example, suppose you are classifying a critical point, you have a Hessian matrix with positive determinant and positive entry in the upper left corner, but you have forgotten whether this implies a minimum or a maximum. Then, think of the representative $M = I$ of this situation – this gives the function $h(x, y) = x^2 + y^2$, which is the well-known paraboloid, which has a *minimum* at $(0, 0)$.

Exercise 2.46. (i) Work through the formative coursework sheets and your material from the lectures, and review the 1D second-derivative test from last term if necessary.

(ii) Convince yourself that theorem 2.41 is true; or, if you like, try to prove it formally. Ask your tutor or lecturer for help if necessary.

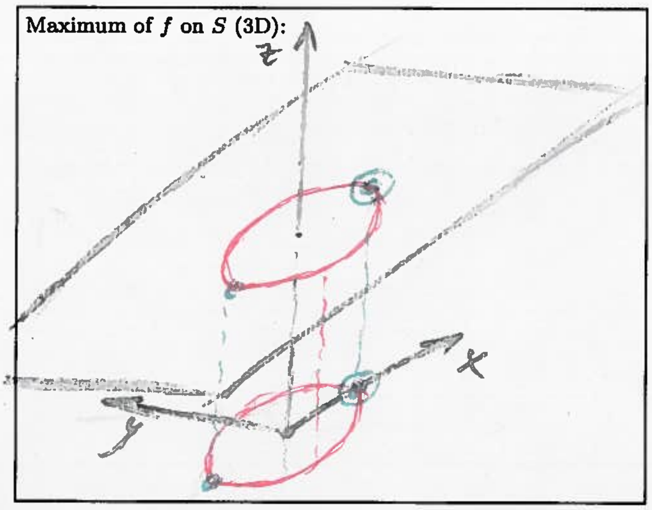
(iii) Make sure to properly work through the computations (iii) of examples 2.42 and 2.44 that are only sketched. For 2.44 (iii), use graphing software to try to find out what the point P is.

(iv) Give examples of (1) a local minimum, (2) a local maximum, and (3) a critical point that is neither, that cannot be classified with the second-derivative test[15].

[15]Hint: $f(x, y) = x^6 + y^4$ should cover one of those cases.

2.6 Extrema under Constraints: Lagrange Multipliers

Example 2.47. Find the maximum of the function $z = f(x, y) = 2x + y$ on the circle $S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.
Sol.: First note that the graph of f is a plane, and therefore f does not have any extrema at all on its full domain $\mathbb{R}^2$. However, we do have a maximum when restricting to the circle:



To find it, we can parametrise the circle as

$$(x(t), y(t)) = (\cos(t), \sin(t)),$$

and then define the function $F(t) = f(x(t), y(t))$. This gives

$$F(t) = 2\cos(t) + \sin(t),$$
$$F'(t) = -2\sin(t) + \cos(t).$$

Setting $F'(t)$ equal to zero, we obtain

$$\tan(t_0) = \frac{1}{2},$$

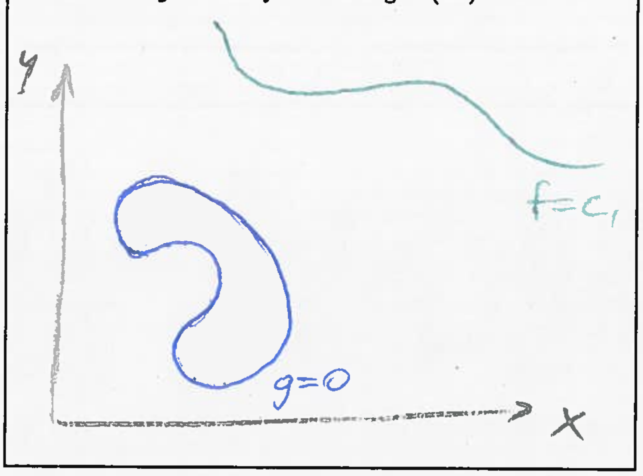
which has solutions $t_0 \approx 0.464 + k\pi$. For $k = 0$ and $k = 1$, we obtain the points

$$(x, y) \approx (\cos(0.464), \sin(0.464)) \approx (0.894, 0.448),$$
$$(x, y) \approx (\cos(3.606), \sin(3.606)) \approx (-0.894, -0.448),$$

and all other k will yield one of those two points again. Evaluating, we obtain the answer: f takes the maximum value 2.236 on the circle S at the point $(x, y) = (0.894, 0.448)$.

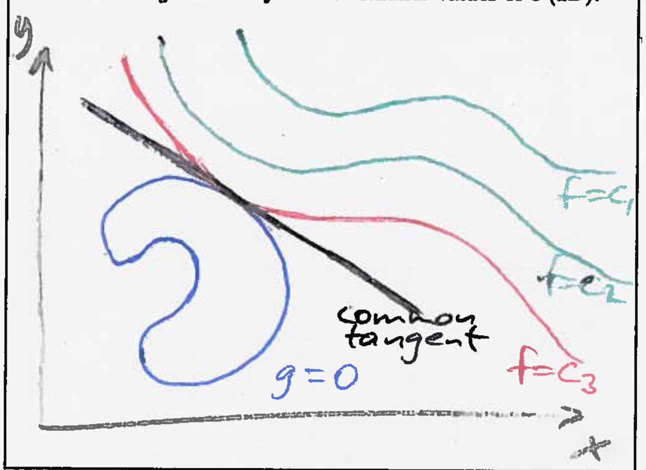
Remark 2.48. Now suppose the set S is given as a level set of a function g, and is less regular and can not easily be parametrised. The following figure shows such a set S in the domain of f, and further a level set $f(x, y) = c$, where we choose a constant $c = c_1$ that is larger than any of the values f takes on S.

The level sets $g = 0$ and $f = c$ for large c (2D):



If we now choose a slightly smaller constant c_2, the curve $f = c$ will move towards $g = 0$. We continue this process until the first contact between the two curves is made, say for $c = c_3$. This contact point is a local maximum – convince yourself of that[16]!

The level sets $g = 0$ and $f = c$ for smaller values of c (2D):



Note that the two level curves have a common tangent line in the figure – convince yourself that this is always the case[17]. Recall that gradient vectors are orthogonal to level sets and their tangent lines or planes – it must therefore be the case that the gradients of f and g are parallel:

$$\nabla f = \lambda \nabla g, \qquad \text{for some } \lambda \in \mathbb{R}.$$

[16]Hint: Can any of the other points in S have larger function values?

[17]Hint: To understand this, try to sketch a *first-contact* scenario where the tangent lines at the contact point intersect (you will not succeed; be aware that level curves of smooth functions do not have ends – they are either closed curves or they go out to infinity).