Lecture 10

Optimization problems for multivariable functions

Local maxima and minima - Critical points

(Relevant section from the textbook by Stewart: 14.7)

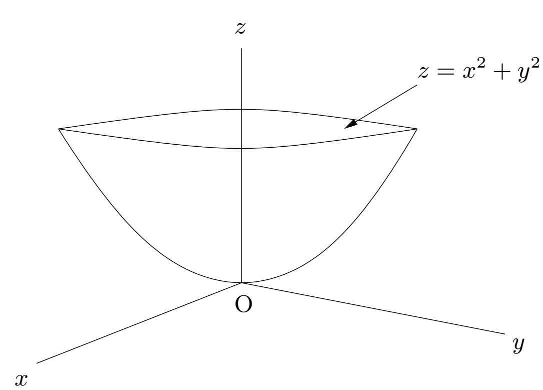
Our goal is to now find maximum and/or minimum values of functions of several variables, e.g., $f(x, y)$ over prescribed domains. As in the case of single-variable functions, we must first establish the notion of critical points of such functions.

Recall that a *critical point* of a function $f(x)$ of a single real variable is a point x for which either (i) $f'(x) = 0$ or (ii) $f'(x)$ is undefined. Critical points are possible candidates for points at which $f(x)$ attains a maximum or minimum value over an interval.

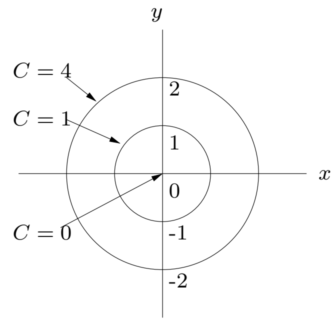
Also recall that if $f'(x) = 0$, it could be a (i) local minimum, (ii) local maximum or (iii) point of inflection. We can determine the nature of this critical point from a look at $f''(x)$, provided it exists.

Up to now, we have encountered three types of critical points for functions $f(x, y)$ of two variables:

1. **Local minima:** The point $(0, 0)$ is a *local minimum* for the function $f(x, y) = x^2 + y^2$, the graph of which is sketched below.



A plot of the countours/level sets of this function will also help us to understand the behaviour of this function around its local minimum. Such a plot, originally presented in Lecture 4, is



Some level sets of $z = x^2 + y^2$

shown again below.

The level sets of $f(x, y)$ satisfy the equation,

$$x^2 + y^2 = C. \quad (1)$$

As such, they are concentric circles of radius $\sqrt{C}$ centered at $(0, 0)$. As C approaches zero from above, these circles get smaller. The level set corresponding to $C = 0$ is the point $(0, 0)$, which represents the minimum value of f achieved at $(0, 0)$.

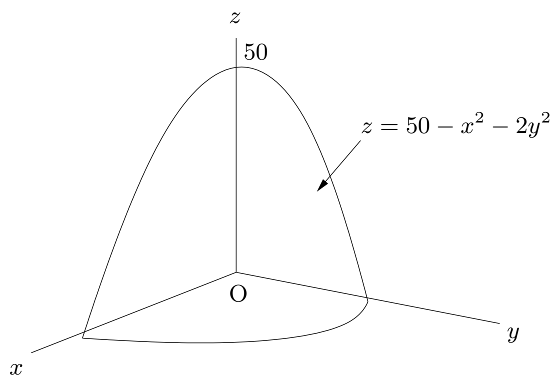
The definition of a local minimum seems quite straightforward but we state it here for the sake of completeness. (You'll also find it in the textbook.)

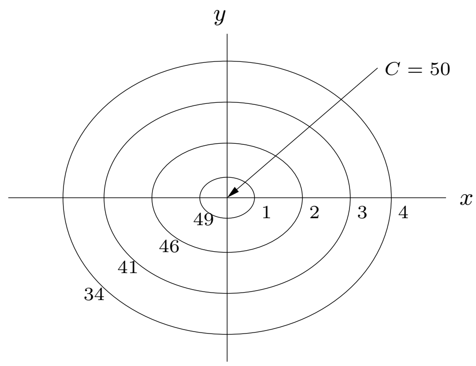
> A point (a, b) is a **local minimum** of the function $f(x, y)$ if there exists a circle C_r of radius $r > 0$ centered at (a, b) such that
>
> $$f(x, y) \geq f(a, b) \quad \text{for all} \quad (x, y) \quad \text{lying inside} \quad C_r. \quad (2)$$

Notes: In many books, the term "relative minimum" is used instead of "local minimum." The exact radius r of the circle is not important here. What is important is that a circular region of radius $r > 0$ exists.

2. **Local maxima:** The point $(0, 0)$ is a *local maximum* for the function $f(x, y) = 50 - x^2 - 2y^2$, the graph of which is sketched below. (This was the hotplate function studied earlier.)

   Once again, a plot of the contours for this function may be helpful to see how they get smaller and converge toward the single point at $(0, 0)$ which now represents a local maximum:





A point (a, b) is a local maximum of the function $f(x, y)$ if there exists a circle C_r of radius $r > 0$ centered at (a, b) such that

$$f(x, y) \leq f(a, b) \quad \text{for all} \quad (x, y) \quad \text{lying inside} \quad C_r. \tag{3}$$

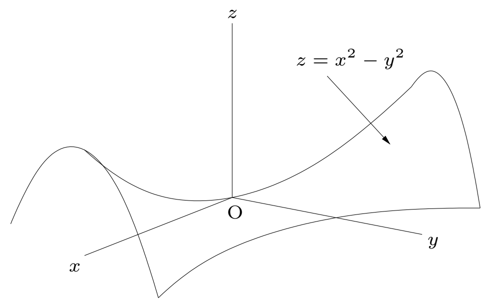
Note: In many books, the term "relative maximum" is used instead of "local maximum."

There is one other special kind of critical point:

3. **Saddle point:** An example is the point $(0, 0)$ for the function $f(x, y) = x^2 - y^2$. We sketch a graph of f near $(0, 0)$. Two noteworthy points can be made from this graph:

   (a) Consider the set of points $(x, y) = (x, 0)$ near $(0, 0)$. Then $f(x, 0) = x^2$ on this curve. $(0, 0)$ is seen to be a local minimum of $f(x, y)$ along this curve.

   (b) Now consider the set of points $(x, y) = (0, y)$ near $(0, 0)$. Then $f(0, y) = -y^2$ on this curve. $(0, 0)$ is now seen to be a local maximum of $f(x, y)$ along this curve.



The level sets of this function satisfy the equation,

$$x^2 - y^2 = C. \tag{4}$$

We consider three cases for C:

(a) $C = 0$: The solution to (4) is $y = \pm x$.

(b) $C > 0$: For example, when $C = 1$, the solution to

$$x^2 - y^2 = 1, \tag{5}$$

is a pair of rectangular hyperbolae that pass through the points $(1, 0)$ and $(-1, 0)$.

(c) $C < 0$: For example, when $C = -1$, the solution to

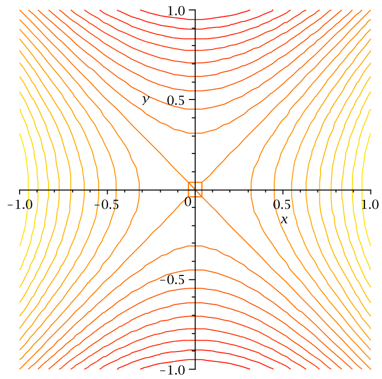
$$x^2 - y^2 = -1, \tag{6}$$

is a pair of rectangular hyperbolae that pass through the points $(0, 1)$ and $(0, -1)$.

A MAPLE plot of some contours of this function in the neighbourhood of the saddle point $(0, 0)$ is presented below. (A similar plot was presented in the Appendix on MAPLE commands for plotting multivariable functions and vector fields.)

Noteworthy differences between contours near local maxima/minima and saddle points: As seen above, is a quite striking difference between the behaviour of contours near local maxima/minima and contours near saddle points. In the former, the contours/level sets are concentric curves, whereas in the latter, they are hyperbolic in shape, with one set of curves, namely those that correspond to the value of the function at the saddle point, intersecting at the saddle point.

In all three cases studied above, the tangent plane to the graph $z = f(x, y)$ at the critical point is horizontal, i.e., parallel to the xy-plane. We'll see why in a moment.



In general, a point (a,b) is said to be a *critical point* of the function $f(x,y)$ if either

$$\frac{\partial f}{\partial x}(a,b) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(a,b) = 0, \tag{7}$$

or one or both of these partial derivatives does not exist at (a,b). Note that the above condition can be written more compactly as

$$\vec{\nabla} f(a,b) = 0 \quad \text{or fails to exist.} \tag{8}$$

Recall that the linearization of a function $f(x,y)$ at a point (a,b) is defined as

$$L_{(a,b)}(x,y) = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b). \tag{9}$$

The plane

$$z = L_{(a,b)}(x,y), \tag{10}$$

is tangent to the graph of $f(x,y)$, i.e., the surface $z = f(x,y)$, at $(x,y) = (a,b)$. At a critical point for which the partial derivatives vanish, as in the three examples discussed above, the linearization becomes the plane

$$z = f(a,b), \tag{11}$$

which is horizontal, i.e., parallel to the xy-plane.

This is also consistent with the fact that if $\vec{\nabla} f(a,b) = 0$, the *directional derivative* of f at (a,b) is zero *in any direction* $\hat{\mathbf{u}}$ since

$$D_{\hat{\mathbf{u}}} f(a,b) = \vec{\nabla} f(a,b) \cdot \hat{\mathbf{u}} = \mathbf{0} \cdot \hat{\mathbf{u}} = 0. \tag{12}$$

Example: Find all critical points of the function

$$f(x,y) = x^2 y - 2xy^2 + 3xy + 4. \tag{13}$$

Solution: To find critical points, we find (x, y) that satisfy $\vec{\nabla} f(x, y) = 0$, i.e.,

$$\begin{aligned} \frac{\partial f}{\partial x} = 2xy - 2y^2 + 3y &= 0 \\ \frac{\partial f}{\partial y} = x^2 - 4xy + 3x &= 0 \end{aligned} \tag{14}$$

First of all, we note that the partial derivatives f_x and f_y are defined for all (x, y). We can factorize these equations further to give

$$\begin{aligned} y(2x - 2y + 3) &= 0 \\ x(x - 4y + 3) &= 0 \end{aligned} \tag{15}$$

These two equations must be satisfied simultaneously. There are four possibilities:

1. $x = 0$ and $y = 0$, implying that $(0, 0)$ is a critical point.

2. $y = 0$ and $x - 4y + 3 = 0$, implying that $x = -3$. This yields the critical point $(-3, 0)$.

3. $x = 0$ and $2x - 2y + 3 = 0$, implying that $y = 3/2$. This yields the critical point $(0, 3/2)$.

4. $2x - 2y + 3 = 0$ and $x - 4y + 3 = 0$. This linear system of equations can be solved to give the critical point $(-1, 1/2)$.

To summarize, there are four critical points. While we are here, let's evaluate the function at these critical points:

1. $f(0, 0) = 4$,

2. $f(-3, 0) = 4$,

3. $f(0, 3/2) = 4$,

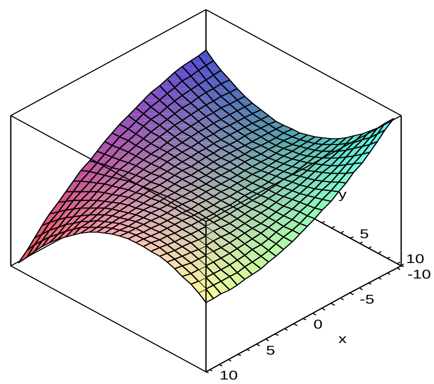
4. $f(-1, 1/2) = 7/2$.

The next, natural question is, "What is the nature of each critical point (a, b)?" Is it a local minimum? A local maximum? A saddle point?

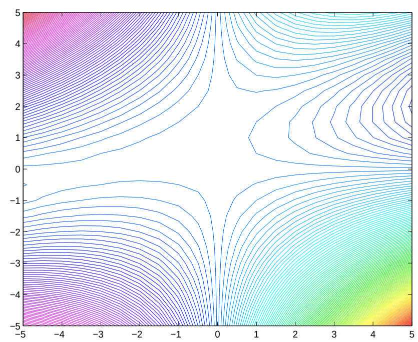
An examination of the values of the function $f(x, y)$ does not answer the question. Clearly, the first three values of the function, all of which are equal to 4, are greater than $f(-1, 1/2)$. One might suspect that these three values represent local maxima and that the latter value of 7/2 might be a local minimum. But this is still speculation.

One could result to some kind of computer-based method to determine the nature of each critical point. For example, we could examine the values of $f(x,y)$ over a sufficiently tiny circle C_r with radius $r > 0$ and centered at (a,b). That would, of course, take a little work.

Another possibility is to look at plots of this function, i.e., the graph $z = f(x,y)$ as well as contour plots of the level sets of $f(x,y)$. A plot of this function, over the region $x, y \in [-10, 10]$ and generated in MAPLE, is presented below in the left figure below. Unfortunately, the natures of these critical points are not readily seen from the figure. The figure at the right is the result of a MAPLE contour plot of the function over the region $x, y \in [-5, 5]$. In this case, 100 contours were plotted. You can see that even with 100 contours, the critical points have not been identified – the function $f(x,y)$ is very shallow in the region around $(0,0)$ in which the critical points are contained.



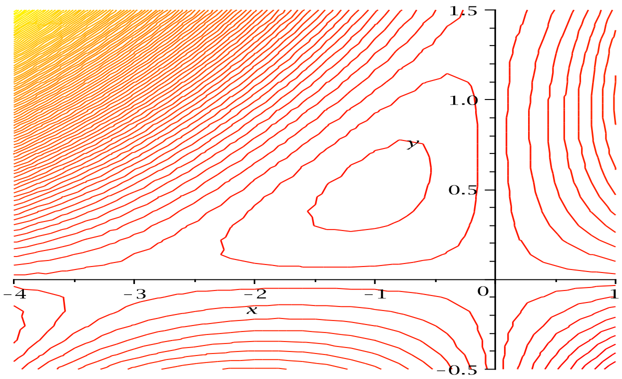
(a) Graph $z = f(x,y)$



(b) Contours of $f(x,y)$

Left: Graph $z = f(x,y)$ of the function $f(x,y) = x^2y - 2xy^2 + 3xy + 4$ over the region $x, y \in [-10, -10]$.
Right: Contours of $f(x,y)$ over the region $x, y \in [-5, 5]$.

If we magnify the region around the critical points, we might be able to get an idea of their nature. A magnified plot of the contours of $f(x,y)$ over the region $x \in [-4, 1]$, $y \in [-1/2, 3/2]$ is presented below. This plot leads us to suspect that the point $(-1, 1/2)$ is a local minimum (since its value is less than 4, the value at the other points) and that the other points, $(0,0)$, $(-3,0)$ and $(0,3/2)$, at which $f(x,y) = 4$ are saddle points. In the next lecture, we investigate a method of determining the nature of critical points analytically from the second derivatives of f evaluated at these points.



Contours of $f(x, y)$ over the region $-4 \leq x \leq 1$, $-0.5 \leq y \leq 1.5$

Lecture 11

Optimization problems for multivariable functions (cont'd)

Second derivative test for critical points of $f(x, y)$

(Relevant section from the textbook by Stewart: 14.7)

We now discuss a "second derivative test" to determine whether critical points of functions $f(x, y)$ are local minima, local maxima or saddle points. It is first instructive, however, to review the second derivative test from single-variable calculus.

Second derivative test - single-variable case: Recall that the critical point of a function $f(x)$ is a point a for which $f'(a) = 0$ or $f'(a)$ is undefined. In the former case, the critical point can be a local minimum, maximum point of inflection. In the case that the second derivative $f''(a)$ exists then:

1. If $f''(a) > 0$, then a is a local minimum,
2. If $f''(a) < 0$, then a is a local maximum,
3. If $f''(a) = 0$, then no conclusion can be made.

These together comprise the *second derivative test* for the critical point a.

For functions of several variables, e.g., $f(x, y)$, the situation is more complicated since we have to account for the behaviour of the function in more than one direction. As a result, we'll have to use all of the possible second-order partial derivatives, i.e., f_{xx}, f_{xy} and f_{yy}.

Second derivative test for critical points of functions $f(x, y)$:

Suppose that (a, b) is a critical point of $f(x, y)$, with $f_x(a, b) = f_y(a, b) = 0$. Furthermore, assume that the partial derivatives f_{xx}, f_{yy} and f_{xy} are continuous at (a, b). Define

$$A = f_{xx}(a, b), \qquad B = f_{xy}(a, b), \qquad C = f_{yy}(a, b), \tag{16}$$

and

$$D = D(a, b) = AC - B^2. \tag{17}$$

If

1. $D > 0$ and $A > 0$, then (a, b) is a local minimum.
2. $D > 0$ and $A < 0$, then (a, b) is a local maximum,
3. $D < 0$, then (a, b) is a saddle point,
4. $D = 0$, then no conclusion can be drawn.

In Case 1, you might be wondering why we check that $A > 0$ and not $C > 0$. The answer is that it doesn't matter: If $D > 0$, then A and C must have the same sign, since AC must be positive. (The term $-B^2$ is negative.) The same is true in Case 2.

A useful method for remembering the formula for D is to write it as a determinant:

$$D = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2. \tag{18}$$

We shall encounter this matrix again shortly.

Important note: In class, I introduced this test in terms of the quantity $B^2 - AC$, which is commonly used in other textbooks (including the textbook used in this course last year). After the lecture, I discovered that Stewart uses the above formulation in terms of $AC - B^2$. As such I have rewritten the formulation above to be consistent with Stewart's textbook and used it in later lectures. In fact, it is a more useful formulation, and easier to understand, because of the matrix of which D is a determinant.

Let us return to the functions examined in the previous lecture:

1. The function

$$f(x, y) = x^2 + y^2. \tag{19}$$

Here

$$\frac{\partial f}{\partial x} = 2x, \qquad \frac{\partial f}{\partial y} = 2y. \tag{20}$$

The condition for a critical point $f_x = f_y = 0$ is satisfied only by $(0, 0)$. The second derivatives at $(0, 0)$ are

$$A = f_{xx}(0, 0) = 2, \qquad B = f_{xy}(0, 0) = 0, \qquad C = f_{yy}(0, 0) = 2. \tag{21}$$

Then $D = AC - B^2 = 4 > 0$. Since $A = 2 > 0$, the second derivative test tells us that $(0,0)$ is a **local minimum** which, of course, is true.

2. The function

$$f(x,y) = 50 - x^2 - 2y^2. \tag{22}$$

Here

$$\frac{\partial f}{\partial x} = -2x, \qquad \frac{\partial f}{\partial y} = -4y. \tag{23}$$

The condition for a critical point $f_x = f_y = 0$ is satisfied only by $(0,0)$. The second derivatives at $(0,0)$ are

$$A = f_{xx}(0,0) = -2, \quad B = f_{xy}(0,0) = 0, \quad C = f_{yy}(0,0) = -4. \tag{24}$$

Then $D = AC - B^2 = 8 > 0$. Since $A = -2 < 0$, it the second derivative test tells us that $(0,0)$ is a **local maximum** which we know to be true.

3. The function

$$f(x,y) = x^2 - y^2. \tag{25}$$

Here

$$\frac{\partial f}{\partial x} = 2x, \qquad \frac{\partial f}{\partial y} = -2y. \tag{26}$$

The condition for a critical point $f_x = f_y = 0$ is satisfied only by $(0,0)$. The second derivatives at $(0,0)$ are

$$A = f_{xx}(0,0) = 2, \quad B = f_{xy}(0,0) = 0, \quad C = f_{yy}(0,0) = -2. \tag{27}$$

Then $D = AC - B^2 = -4 < 0$. The second derivative test tells us that $(0,0)$ is a **saddle point** which, once again, is true.

4. The function

$$f(x,y) = x^2y - 2xy^2 + 3xy + 4. \tag{28}$$

Recall that we found four critical points: $(0,0)$, $(-3,0)$, $(0,3/2)$ and $(-1,1/2)$. First we compute the necessary second-order derivatives at a general point (x,y):

$$A = f_{xx}(x,y) = 2y, \qquad B = f_{xy}(x,y) = 2x - 4y + 3, \qquad C = f_{yy}(x,y) = -4x. \tag{29}$$

(a) Critical point $(0,0)$: Here, $A = 0$, $B = 3$ and $C = 0$ so that $D = AC - B^2 = -9 < 0$. Saddle point.

(b) Critical point $(-3, 0)$: Here, $A = 0$, $B = -3$ and $C = 12$ so that $D = AC - B^2 = -9 < 0$. Saddle point.

(c) Critical point $(0, 3/2)$: Here, $A = 3$, $B = -3$ and $C = 0$ so that $D = A - B^2 = -9 < 0$. Saddle point.

(d) Critical point $(-1, 1/2)$: Here, $A = 1$, $B = -1$ and $C = 4$ so that $D = A - B^2 = 3 > 0$. Since $A > 0$, $(-1, 1/2)$ is a local minimum.

Recall that the value of the function $f(x, y)$ at the first three critical points was 4. These three critical points have turned out to be saddle points. And at the final critical point, $f(-1, 1/2) = 7/2$. This point is a local minimum. These conclusions are consistent with the contour plot of f presented in the previous lecture.

A more practical optimization problem: Find three positive numbers whose sum is 100 and whose product is a maximum. (This is Question No. 43, Section 14.7, p. 932 of Stewart, Sixth Edition.)

This is actually an example of a **constrained optimization problem**: We have to maximize the function

$$f(x, y, z) = xyz \tag{30}$$

subject to the constraint

$$x + y + z = 100. \tag{31}$$

As one student pointed out in class, we should be able to solve this problem using the method of "Lagrange multipliers," which has not yet been covered in this course. (It will be done next week.) We can incorporate the constraint by using it to express one variable in terms of the other two. In this way, we reduce the dimensionality of the problem from three to two. For example, let's use the constraint to express z in terms of x and y,

$$z = 100 - x - y. \tag{32}$$

Substitution into the function $f(x, y, z)$ produces a function $g(x, y)$:

$$g(x, y) = xy(100 - x - y) = 100xy - x^2y - xy^2. \tag{33}$$

The goal is to maximize this function of two variables. From the x and y values that maximize it, we can compute z. The first step is to find the critical points of g, which must satisfy both of the following equations:

$$\begin{aligned} \frac{\partial g}{\partial x} &= 100y - 2xy - y^2 = 0 \\ \frac{\partial g}{\partial y} &= 100x - x^2 - 2xy = 0 \end{aligned} \tag{34}$$

These equations can be factored:

$$\begin{aligned} y[100 - 2x - y] &= 0 \\ x[100 - x - 2y] &= 0. \end{aligned} \tag{35}$$

As in a previous example, there are four cases to consider:

1. $x = y = 0$, which is actually unacceptable, since the numbers are supposed to be positive. In this case, the product of x y and z would be zero.

2. $y = 0$ and $x + 2y = 100$, implying that $x = 100$. Once again, this is unacceptable.

3. $x = 0$ and $2x + y = 100$, implying that $y = 100$. Also unacceptable.

4. The solution to the linear homogeneous system of equations,

$$\begin{aligned} 2x + y &= 100 \\ x + 2y &= 100. \end{aligned} \tag{36}$$

   The determinant of this system is $D = 4 - 1 = 3$. Because it is nonzero, the solution to this system is unique. We can use a variety of methods to solve this simple problem: the answer is $x = y = \frac{100}{3}$.

Since Case 4 yields is the only one, it is almost certainly the solution to our problem. But let's verify that it corresponds to a maximum value of $g(x, y)$ by performing a second derivative test. The second derivatives of g are

$$g_{xx}(x, y) = -2y, \quad g_{xy}(x, y) = 100 - 2x - 2y, \quad g_{yy}(x, y) = -2x. \tag{37}$$

Even though it is not necessary to examine the unacceptable critical points, we shall do so, just to get an idea of the behaviour of the function $g(x, y)$.

1. Critical point $(0,0)$: $A = g_{xx}(0,0) = 0$, $B = g_{xy}(0,0) = 100$, $C = g_{yy}(0,0) = 0$. Then $D = AC - B^2 = -100^2 < 0$, corresponding to a saddle point.

2. Critical point $(100,0)$: $A = g_{xx}(100,0) = 0$, $B = g_{xy}(0,0) = -100$, $C = g_{yy}(0,0) = -200$. Then $D = AC - B^2 = -100^2 < 0$, corresponding to a saddle point.

3. Critical point $(0,100)$: $A = g_{xx}(100,0) = -200$, $B = g_{xy}(0,0) = -100$, $C = g_{yy}(0,0) = 0$. Then $D = AC - B^2 = -100^2 < 0$, corresponding to a saddle point.

4. Critical point $(100/3, 100/3)$: $A = g_{xx}(100,0) = -200/3$, $B = g_{xy}(0,0) = 100/3$, $C = g_{yy}(0,0) = -200/3$. Then $D = AC - B^2 = (200/3)^2 - (100/3)^2 > 0$ and $A < 0$, corresponding to a local maximum, as desired.

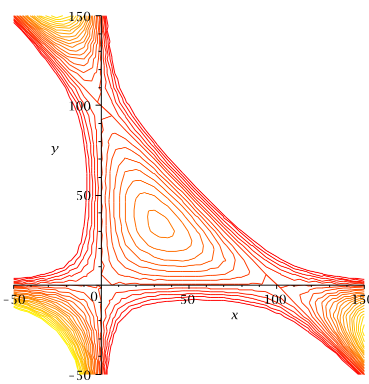
Corresponding to this final set of x and y values is the z value $\frac{100}{3}$. Thus the three numbers that add up to 100 and whose product is a maximum are

$$x = y = z = \frac{100}{3}. \tag{38}$$

In the figure below is shown a contour plot of the two-variable function $g(x,y)$ that we had to maximize to solve this optimization problem. As determined above, the three critical points $(0,0)$, $(100,0)$ and $(0,100)$ are saddle points of $g(x,y)$. There is another interesting phenomenon: there is a contour line connecting the points $(0,100)$ and $(100,0)$. The equation of this line is

$$x + y = 100. \tag{39}$$

A look at $g(x,y)$ in Eq. (33) shows that $g(x,y)$ must be zero at all points on this line.



A brief explanation of the second derivative test

(This is an expanded version of what was presented in class, in particular, the discussion about the eigenvalues of the matrix $\mathbf{H}$.)

The formulas comprising the second derivative test probably look rather strange. A proof is given in the text by Stewart, Sixth Edition on Page 930 of Section 14.7. In what follows, we'll show briefly below how these formulas can be obtained from a consideration of the Taylor series of $f(x,y)$ about a critical point. But first, let's review the one-variable case to give an idea of where we are going.

Recall that the Taylor expansion of a function $f(x)$ about a point $x = a$, up to the second derivative term takes the form

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots. \tag{40}$$

In the case that $x = a$ is a critical point, $f'(a)$ vanishes. We rewrite the above equation as follows,

$$f(x) - f(a) = \frac{1}{2}f''(a)(x-a)^2 + \cdots. \tag{41}$$

For x sufficiently close to a, the contributions from higher order terms, i.e., the "dots", can be neglected. As such, if $f''(a) > 0$, we have that $f(x) > f(a)$, implying that $x = a$ is a local minimum. If $f''(a) < 0$¡ then $f(x) < f(a)$, implying that $x = a$ is a local maximum. This is the basis of the second derivative test for single-variable functions.

We now move to the two-variable case. First, the Taylor expansion of $f(x,y)$ about a general point (a,b), up to all terms that employ second derivatives, has the following form,

$$\begin{aligned} f(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &\quad + \frac{1}{2}f_{xx}(a,b)(x-a)^2 + \frac{1}{2}f_{yy}(a,b)(y-b)^2 + f_{xy}(a,b)(x-a)(y-b) + \cdots \end{aligned} \tag{42}$$

In the case that (a,b) is a critical point, the first derivatives $f_x(a,b)$ and $f_y(a,b)$ vanish. We rewrite the above equation slightly to give

$$f(x,y) - f(a,b) = \frac{1}{2}A(x-a)^2 + \frac{1}{2}C(y-b)^2 + B(x-a)(y-b) + \cdots. \tag{43}$$

For (x,y) sufficiently close to (a,b) – in other words, the circle C_r has sufficiently small radius r – we can ignore the remainder terms $+\cdots$. We'll rewrite the result as

$$f(x,y)-f(a,b)=\begin{bmatrix} x-a & y-b \end{bmatrix}\frac{1}{2}\begin{bmatrix} A & B \\ B & C \end{bmatrix}\begin{bmatrix} x-a \\ y-b \end{bmatrix}=\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}. \tag{44}$$

Here, $\mathbf{h}^T=(x-a,y-b)=(\Delta x,\Delta y)$ represents a tiny displacement from the point (a,b). The term $\mathbf{h}^T\mathbf{H}\mathbf{h}$ is called a *quadratic form*. The matrix $\mathbf{H}$ is called the "Hessian" matrix associated with this quadratic form. Note that $D=\text{Det }\mathbf{H}$ is the determinant encountered in the second derivative test. We consider three cases:

1. If (a,b) is a local **minimum**, then the right-hand side must be **positive** for all (x,y) in a neighbourhood of (a,b) or, equivalently, all displacements $\mathbf{h}=(h_1,h_2)$ in a neighbourhood of $(0,0)$. From linear algebra, this is guaranteed if the two eigenvalues λ_1 and λ_2 of the matrix $\mathbf{H}$ are **positive**.

2. If (a,b) is a local **maximum**, then the right-hand side must be **negative** for all (x,y) in a neighbourhood of (a,b) or, equivalently, all displacements $\mathbf{h}=(h_1,h_2)$ in a neighbourhood of $(0,0)$. From linear algebra, this is guaranteed if the two eigenvalues λ_1 and λ_2 of the matrix $\mathbf{H}$ are **negative**.

3. If (a,b) is a **saddle point**, then the right-hand side must be

   (a) positive for all displacements $\mathbf{h}=(h_1,h_2)$ in a neighbourhood of $(0,0)$ and along a line emanating from $(0,0)$, and

   (b) negative for all displacements $\mathbf{h}=(h_1,h_2)$ in a neighbourhood of $(0,0)$ and along a another line emanating from $(0,0)$.

   From linear algebra, these two conditions are guaranteed if one of the two eigenvalues λ_1 and λ_2 of the matrix $\mathbf{H}$ is **negative** and the other is **positive**.

Now the eigenvalues λ_i of a 2×2 matrix $\mathbf{H}$ satisfy the characteristic equation

$$\lambda^2-\text{Tr }\mathbf{H}\lambda+\text{Det }\mathbf{H}=0, \tag{45}$$

where $\text{Tr}\mathbf{H}=A+C$ is the "trace" of the Hessian matrix $\mathbf{H}$ and $\text{Det }\mathbf{H}=AC-B^2$. From linear algebra, we also know that the roots of this equation λ_i satisfy

$$\lambda_1+\lambda_2=\text{Tr }\mathbf{H}, \qquad \lambda_1\lambda_2=\text{Det }\mathbf{H}. \tag{46}$$

If the eigenvalues λ_1 and λ_2 are nonzero and have opposite sign, i.e., one is positive and the other is negative, then $\lambda_1\lambda_2 < 0$, implying that Det $\mathbf{H} < 0$. This implies that Det $\mathbf{H} = AC - B^2 < 0$, corresponding to Case 3 above.

If the eigenvalues λ_1 and λ_2 are either both positive or both negative, then Det $\mathbf{H} = AC - B^2 > 0$, corresponding to Cases 1 and 2 above.

To determine whether both eigenvalues are positive or negative, it is sufficient to simply check one special direction of the displacement from the origin $(0,0)$: We could check $\Delta x = x - a \neq 0$ and $\Delta y = y - b = 0$. In this case, from Eq. (43),

$$f(x,y) - f(a,b) = \frac{1}{2}A(x-a)^2. \tag{47}$$

1. If $A > 0$, then $f(x,y) > f(a,b)$, implying that (a,b) is a local minimum.

2. If $A < 0$, then $f(x,y) < f(a,b)$, implying that (a,b) is a local maximum.

If we checked the special direction $\Delta x = x - a = 0$ and $\Delta y = y - b \neq 0$, then Eq. (43) becomes

$$f(x,y) - f(a,b) = \frac{1}{2}C(y-a)^2. \tag{48}$$

1. If $C > 0$, then $f(x,y) > f(a,b)$, implying that (a,b) is a local minimum.

2. If $C < 0$, then $f(x,y) < f(a,b)$, implying that (a,b) is a local maximum.

This explains why either A or C can be used in the second derivative test in Cases 1 and 2.

Lecture 12

Extreme values of functions over restricted domains - absolute maxima and minima

(Relevant section from the textbook by Stewart: 14.7)

We have now arrived at the most important aspect of optimization – determining the maximum and minimum values attained by a function $f(\mathbf{r}$ over a bounded region $R \subset \mathbf{R}^n$ of interest. This is analogous to the problem from first-year calculus of finding the absolute maximum and minimum values of a function $f(x)$ over an interval $[a, b]$. In what follows, we'll restrict our discussion to functions of two variables, i.e., $f(x, y)$. The definitions below easily extend to functions of several variables.

1. The *absolute maximum* of a function $f(x, y)$ on a region $R \subseteq \mathbf{R^2}$ is the largest value M attained by $f(x, y)$ on R, i.e.,
$$f(x, y) \leq M \quad \text{for all} \quad (x, y) \in R. \tag{49}$$
The point(s) (a, b) at which f attains this maximum value is(are) called absolute maximum point(s).

2. The *absolute minimum* of a function $f(x, y)$ on a region $R \subseteq \mathbf{R^2}$ is the least value m attained by $f(x, y)$ on R, i.e.,
$$f(x, y) \geq m \quad \text{for all} \quad (x, y) \in R. \tag{50}$$
The point(s) (a, b) at which f attains this minimum value is(are) called absolute minimum point(s).

Now recall the procedure for finding absolute maximum and minimum values of a function $f(x)$ over an interval $[a, b]$:

1. Determine all critical points of $f(x)$ in $[a, b]$ and evaluate f at these points.
2. Evaluate $f(x, y)$ at the endpoints of $[a, b]$, i.e., $f(a)$ and $f(b)$.
3. Select the largest and smallest values of $f(x)$ attained at the points examined in Steps 1 and 2.

For functions $f(x, y)$, the region R will be a two-dimensional region of $\mathbf{R^2}$, for example, the region contained inside a circle or a rectangle. Such regions R will generally not have endpoints but will be enclosed by *boundary curves*. (By the way, the boundary of a region $R \in \mathbf{R}^n$ is denoted mathematically as "∂R.") The analogous procedure of finding the absolute maximum and minimum values of $f(x, y)$ over region R will be as follows:

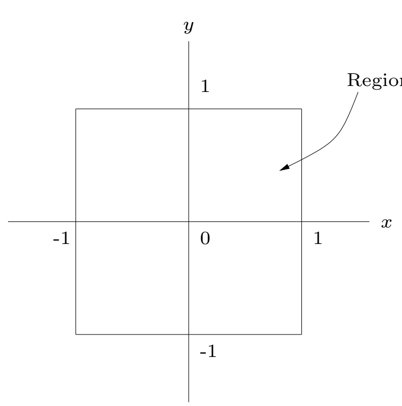
1. Determine all critical points of $f(x, y)$ in R and evaluate f at these points. (Note that we really don't need to spend time determining whether these critical points are local maxima, minima or saddle points - it is the value of the function $f(x, y)$ that is usually more important.)

2. Determine the maximum and minimum values achieved by $f(x, y)$ over the boundary curve(s) of R.

3. Select the largest and smallest values of $f(x, y)$ attained at the points examined in Steps 1 and 2.

Note: There is one important theoretical technicality. How do we know that f attains an absolute maximum or absolute minimum on a region R? In most of the examples that we shall encounter, R will be a *closed* and *bounded* region of $\mathbf{R}^2$, e.g., the interior of a rectangle, circle or ellipse. In such cases, if $f(x, y)$ is a continuous function of x and y, then it must attain absolute maximum and minimum values on R. (Recall the case of continuous functions $f(x)$ on closed intervals $[a, b]$ in first-year calculus.)

Example: Find the maximum and minimum values of the function $f(x, y) = x^2 + xy + y^2$ over the square region R defined by $-1 \leq x \leq 1$, $-1 \leq y \leq 1$.



Before we proceed with the detailed analysis, it is sometimes instructive to step back a bit and see if some "intuition" can give us an idea of where and what kind of answer to expect. For example, in this case, note that x^2 and y^2 are always positive (except, of course, at $(0,0)$). If x and y have the same sign, then the term xy will be positive, and therefore add to the positive term $x^2 + y^2$. On the contrary, if x and y have opposite signs, then the term xy will be negative and detract from the positive term $x^2 + y^2$. We therefore suspect that the maximum of f might be achieved at or around the upper right and lower left corners of the square region.

1. **Step 1:** Solve for all critical points of $f(x, y)$ that lie in R. They must satisfy

$$\begin{aligned} f_x &= 2x + y = 0 \\ f_y &= x + 2y = 0. \end{aligned} \tag{51}$$

   The only solution of this system is $(0,0)$. At this point $f(0,0) = 0$.

   Just for interest's sake, we shall perform a second derivative test of this point:

$$A = f_{xx}(0,0) = 2, \qquad B = f_{xy}(0,0) = 1, \qquad C = f_{yy}(0,0) = 2, \tag{52}$$

   so that $D = AC - B^2 = 3 > 0$. Since $A > 0$, $(0,0)$ is a local minimum.

   Note that we didn't really have to perform this test – an examination of the values of f at its critical points and on the boundary is sufficient.

2. **Step 2:** We now determine the maximum and minimum values achieved by $f(x, y)$ on the boundary of R.

   (a) The line $y = 1$, with $-1 \le x \le 1$. On this line, the function $f(x, y)$ is given by $f(x, 1) = x^2 + x + 1$, which we shall call $g(x)$. The problem is now to find the max and min values of $g(x)$ on $[-1, 1]$, a first-year calculus problem.

   Since $g'(x) = 2x+1$, the critical point of $g(x)$ is at $x = -1/2$. At this point $g(-1/2) = 3/4$.

   We must also check the endpoints $x = \pm 1$: $g(-1) = 1$ and $g(1) = 3$.

   (b) The line $y = -1$, with $-1 \le x \le 1$. Here, $f(x, y)$ is given by $f(x, -1) = x^2 - x + 1$, which we shall call $h(x)$. Now find the max and min values of $g(x)$ on $[-1, 1]$.

   Since $h'(x) = 2x - 1$, the critical point of $h(x)$ is at $x = 1/2$. At this point $h(1/2) = 3/4$.

   We also check the endpoints $x = \pm 1$: $h(-1) = 3$ and $h(1) = 1$.

(c) The line $x = 1$, with $-1 \leq y \leq 1$. Here, $f(x, y)$ is given by $f(1, y) = 1 + y + y^2$, which we shall call $k(y)$. We must now find the max and min values of $k(y)$ on $[-1, 1]$. This problem turns out to be identical to the first case, except that x is now called y. But just to be complete, we'll work it out in detail.

$k'(y) = 2y + 1$, the critical point of $k(y)$ is at $y = -1/2$. At this point $k(-1/2) = 3/4$.

We also check the endpoints $y = \pm 1$: $k(-1) = 1$ and $k(1) = 3$.

(d) The line $x = -1$, with $-1 \leq y \leq 1$. Here, $f(x, y)$ is given by $f(-1, y) = 1 - y + y^2$, which we shall call $l(y)$. Now find the max and min values of $g(x)$ on $[-1, 1]$.

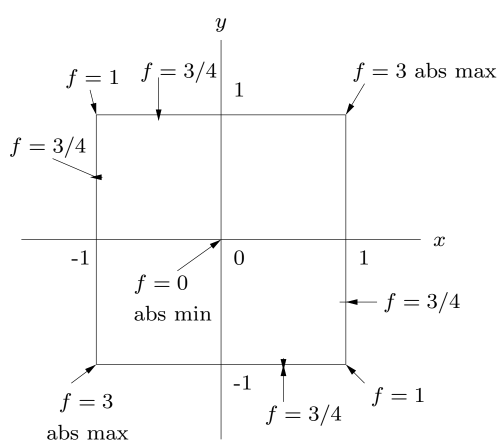
Since $l'(y) = 2x - 1$, the critical point of $l(y)$ is at $x = 1/2$. At this point $l(1/2) = 3/4$.

We also check the endpoints $x = \pm 1$: $l(-1) = 3$ and $l(1) = 1$.

Reviewing all of the above results, we have found the following: On the region R, the function $f(x, y) = x^2 + xy + y^2$ achieves

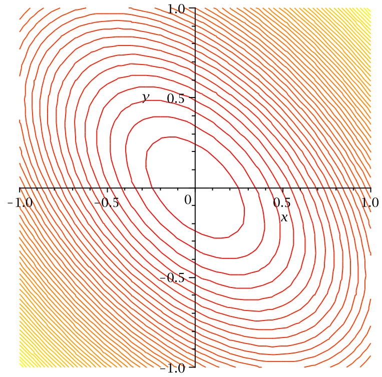
1. the absolute maximum value 3 at $(1, 1)$ and $(-1, -1)$
2. the absolute minimum value 0 at $(0, 0)$.

These results, which agree with our earlier intuition, are summarized in the figure below.



Summary of results of this example: Max and min values of $f(x, y) = x^2 + xy + y^2$ over square region $-1 \leq x, y \leq 1$.

A contour plot of $f(x, y)$ (produced by MAPLE using the *contourplot* command) is also presented below. The behaviour of the contours explains the results of this example.



Contour plot of $f(x, y) = x^2 + xy + y^2$.

We continue with another example of finding the absolute maxima and minima of a function over a region D.

Example 2: Find the absolute max/min values of $f(x, y) = x^2 + xy + y^2$ over the circular region $R = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

This is the same function as considered in the previous lecture but the region R is different. Here, R is enclosed by the circle $x^2 + y^2 = 1$.

1. **Step 1:** Determine critical points of f in R. $(0, 0)$ is the only critical point of f. It also lies in the region R. Here $f(0, 0) = 0$.

2. **Step 2:** Examine f over the boundary of R, the circle $x^2 + y^2 = 1$. Perhaps the easiest way to proceed is to parametrize the curve. (See note below.) Because of the circular symmetry, we use

$$x(t) = \cos t, \qquad y(t) = \sin t, \quad 0 \leq t \leq 2\pi. \tag{53}$$

On this curve, we can define the function

$$\begin{aligned} g(t) &= f(x(t), y(t)) \\ &= \cos^2 t + \cos t \sin t + \sin^2 t \\ &= 1 + \frac{1}{2}\sin 2t, \qquad 0 \leq t \leq 2\pi. \end{aligned} \tag{54}$$

This is once again a first-year calculus problem. There are at least two methods to determine the max/min vaues of $g(t)$ over the interval $[0, 2\pi]$.

(a) **Method No. 1:** The function $\sin 2t$ makes two complete oscillations over $[0, 2\pi]$ and achieves a maximum value of 1 at $t = \pi/4, 5\pi/4$ and a minumum value of -1 at $t = 3\pi/4, 7\pi/4$. Since we take one-half of these values and add them to 1 to produce $g(t)$, it follows that $g(t)$ achieves a maximum value of 3/2 at $t = \pi/4, 5\pi/4$ and a minimum value of 1/2 at $t = 3\pi/4, 7\pi/4$.

(b) **Method No. 2** The more traditional calculus examination of $g(t)$. Since $g'(t) = \cos 2t$, the critical points of $g(t)$ lie at points where $2t = \pi/2, 3\pi/2, 5\pi/2, 7\pi/2$. Thus, critical points exist at $t = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$.

i. At $t = \pi/4$, $g(\pi/4) = 1 + 1/2 = 3/2$. This occurs at $x = \cos(\pi/4) = 1/\sqrt{2}$, $y = \sin(\pi/4) = 1/\sqrt{2}$.

ii. Likewise, at $t = 5\pi/4$, $g(5\pi/4) = 1 + 1/2 = 3/2$. This occurs at $x = \cos(5\pi/4) = -1/\sqrt{2}$, $y = \sin(5\pi/4) = -1/\sqrt{2}$.

iii. At $t = 3\pi/4$, $g(3\pi/4) = 1 - 1/2 = 1/2$. This occurs at $x = \cos(3\pi/4) = 1\sqrt{2}$, $y = \sin(3\pi/4) = 1\sqrt{2}$.

iv. At $t = 7\pi/4$, $g(7\pi/4) = 1 - 1/2 = 1/2$. This occurs at $x = \cos(7\pi/4) = 1\sqrt{2}$, $y = \sin(7\pi/4) = -1\sqrt{2}$.

All of these results are in accord with those of Method No. 1.

The result of this entire procedure: f achieves an absolute maximum value of 3/2 at the points $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, -1/\sqrt{2})$ and an absolute minimum value of 0 at $(0, 0)$.

Note: In this example, instead of parametrizing the curve in terms of a parameter t, we could have expressed one variable, e.g., y in terms of the other, i.e., x. The boundary curve then becomes composed of two curves,

$$y_1 = \sqrt{1 - x^2}, \qquad y_2 = -\sqrt{1 - x^2}, \qquad -1 \leq x \leq 1, \tag{55}$$

which are the upper and lower semicircular curves comprising the circle. On the top curve, the function $f(x, y)$ becomes

$$f(x, y) \quad = \quad f(x, \sqrt{1 - x^2})$$

$$
\begin{aligned}
&= x^2 + x\sqrt{1-x^2} + 1 - x^2 \\
&= 1 + x\sqrt{1-x^2}, \qquad -1 \leq x \leq 1. \quad (56)
\end{aligned}
$$

And on the bottom curve,

$$f(x,y) = 1 - x\sqrt{1-x^2}. \quad (57)$$

We may then analyze these functions of the single variable x to find maximum and minimum values for $x \in [-1,1]$.