

2. Functions of Several Variables

Functions of Several Variables

2.1 Functions of Two or Three Variables

In Section 1.8 we discussed vector-valued functions of a single real variable. We will now examine real-valued functions of a point (or vector) in $\mathbb{R}^2$ or $\mathbb{R}^3$. For the most part these functions will be defined on sets of points in $\mathbb{R}^2$, but there will be times when we will use points in $\mathbb{R}^3$, and there will also be times when it will be convenient to think of the points as vectors (or terminal points of vectors).

A **real-valued function** f defined on a subset D of $\mathbb{R}^2$ is a rule that assigns to each point (x, y) in D a real number $f(x, y)$. The largest possible set D in $\mathbb{R}^2$ on which f is defined is called the **domain** of f , and the **range** of f is the set of all real numbers $f(x, y)$ as (x, y) varies over the domain D . A similar definition holds for functions $f(x, y, z)$ defined on points (x, y, z) in $\mathbb{R}^3$.

width height 0.5pt

Example 2.1

The domain of the function

$$f(x, y) = xy$$

is all of $\mathbb{R}^2$, and the range of f is all of $\mathbb{R}$.

width height 0.5pt

Example 2.2

The domain of the function

$$f(x, y) = \frac{1}{x - y}$$

is all of $\mathbb{R}^2$ except the points (x, y) for which $x = y$. That is, the domain is the set $D = \{(x, y) : x \neq y\}$. The range of f is all real numbers except 0.

width height 0.5pt

Example 2.3

The domain of the function

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

is the set $D = \{(x, y) : x^2 + y^2 \leq 1\}$, since the quantity inside the square root is nonnegative if and only if $1 - (x^2 + y^2) \geq 0$. We see that D consists of all points on and inside the unit circle in $\mathbb{R}^2$ (D is sometimes called the *closed unit disk*). The range of f is the interval $[0, 1]$ in $\mathbb{R}$.

width height 0.5pt

Example 2.4

The domain of the function

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

is all of $\mathbb{R}^3$, and the range of f is all positive real numbers.

width height 0.5pt

A function $f(x, y)$ defined in $\mathbb{R}^2$ is often written as $z = f(x, y)$, as was mentioned in Section 1.1, so that the **graph** of $f(x, y)$ is the set $\{(x, y, z) : z = f(x, y)\}$ in $\mathbb{R}^3$. So we see that this graph is a surface in $\mathbb{R}^3$, since it satisfies an equation of the form $F(x, y, z) = 0$ (namely, $F(x, y, z) = f(x, y) - z$). The traces of this surface in the planes $z = c$, where c varies over $\mathbb{R}$, are called the **level curves** of the function. Equivalently, the level curves are the solution sets of the equations $f(x, y) = c$, for c in $\mathbb{R}$. Level curves are often projected onto the xy -plane to give an idea of the various “elevation” levels of the surface (as is done in topography).

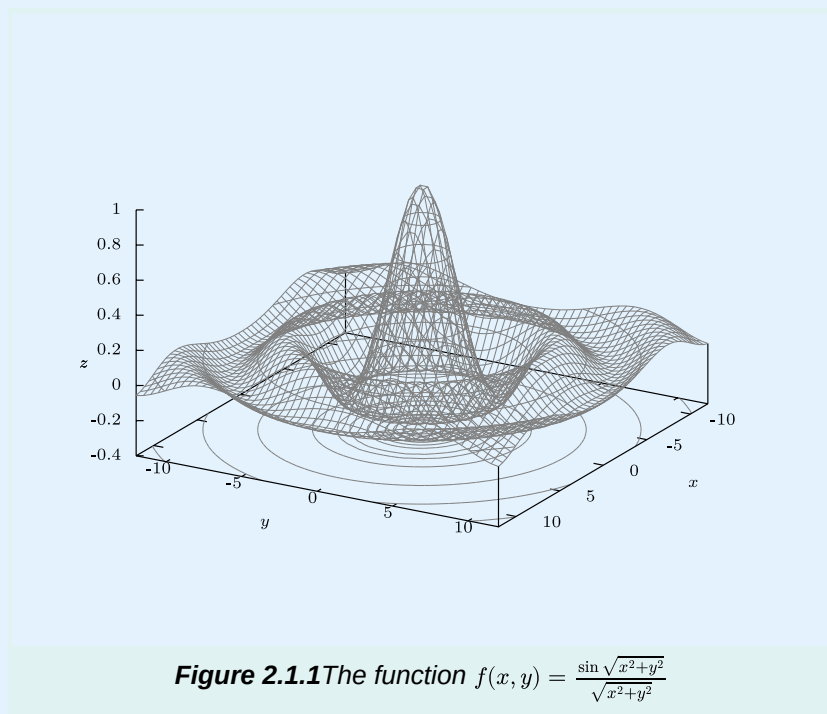
width height 0.5pt

Example 2.5

The graph of the function

$$f(x, y) = \frac{\sin \sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}}$$

is shown below. Note that the level curves (shown both on the surface and projected onto the xy -plane) are groups of concentric circles.



width height 0.5pt

You may be wondering what happens to the function in Example 2.5 at the point $(x, y) = (0, 0)$, since both the numerator and denominator are 0 at that point. The function is not defined at $(0, 0)$, but the *limit* of the function exists (and equals 1) as (x, y) *approaches* $(0, 0)$. We will now state explicitly what is meant by the limit of a function of two variables.

Definition 2.1

Let (a, b) be a point in $\mathbb{R}^2$, and let $f(x, y)$ be a real-valued function defined on some set containing (a, b) (but not necessarily defined at (a, b) itself). Then we say that the **limit** of $f(x, y)$ equals L as (x, y) approaches (a, b) , written as

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = L, \quad (2.1)$$

if given any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x, y) - L| < \epsilon \quad \text{whenever} \quad 0 < \sqrt{(x - a)^2 + (y - b)^2} < \delta.$$

A similar definition can be made for functions of three variables. The idea behind the above definition is that the values of $f(x, y)$ can get arbitrarily close to L (i.e. within ϵ of L) if we pick (x, y) sufficiently close to (a, b) (i.e. inside a circle centered at (a, b) with some sufficiently small radius δ).

If you recall the “epsilon-delta” proofs of limits of real-valued functions of a single variable, you may remember how awkward they can be, and how they can usually only be done easily for simple functions. In general, the multivariable cases are at least equally awkward to go through, so we will not bother with such proofs. Instead, we will simply state that when the function $f(x, y)$ is given by a single formula and is defined at the point (a, b) (e.g. is not some indeterminate form like $0/0$) then you can just substitute $(x, y) = (a, b)$ into the formula for $f(x, y)$ to find the limit.

width height 0.5pt

Example 2.6

$$\lim_{(x,y) \rightarrow (1,2)} \frac{xy}{x^2 + y^2} = \frac{(1)(2)}{1^2 + 2^2} = \frac{2}{5}$$

since $f(x, y) = \frac{xy}{x^2 + y^2}$ is properly defined at the point $(1, 2)$.

width height 0.5pt

The major difference between limits in one variable and limits in two or more variables has to do with how a point is approached. In the single-variable case, the statement “ $x \rightarrow a$ ” means that x gets closer to the value a from two possible directions along the real number line (see Figure 2.1.2(a)). In two dimensions, however, (x, y) can approach a point (a, b) along an infinite number of paths (see Figure 2.1.2(b)).

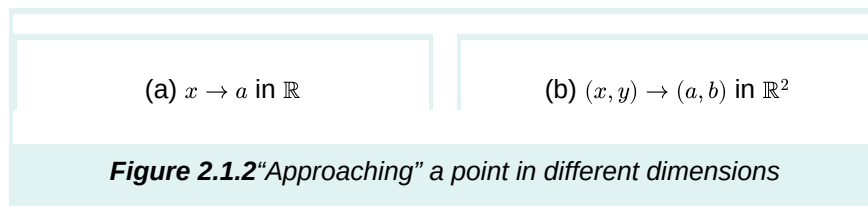


Figure 2.1.2 “Approaching” a point in different dimensions

Example 2.7

$$\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2} \text{ does not exist}$$

Note that we can not simply substitute $(x, y) = (0, 0)$ into the function, since doing so gives an indeterminate form $0/0$. To show that the limit does not exist, we will show that the function approaches different values as (x, y) approaches $(0, 0)$ along different paths in $\mathbb{R}^2$. To see this, suppose that $(x, y) \rightarrow (0, 0)$ along the positive x -axis, so that $y = 0$ along that path. Then

$$f(x, y) = \frac{xy}{x^2 + y^2} = \frac{x \cdot 0}{x^2 + 0^2} = 0$$

along that path (since $x > 0$ in the denominator). But if $(x, y) \rightarrow (0, 0)$ along the straight line $y = x$ through the origin, for $x > 0$, then we see that

$$f(x, y) = \frac{xy}{x^2 + y^2} = \frac{x^2}{x^2 + x^2} = \frac{1}{2}$$

which means that $f(x, y)$ approaches different values as $(x, y) \rightarrow (0, 0)$ along different paths. Hence the limit does not exist.

width height 0.5pt

Limits of real-valued multivariable functions obey the same algebraic rules as in the single-variable case, as shown in the following theorem, which we state without proof.

Theorem 2.1

Suppose that $\lim_{(x,y) \rightarrow (a,b)} f(x,y)$ and $\lim_{(x,y) \rightarrow (a,b)} g(x,y)$ both exist, and that k is some scalar. Then:

1. $\lim_{(x,y) \rightarrow (a,b)} [f(x,y) \pm g(x,y)] = \left[\lim_{(x,y) \rightarrow (a,b)} f(x,y) \right] \pm \left[\lim_{(x,y) \rightarrow (a,b)} g(x,y) \right]$
2. $\lim_{(x,y) \rightarrow (a,b)} k f(x,y) = k \left[\lim_{(x,y) \rightarrow (a,b)} f(x,y) \right]$
3. $\lim_{(x,y) \rightarrow (a,b)} [f(x,y)g(x,y)] = \left[\lim_{(x,y) \rightarrow (a,b)} f(x,y) \right] \left[\lim_{(x,y) \rightarrow (a,b)} g(x,y) \right]$
4. $\lim_{(x,y) \rightarrow (a,b)} \frac{f(x,y)}{g(x,y)} = \frac{\lim_{(x,y) \rightarrow (a,b)} f(x,y)}{\lim_{(x,y) \rightarrow (a,b)} g(x,y)}$ if $\lim_{(x,y) \rightarrow (a,b)} g(x,y) \neq 0$
5. If $|f(x,y) - L| \leq g(x,y)$ for all (x,y) and if $\lim_{(x,y) \rightarrow (a,b)} g(x,y) = 0$, then $\lim_{(x,y) \rightarrow (a,b)} f(x,y) = L$.

Note that in part (e), it suffices to have $|f(x,y) - L| \leq g(x,y)$ for all (x,y) “sufficiently close” to (a,b) (but excluding (a,b) itself).

Example 2.8

Show that

$$\lim_{(x,y) \rightarrow (0,0)} \frac{y^4}{x^2 + y^2} = 0.$$

Since substituting $(x,y) = (0,0)$ into the function gives the indeterminate form $0/0$, we need an alternate method for evaluating this limit. We will use Theorem 2.1(e). First, notice that $y^4 = (\sqrt{y^2})^4$ and so $0 \leq y^4 \leq (\sqrt{x^2 + y^2})^4$ for all (x,y) . But $(\sqrt{x^2 + y^2})^4 = (x^2 + y^2)^2$. Thus, for all $(x,y) \neq (0,0)$ we have

$$\left| \frac{y^4}{x^2 + y^2} \right| \leq \frac{(x^2 + y^2)^2}{x^2 + y^2} = x^2 + y^2 \rightarrow 0 \text{ as } (x,y) \rightarrow (0,0).$$

Therefore $\lim_{(x,y) \rightarrow (0,0)} \frac{y^4}{x^2 + y^2} = 0$.

width height 0.5pt

Continuity can be defined similarly as in the single-variable case.

Definition 2.2

A real-valued function $f(x, y)$ with domain D in $\mathbb{R}^2$ is **continuous** at the point (a, b) in D if $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$. We say that $f(x, y)$ is a **continuous function** if it is continuous at every point in its domain D .

Unless indicated otherwise, you can assume that all the functions we deal with are continuous. In fact, we can modify the function from Example 2.8 so that it is continuous on all of $\mathbb{R}^2$.

width height 0.5pt

Example 2.9

Define a function $f(x, y)$ on all of $\mathbb{R}^2$ as follows:

$$f(x, y) = \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ \frac{y^4}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \end{cases}$$

Then $f(x, y)$ is well-defined for all (x, y) in $\mathbb{R}^2$ (i.e. there are no indeterminate forms for any (x, y)), and we see that

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = \frac{b^4}{a^2 + b^2} = f(a, b) \quad \text{for } (a, b) \neq (0, 0).$$

So since

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0 = f(0, 0) \quad \text{by Example 2.8,}$$

then $f(x, y)$ is continuous on all of $\mathbb{R}^2$.

width height 0.5pt

Exercises**A**

For Exercises 1-6, state the domain and range of the given function.

2

1. $f(x, y) = x^2 + y^2 - 1$

2. $f(x, y) = \frac{1}{x^2 + y^2}$

2

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

4. $f(x, y) = \frac{x^2 + 1}{y}$

2

5. $f(x, y, z) = \sin(xyz)$

6. $f(x, y, z) = \sqrt{(x-1)(yz-1)}$

For Exercises 7-18, evaluate the given limit.

2

$$7. \lim_{(x,y) \rightarrow (0,0)} \cos(xy)$$

$$8. \lim_{(x,y) \rightarrow (0,0)} e^{xy}$$

2

$$9. \lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$$

$$10. \lim_{(x,y) \rightarrow (0,0)} \frac{xy^2}{x^2 + y^4}$$

2

$$11. \lim_{(x,y) \rightarrow (1,-1)} \frac{x^2 - 2xy + y^2}{x - y}$$

$$12. \lim_{(x,y) \rightarrow (0,0)} \frac{xy^2}{x^2 + y^2}$$

2

$$13. \lim_{(x,y) \rightarrow (1,1)} \frac{x^2 - y^2}{x - y}$$

$$14. \lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - 2xy + y^2}{x - y}$$

2

$$15. \lim_{(x,y) \rightarrow (0,0)} \frac{y^4 \sin(xy)}{x^2 + y^2}$$

$$16. \lim_{(x,y) \rightarrow (0,0)} (x^2 + y^2) \cos\left(\frac{1}{xy}\right)$$

2

$$17. \lim_{(x,y) \rightarrow (0,0)} \frac{x}{y}$$

$$18. \lim_{(x,y) \rightarrow (0,0)} \cos\left(\frac{1}{xy}\right)$$

B

19. Show that $f(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$, for $\sigma > 0$, is constant on the circle of radius $r > 0$ centered at the origin. This function is called a *Gaussian blur*, and is used as a filter in image processing software to produce a “blurred” effect.
20. Suppose that $f(x, y) \leq f(y, x)$ for all (x, y) in $\mathbb{R}^2$. Show that $f(x, y) = f(y, x)$ for all (x, y) in $\mathbb{R}^2$.
21. Use the substitution $r = \sqrt{x^2 + y^2}$ to show that

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\sin \sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} = 1 .$$

(Hint: You will need to use L'Hôpital's Rule for single-variable limits.)

C

22. Prove Theorem 2.1(a) in the case of addition. (Hint: Use Definition 2.1.)
23. Prove Theorem 2.1(b).

2.2 Partial Derivatives

Now that we have an idea of what functions of several variables are, and what a limit of such a function is, we can start to develop an idea of a derivative of a function of two or more variables. We will start with the notion of a *partial derivative*.

Definition 2.3

Let $f(x, y)$ be a real-valued function with domain D in $\mathbb{R}^2$, and let (a, b) be a point in D . Then the **partial derivative of f at (a, b) with respect to x** , denoted by $\frac{\partial f}{\partial x}(a, b)$, is defined as

$$\frac{\partial f}{\partial x}(a, b) = \lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h} \quad (2.2)$$

and the **partial derivative of f at (a, b) with respect to y** , denoted by $\frac{\partial f}{\partial y}(a, b)$, is defined as

$$\frac{\partial f}{\partial y}(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b + h) - f(a, b)}{h} . \quad (2.3)$$

Note: The symbol ∂ is pronounced “del”.^[1]

Recall that the derivative of a function $f(x)$ can be interpreted as the rate of change of that function in the (positive) x direction. From the definitions above, we can see that the partial derivative of a function $f(x, y)$ with respect to x or y is the rate of change of $f(x, y)$ in the (positive) x or y direction, respectively. What this means is that the partial derivative of a function $f(x, y)$ with respect to x can be calculated by treating the y variable as a *constant*, and then simply differentiating $f(x, y)$ as if it were a function of x alone, using the usual rules from single-variable calculus. Likewise, the partial derivative of $f(x, y)$ with respect to y is obtained by treating the x variable as a constant and then differentiating $f(x, y)$ as if it were a function of y alone.

width height 0.5pt

Example 2.10

Find $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ for the function $f(x, y) = x^2y + y^3$.

Solution: Treating y as a constant and differentiating $f(x, y)$ with respect to x gives

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

and treating x as a constant and differentiating $f(x, y)$ with respect to y gives

$$\frac{\partial f}{\partial y}(x, y) = x^2 + 3y^2.$$

width height 0.5pt We will often simply write $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ instead of $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$.

width height 0.5pt

Example 2.11

Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ for the function $f(x, y) = \frac{\sin(xy^2)}{x^2 + 1}$.

Solution: Treating y as a constant and differentiating $f(x, y)$ with respect to x gives

$$\frac{\partial f}{\partial x} = \frac{(x^2 + 1)(y^2 \cos(xy^2)) - (2x) \sin(xy^2)}{(x^2 + 1)^2}$$

and treating x as a constant and differentiating $f(x, y)$ with respect to y gives

$$\frac{\partial f}{\partial y} = \frac{2xy \cos(xy^2)}{x^2 + 1}.$$

width height 0.5pt

Since both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are themselves functions of x and y , we can take *their* partial derivatives with respect to x and y . This yields the *higher-order partial derivatives*:

$$\begin{array}{ll} \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right) & \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right) \\ \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) & \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \\ \frac{\partial^3 f}{\partial x^3} = \frac{\partial}{\partial x} \left(\frac{\partial^2 f}{\partial x^2} \right) & \frac{\partial^3 f}{\partial y^3} = \frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial y^2} \right) \\ \frac{\partial^3 f}{\partial y \partial x^2} = \frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial x^2} \right) & \frac{\partial^3 f}{\partial x \partial y^2} = \frac{\partial}{\partial x} \left(\frac{\partial^2 f}{\partial y^2} \right) \\ \frac{\partial^3 f}{\partial y^2 \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial y \partial x} \right) & \frac{\partial^3 f}{\partial x^2 \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial^2 f}{\partial x \partial y} \right) \\ \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial}{\partial x} \left(\frac{\partial^2 f}{\partial y \partial x} \right) & \frac{\partial^3 f}{\partial y \partial x \partial y} = \frac{\partial}{\partial y} \left(\frac{\partial^2 f}{\partial x \partial y} \right) \\ & \vdots \end{array}$$

width height 0.5pt

Example 2.12

Find the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, $\frac{\partial^2 f}{\partial y \partial x}$ and $\frac{\partial^2 f}{\partial x \partial y}$ for the function $f(x, y) = e^{x^2 y} + xy^3$.

Solution: Proceeding as before, we have

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2xye^{x^2 y} + y^3 & \frac{\partial f}{\partial y} &= x^2 e^{x^2 y} + 3xy^2 \\ \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}(2xye^{x^2 y} + y^3) & \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}(x^2 e^{x^2 y} + 3xy^2) \\ &= 2ye^{x^2 y} + 4x^2 y^2 e^{x^2 y} & &= x^4 e^{x^2 y} + 6xy \\ \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y}(2xye^{x^2 y} + y^3) & \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}(x^2 e^{x^2 y} + 3xy^2) \\ &= 2xe^{x^2 y} + 2x^3 ye^{x^2 y} + 3y^2 & &= 2xe^{x^2 y} + 2x^3 ye^{x^2 y} + 3y^2\end{aligned}$$

width height 0.5pt

Higher-order partial derivatives that are taken with respect to different variables, such as $\frac{\partial^2 f}{\partial y \partial x}$ and $\frac{\partial^2 f}{\partial x \partial y}$, are called **mixed partial derivatives**. Notice in the above example that $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$. It turns that this will usually be the case. Specifically, whenever both $\frac{\partial^2 f}{\partial y \partial x}$ and $\frac{\partial^2 f}{\partial x \partial y}$ are continuous at a point (a, b) , then they are equal at that point.^[2] All the functions we will deal with will have continuous partial derivatives of all orders, so you can assume in the remainder of the text that

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} \text{ for all } (x, y) \text{ in the domain of } f.$$

In other words, it doesn't matter in which order you take partial derivatives. This applies even to mixed partial derivatives of order 3 or higher.

The notation for partial derivatives varies. All of the following are equivalent:

$$\begin{aligned}
\frac{\partial f}{\partial x} &: f_x(x, y), \quad f_1(x, y), \quad D_x(x, y), \quad D_1(x, y) \\
\frac{\partial f}{\partial y} &: f_y(x, y), \quad f_2(x, y), \quad D_y(x, y), \quad D_2(x, y) \\
\frac{\partial^2 f}{\partial x^2} &: f_{xx}(x, y), \quad f_{11}(x, y), \quad D_{xx}(x, y), \quad D_{11}(x, y) \\
\frac{\partial^2 f}{\partial y^2} &: f_{yy}(x, y), \quad f_{22}(x, y), \quad D_{yy}(x, y), \quad D_{22}(x, y) \\
\frac{\partial^2 f}{\partial y \partial x} &: f_{xy}(x, y), \quad f_{12}(x, y), \quad D_{xy}(x, y), \quad D_{12}(x, y) \\
\frac{\partial^2 f}{\partial x \partial y} &: f_{yx}(x, y), \quad f_{21}(x, y), \quad D_{yx}(x, y), \quad D_{21}(x, y)
\end{aligned}$$

Exercises

A

For Exercises 1-16, find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

2

1. $f(x, y) = x^2 + y^2$

2. $f(x, y) = \cos(x + y)$

2

3. $f(x, y) = \sqrt{x^2 + y + 4}$

4. $f(x, y) = \frac{x + 1}{y + 1}$

2

5. $f(x, y) = e^{xy} + xy$

6. $f(x, y) = x^2 - y^2 + 6xy + 4x - 8y + 2$

2

7. $f(x, y) = x^4$

8. $f(x, y) = x + 2y$

2

9. $f(x, y) = \sqrt{x^2 + y^2}$

10. $f(x, y) = \sin(x + y)$

2

11. $f(x, y) = \sqrt[3]{x^2 + y + 4}$

12. $f(x, y) = \frac{xy + 1}{x + y}$

2

13. $f(x, y) = e^{-(x^2 + y^2)}$

14. $f(x, y) = \ln(xy)$

2

15. $f(x, y) = \sin(xy)$

16. $f(x, y) = \tan(x + y)$

For Exercises 17-26, find $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$ and $\frac{\partial^2 f}{\partial y \partial x}$ (use Exercises 1-8, 14, 15).

2

17. $f(x, y) = x^2 + y^2$

18. $f(x, y) = \cos(x + y)$

2

19. $f(x, y) = \sqrt{x^2 + y + 4}$

20. $f(x, y) = \frac{x + 1}{y + 1}$

2

21. $f(x, y) = e^{xy} + xy$

22. $f(x, y) = x^2 - y^2 + 6xy + 4x - 8y + 2$

2

23. $f(x, y) = x^4$

24. $f(x, y) = x + 2y$

2

25. $f(x, y) = \ln(xy)$

26. $f(x, y) = \sin(xy)$

B

27. Show that the function $f(x, y) = \sin(x + y) + \cos(x - y)$ satisfies the *wave equation*

$$\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} = 0 .$$

The wave equation is an example of a *partial differential equation*.

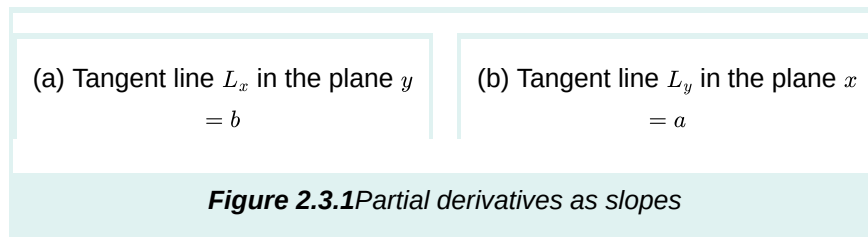
28. Let u and v be twice-differentiable functions of a single variable, and let $c \neq 0$ be a constant. Show that $f(x, y) = u(x + cy) + v(x - cy)$ is a solution of the *general one-dimensional wave equation*^[3]

$$\frac{\partial^2 f}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 f}{\partial y^2} = 0 .$$

-
1. It is not a Greek letter. The symbol was first used by the mathematicians A. Clairaut and L. Euler around 1740, to distinguish it from the letter d used for the “usual” derivative. ↩
 2. See pp. 214-216 in tm for a proof. ↩
 3. Conversely, it turns out that *any* solution must be of this form. See Ch. 1 in wei. ↩

2.3 Tangent Plane to a Surface

In the previous section we mentioned that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ can be thought of as the rate of change of a function $z = f(x, y)$ in the positive x and y directions, respectively. Recall that the derivative $\frac{dy}{dx}$ of a function $y = f(x)$ has a geometric meaning, namely as the slope of the tangent line to the graph of f at the point $(x, f(x))$ in $\mathbb{R}^2$. There is a similar geometric meaning to the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ of a function $z = f(x, y)$: given a point (a, b) in the domain D of $f(x, y)$, the trace of the surface described by $z = f(x, y)$ in the plane $y = b$ is a curve in $\mathbb{R}^3$ through the point $(a, b, f(a, b))$, and the slope of the tangent line L_x to that curve at that point is $\frac{\partial f}{\partial x}(a, b)$. Similarly, $\frac{\partial f}{\partial y}(a, b)$ is the slope of the tangent line L_y to the trace of the surface $z = f(x, y)$ in the plane $x = a$ (see Figure 2.3.1).



Since the derivative $\frac{dy}{dx}$ of a function $y = f(x)$ is used to find the tangent line to the graph of f (which is a curve in $\mathbb{R}^2$), you might expect that partial derivatives can be used to define a *tangent plane* to the graph of a surface $z = f(x, y)$. This indeed turns out to be the case. First, we need a definition of a tangent plane. The intuitive idea is that a tangent plane “just touches” a surface at a point. The formal definition mimics the intuitive notion of a tangent line to a curve.

Definition 2.4

Let $z = f(x, y)$ be the equation of a surface S in $\mathbb{R}^3$, and let $P = (a, b, c)$ be a point on S . Let T be a plane which contains the point P , and let $Q = (x, y, z)$ represent a generic point on the surface S . If the (acute) angle between the vector $\overrightarrow{PQ}$ and the plane T approaches zero as the point Q approaches P along the surface S , then we call T the **tangent plane** to S at P .

Note that since two lines in $\mathbb{R}^3$ determine a plane, then the two tangent lines to the surface $z = f(x, y)$ in the x and y directions described in Figure 2.3.1 are contained in the tangent plane at that point, *if the tangent plane exists at that point*. The existence of those two tangent lines does not by itself guarantee the existence of the tangent plane. It is possible that if we take the trace of the surface in the plane $x - y = 0$ (which makes a 45° angle with the positive x -axis), the resulting curve in that plane may have a tangent line which is not in the plane determined by the other two tangent lines, or it may not have a tangent line at all at that point. Luckily, it turns out^[1] that if $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist in a region around a point (a, b) and are continuous at (a, b) then the tangent plane to the surface $z = f(x, y)$ will exist at the point $(a, b, f(a, b))$. In this text, those conditions will always hold.

Suppose that we want an equation of the tangent plane T to the surface $z = f(x, y)$ at a point $(a, b, f(a, b))$. Let L_x and L_y be the tangent lines to the traces of the surface in the planes $y = b$ and $x = a$, respectively (as in Figure 2.3.2), and suppose that the conditions for T to exist do hold. Then the equation for T is

$$A(x - a) + B(y - b) + C(z - f(a, b)) = 0 \quad (2.4)$$

where $\mathbf{n} = (A, B, C)$ is a normal vector to the plane T . Since T contains the lines L_x and L_y , then all we need are vectors $\mathbf{v}_x$ and $\mathbf{v}_y$ that are parallel to L_x and L_y , respectively, and then let $\mathbf{n} = \mathbf{v}_x \times \mathbf{v}_y$.

Since the slope of L_x is $\frac{\partial f}{\partial x}(a, b)$, then the vector $\mathbf{v}_x = (1, 0, \frac{\partial f}{\partial x}(a, b))$ is parallel to L_x (since $\mathbf{v}_x$ lies in the xz -plane and lies in a line with slope $\frac{\frac{\partial f}{\partial x}(a, b)}{1} = \frac{\partial f}{\partial x}(a, b)$). See Figure 2.3.3). Similarly, the vector $\mathbf{v}_y = (0, 1, \frac{\partial f}{\partial y}(a, b))$ is parallel to L_y . Hence, the vector

$$\mathbf{n} = \mathbf{v}_x \times \mathbf{v}_y = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & \frac{\partial f}{\partial x}(a, b) \\ 0 & 1 & \frac{\partial f}{\partial y}(a, b) \end{vmatrix} = -\frac{\partial f}{\partial x}(a, b) \mathbf{i} - \frac{\partial f}{\partial y}(a, b) \mathbf{j} + \mathbf{k}$$

is normal to the plane T . Thus the equation of T is

$$-\frac{\partial f}{\partial x}(a, b)(x - a) - \frac{\partial f}{\partial y}(a, b)(y - b) + z - f(a, b) = 0. \quad (2.5)$$

Multiplying both sides by -1 , we have the following result:

The equation of the tangent plane to the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$ is

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) - z + f(a, b) = 0 \quad (2.6)$$

Example 2.13

Find the equation of the tangent plane to the surface $z = x^2 + y^2$ at the point $(1, 2, 5)$.

Solution: For the function $f(x, y) = x^2 + y^2$, we have $\frac{\partial f}{\partial x} = 2x$ and $\frac{\partial f}{\partial y} = 2y$, so the equation of the tangent plane at the point $(1, 2, 5)$ is

$$2(1)(x - 1) + 2(2)(y - 2) - z + 5 = 0, \text{ or} \\ 2x + 4y - z - 5 = 0.$$

width height 0.5pt

In a similar fashion, it can be shown that if a surface is defined implicitly by an equation of the form $F(x, y, z) = 0$, then the tangent plane to the surface at a point (a, b, c) is given by the equation

$$\frac{\partial F}{\partial x}(a, b, c)(x - a) + \frac{\partial F}{\partial y}(a, b, c)(y - b) + \frac{\partial F}{\partial z}(a, b, c)(z - c) = 0. \quad (2.7)$$

Note that formula (2.6) is the special case of formula (2.7) where $F(x, y, z) = f(x, y) - z$.

width height 0.5pt

Example 2.14

Find the equation of the tangent plane to the surface $x^2 + y^2 + z^2 = 9$ at the point $(2, 2, -1)$.

Solution: For the function $F(x, y, z) = x^2 + y^2 + z^2 - 9$, we have $\frac{\partial F}{\partial x} = 2x$, $\frac{\partial F}{\partial y} = 2y$, and $\frac{\partial F}{\partial z} = 2z$, so the equation of the tangent plane at $(2, 2, -1)$ is

$$2(2)(x - 2) + 2(2)(y - 2) + 2(-1)(z + 1) = 0, \text{ or} \\ 2x + 2y - z - 9 = 0.$$

EXERCISES

A

For Exercises 1-6, find the equation of the tangent plane to the surface $z = f(x, y)$ at the point P .

2

1. $f(x, y) = x^2 + y^3$, $P = (1, 1, 2)$

2. $f(x, y) = xy$, $P = (1, -1, -1)$

2

3. $f(x, y) = x^2y$, $P = (-1, 1, 1)$

4. $f(x, y) = xe^y$, $P = (1, 0, 1)$

2

5. $f(x, y) = x + 2y$, $P = (2, 1, 4)$

6. $f(x, y) = \sqrt{x^2 + y^2}$, $P = (3, 4, 5)$

For Exercises 7-10, find the equation of the tangent plane to the given surface at the point P .

2

7. $\frac{x^2}{4} + \frac{y^2}{9} + \frac{z^2}{16} = 1$, $P = \left(1, 2, \frac{2\sqrt{11}}{3}\right)$

8. $x^2 + y^2 + z^2 = 9$, $P = (0, 0, 3)$

2

9. $x^2 + y^2 - z^2 = 0$, $P = (3, 4, 5)$

10. $x^2 + y^2 = 4$, $P = (\sqrt{3}, 1, 0)$

1. See [§ 6.4]tm. ↩

2.4 Directional Derivatives and the Gradient

For a function $z = f(x, y)$, we learned that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ represent the (instantaneous) rate of change of f in the positive x and y directions, respectively. What about other directions? It turns out that we can find the rate of change in *any* direction using a more general type of derivative called a *directional derivative*.

Definition 2.5

Let $f(x, y)$ be a real-valued function with domain D in $\mathbb{R}^2$, and let (a, b) be a point in D . Let $\mathbf{v}$ be a unit vector in $\mathbb{R}^2$. Then the **directional derivative of f at (a, b) in the direction of $\mathbf{v}$** , denoted by $D_{\mathbf{v}}f(a, b)$, is defined as

$$D_{\mathbf{v}}f(a, b) = \lim_{h \rightarrow 0} \frac{f((a, b) + h\mathbf{v}) - f(a, b)}{h} \quad (2.8)$$

Notice in the definition that we seem to be treating the point (a, b) as a vector, since we are adding the vector $h\mathbf{v}$ to it. But this is just the usual idea of identifying vectors with their terminal points, which the reader should be used to by now. If we were to write the vector $\mathbf{v}$ as $\mathbf{v} = (v_1, v_2)$, then

$$D_{\mathbf{v}}f(a, b) = \lim_{h \rightarrow 0} \frac{f(a + hv_1, b + hv_2) - f(a, b)}{h} . \quad (2.9)$$

From this we can immediately recognize that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are special cases of the directional derivative with $\mathbf{v} = \mathbf{i} = (1, 0)$ and $\mathbf{v} = \mathbf{j} = (0, 1)$, respectively. That is, $\frac{\partial f}{\partial x} = D_{\mathbf{i}}f$ and $\frac{\partial f}{\partial y} = D_{\mathbf{j}}f$. Since there are many vectors with the same direction, we use a unit vector in the definition, as that represents a “standard” vector for a given direction.

If $f(x, y)$ has continuous partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ (which will always be the case in this text), then there is a simple formula for the directional derivative:

Theorem 2.2

Let $f(x, y)$ be a real-valued function with domain D in $\mathbb{R}^2$ such that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist and are continuous in D . Let (a, b) be a point in D , and let $\mathbf{v} = (v_1, v_2)$ be a unit vector in $\mathbb{R}^2$. Then

$$D_{\mathbf{v}}f(a, b) = v_1 \frac{\partial f}{\partial x}(a, b) + v_2 \frac{\partial f}{\partial y}(a, b) . \quad (2.10)$$

proof[Proof:] Note that if $\mathbf{v} = \mathbf{i} = (1, 0)$ then the above formula reduces to $D_{\mathbf{v}}f(a, b) = \frac{\partial f}{\partial x}(a, b)$, which we know is true since $D_{\mathbf{i}}f = \frac{\partial f}{\partial x}$, as we noted earlier. Similarly, for $\mathbf{v} = \mathbf{j} = (0, 1)$ the formula reduces to $D_{\mathbf{v}}f(a, b) = \frac{\partial f}{\partial y}(a, b)$, which is true since $D_{\mathbf{j}}f = \frac{\partial f}{\partial y}$. So since $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ are the only unit vectors in $\mathbb{R}^2$ with a zero component, then we need only show the formula holds for unit vectors $\mathbf{v} = (v_1, v_2)$ with $v_1 \neq 0$ and $v_2 \neq 0$. So fix such a vector $\mathbf{v}$ and fix a number $h \neq 0$. Then

$$f(a + hv_1, b + hv_2) - f(a, b) = f(a + hv_1, b + hv_2) - f(a + hv_1, b) + f(a + hv_1, b) - f(a, b) . \quad (2.11)$$

Since $h \neq 0$ and $v_2 \neq 0$, then $hv_2 \neq 0$ and thus any number c between b and $b + hv_2$ can be written as $c = b + \alpha hv_2$ for some number $0 < \alpha < 1$. So since the function $f(a + hv_1, y)$ is a real-valued function of y (since $a + hv_1$ is a fixed number), then the Mean Value Theorem from single-variable calculus can be applied to the function $g(y) = f(a + hv_1, y)$ on the interval $[b, b + hv_2]$ (or $[b + hv_2, b]$ if one of h or v_2 is negative) to find a number $0 < \alpha < 1$ such that

$$\frac{\partial f}{\partial y}(a + hv_1, b + \alpha hv_2) = g'(b + \alpha hv_2) = \frac{g(b + hv_2) - g(b)}{b + hv_2 - b} = \frac{f(a + hv_1, b + hv_2) - f(a + hv_1, b)}{hv_2}$$

and so

$$f(a + hv_1, b + hv_2) - f(a + hv_1, b) = hv_2 \frac{\partial f}{\partial y}(a + hv_1, b + \alpha hv_2).$$

By a similar argument, there exists a number $0 < \beta < 1$ such that

$$f(a + hv_1, b) - f(a, b) = hv_1 \frac{\partial f}{\partial x}(a + \beta hv_1, b).$$

Thus, by equation (2.11), we have

$$\begin{aligned} \frac{f(a + hv_1, b + hv_2) - f(a, b)}{h} &= \frac{hv_2 \frac{\partial f}{\partial y}(a + hv_1, b + \alpha hv_2) + hv_1 \frac{\partial f}{\partial x}(a + \beta hv_1, b)}{h} \\ &= v_2 \frac{\partial f}{\partial y}(a + hv_1, b + \alpha hv_2) + v_1 \frac{\partial f}{\partial x}(a + \beta hv_1, b) \end{aligned}$$

so by formula (2.9) we have

$$\begin{aligned} D_v f(a, b) &= \lim_{h \rightarrow 0} \frac{f(a + hv_1, b + hv_2) - f(a, b)}{h} \\ &= \lim_{h \rightarrow 0} \left[v_2 \frac{\partial f}{\partial y}(a + hv_1, b + \alpha hv_2) + v_1 \frac{\partial f}{\partial x}(a + \beta hv_1, b) \right] \\ &= v_2 \frac{\partial f}{\partial y}(a, b) + v_1 \frac{\partial f}{\partial x}(a, b) \quad \text{by the continuity of } \frac{\partial f}{\partial x} \text{ and } \frac{\partial f}{\partial y}, \text{ so} \\ D_v f(a, b) &= v_1 \frac{\partial f}{\partial x}(a, b) + v_2 \frac{\partial f}{\partial y}(a, b) \end{aligned}$$

after reversing the order of summation.

Note that $D_v f(a, b) = \mathbf{v} \cdot \left(\frac{\partial f}{\partial x}(a, b), \frac{\partial f}{\partial y}(a, b) \right)$. The second vector has a special name:

Definition 2.6

For a real-valued function $f(x, y)$, the **gradient** of f , denoted by ∇f , is the vector

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (2.12)$$

in $\mathbb{R}^2$. For a real-valued function $f(x, y, z)$, the gradient is the vector

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \quad (2.13)$$

in $\mathbb{R}^3$. The symbol ∇ is pronounced “del”.^[1]

Corollary 2.3

$$D_{\mathbf{v}}f = \mathbf{v} \cdot \nabla f$$

width height 0.5pt

Example 2.15

Find the directional derivative of $f(x, y) = xy^2 + x^3y$ at the point $(1, 2)$ in the direction of $\mathbf{v} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$.

Solution: We see that $\nabla f = (y^2 + 3x^2y, 2xy + x^3)$, so

$$D_{\mathbf{v}}f(1, 2) = \mathbf{v} \cdot \nabla f(1, 2) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \cdot (2^2 + 3(1)^2(2), 2(1)(2) + 1^3) = \frac{15}{\sqrt{2}}$$

width height 0.5pt

A real-valued function $z = f(x, y)$ whose partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist and are continuous is called *continuously differentiable*. Assume that $f(x, y)$ is such a function and that $\nabla f \neq 0$. Let c be a real number in the range of f and let $\mathbf{v}$ be a unit vector in $\mathbb{R}^2$ which is tangent to the level curve $f(x, y) = c$ (see Figure 2.4.1).

Figure 2.4.1

The value of $f(x, y)$ is constant along a level curve, so since $\mathbf{v}$ is a tangent vector to this curve, then the rate of change of f in the direction of $\mathbf{v}$ is 0, i.e. $D_{\mathbf{v}}f = 0$. But we know that $D_{\mathbf{v}}f = \mathbf{v} \cdot \nabla f = \|\mathbf{v}\| \|\nabla f\| \cos \theta$, where θ is the angle between $\mathbf{v}$ and ∇f . So since $\|\mathbf{v}\| = 1$ then $D_{\mathbf{v}}f = \|\nabla f\| \cos \theta$. So since $\nabla f \neq \mathbf{0}$ then $D_{\mathbf{v}}f = 0 \Rightarrow \cos \theta = 0 \Rightarrow \theta = 90^\circ$. In other words, $\nabla f \perp \mathbf{v}$, which means that ∇f is *normal* to the level curve.

In general, for any unit vector $\mathbf{v}$ in $\mathbb{R}^2$, we still have $D_{\mathbf{v}}f = \|\nabla f\| \cos \theta$, where θ is the angle between $\mathbf{v}$ and ∇f . At a fixed point (x, y) the length $\|\nabla f\|$ is fixed, and the value of $D_{\mathbf{v}}f$ then varies as θ varies. The largest value that $D_{\mathbf{v}}f$ can take is when $\cos \theta = 1$ ($\theta = 0^\circ$), while the smallest value occurs when $\cos \theta = -1$ ($\theta = 180^\circ$). In other words, the value of the function f increases the fastest in the direction of ∇f (since $\theta = 0^\circ$ in that case), and the value of f decreases the fastest in the direction of $-\nabla f$ (since $\theta = 180^\circ$ in that case). We have thus proved the following theorem:

Theorem 2.4

Let $f(x, y)$ be a continuously differentiable real-valued function, with $\nabla f \neq \mathbf{0}$. Then:

1. The gradient ∇f is normal to any level curve $f(x, y) = c$.
2. The value of $f(x, y)$ increases the fastest in the direction of ∇f .
3. The value of $f(x, y)$ decreases the fastest in the direction of $-\nabla f$.

width height 0.5pt

Example 2.16

In which direction does the function $f(x, y) = xy^2 + x^3y$ increase the fastest from the point $(1, 2)$? In which direction does it decrease the fastest?

Solution: Since $\nabla f = (y^2 + 3x^2y, 2xy + x^3)$, then $\nabla f(1, 2) = (10, 5) \neq \mathbf{0}$. A unit vector in that direction is $\mathbf{v} = \frac{\nabla f}{\|\nabla f\|} = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right)$. Thus, f increases the fastest in the direction of $\left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right)$ and decreases the fastest in the direction of $\left(\frac{-2}{\sqrt{5}}, \frac{-1}{\sqrt{5}} \right)$.

width height 0.5pt

Though we proved Theorem [2.4](#) for functions of two variables, a similar argument can be used to show that it also applies to functions of three or more variables. Likewise, the directional derivative in the three-dimensional case can also be defined by the formula $D_{\mathbf{v}}f = \mathbf{v} \cdot \nabla f$.

width height 0.5pt

Example 2.17

The temperature T of a solid is given by the function $T(x, y, z) = e^{-x} + e^{-2y} + e^{4z}$, where x, y, z are space coordinates relative to the center of the solid. In which direction from the point $(1, 1, 1)$ will the temperature decrease the fastest?

Solution: Since $\nabla f = (-e^{-x}, -2e^{-2y}, 4e^{4z})$, then the temperature will decrease the fastest in the direction of $-\nabla f(1, 1, 1) = (e^{-1}, 2e^{-2}, -4e^4)$.

width height 0.5pt

Exercises

A

For Exercises 1-10, compute the gradient ∇f .

2

1. $f(x, y) = x^2 + y^2 - 1$

2. $f(x, y) = \frac{1}{x^2 + y^2}$

2

3. $f(x, y) = \sqrt{x^2 + y^2 + 4}$

4. $f(x, y) = x^2 e^y$

2

5. $f(x, y) = \ln(xy)$

6. $f(x, y) = 2x + 5y$

2

7. $f(x, y, z) = \sin(xyz)$

8. $f(x, y, z) = x^2 e^{yz}$

2

9. $f(x, y, z) = x^2 + y^2 + z^2$

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

For Exercises 11-14, find the directional derivative of f at the point P in the direction of $\mathbf{v} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$.

2

11. $f(x, y) = x^2 + y^2 - 1, P = (1, 1)$

12. $f(x, y) = \frac{1}{x^2 + y^2}, P = (1, 1)$

2

13. $f(x, y) = \sqrt{x^2 + y^2 + 4}, P = (1, 1)$

14. $f(x, y) = x^2 e^y, P = (1, 1)$

For Exercises 15-16, find the directional derivative of f at the point P in the direction of $\mathbf{v} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$.

2

15. $f(x, y, z) = \sin(xyz)$, $P = (1, 1, 1)$
16. $f(x, y, z) = x^2 e^{yz}$, $P = (1, 1, 1)$
17. Repeat Example 2.16 at the point $(2, 3)$.
18. Repeat Example 2.17 at the point $(3, 1, 2)$.

B

For Exercises 19-26, let $f(x, y)$ and $g(x, y)$ be continuously differentiable real-valued functions, let c be a constant, and let $\mathbf{v}$ be a unit vector in $\mathbb{R}^2$. Show that:

2

19. $\nabla(cf) = c\nabla f$

20. $\nabla(f + g) = \nabla f + \nabla g$

2

21. $\nabla(fg) = f\nabla g + g\nabla f$

22. $\nabla(f/g) = \frac{g\nabla f - f\nabla g}{g^2}$ if $g(x, y) \neq 0$

2

23. $D_{-\mathbf{v}}f = -D_{\mathbf{v}}f$

24. $D_{\mathbf{v}}(cf) = cD_{\mathbf{v}}f$

2

25. $D_{\mathbf{v}}(f + g) = D_{\mathbf{v}}f + D_{\mathbf{v}}g$

26. $D_{\mathbf{v}}(fg) = fD_{\mathbf{v}}g + gD_{\mathbf{v}}f$

27. The function $r(x, y) = \sqrt{x^2 + y^2}$ is the length of the position vector $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$ for each point (x, y) in $\mathbb{R}^2$. Show that $\nabla r = \frac{1}{r}\mathbf{r}$ when $(x, y) \neq (0, 0)$, and that $\nabla(r^2) = 2\mathbf{r}$.

1. Sometimes the notation $\text{grad}(f)$ is used instead of ∇f . ↩

2.5 Maxima and Minima

The gradient can be used to find *extreme points* of real-valued functions of several variables, that is, points where the function has a *local maximum* or *local minimum*. We will consider only functions of two variables; functions of three or more variables require methods using linear algebra.

Definition 2.7

Let $f(x, y)$ be a real-valued function, and let (a, b) be a point in the domain of f . We say that f has a **local maximum** at (a, b) if $f(x, y) \leq f(a, b)$ for all (x, y) inside some disk of positive radius centered at (a, b) , i.e. there is some sufficiently small $r > 0$ such that $f(x, y) \leq f(a, b)$ for all (x, y) for which $(x - a)^2 + (y - b)^2 < r^2$.

Likewise, we say that f has a **local minimum** at (a, b) if $f(x, y) \geq f(a, b)$ for all (x, y) inside some disk of positive radius centered at (a, b) .

If $f(x, y) \leq f(a, b)$ for all (x, y) in the domain of f , then f has a **global maximum** at (a, b) . If $f(x, y) \geq f(a, b)$ for all (x, y) in the domain of f , then f has a **global minimum** at (a, b) .

Suppose that (a, b) is a local maximum point for $f(x, y)$, and that the first-order partial derivatives of f exist at (a, b) . We know that $f(a, b)$ is the largest value of $f(x, y)$ as (x, y) goes in all directions from the point (a, b) , in some sufficiently small disk centered at (a, b) . In particular, $f(a, b)$ is the largest value of f in the x direction (around the point (a, b)), that is, the single-variable function $g(x) = f(x, b)$ has a local maximum at $x = a$. So we know that $g'(a) = 0$. Since $g'(x) = \frac{\partial f}{\partial x}(x, b)$, then $\frac{\partial f}{\partial x}(a, b) = 0$. Similarly, $f(a, b)$ is the largest value of f near (a, b) in the y direction and so $\frac{\partial f}{\partial y}(a, b) = 0$. We thus have the following theorem:

Theorem 2.5

Let $f(x, y)$ be a real-valued function such that both $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ exist. Then a necessary condition for $f(x, y)$ to have a local maximum or minimum at (a, b) is that $\nabla f(a, b) = 0$.

Note: Theorem [2.5](#) can be extended to apply to functions of three or more variables.

A point (a, b) where $\nabla f(a, b) = 0$ is called a **critical point** for the function $f(x, y)$. So given a function $f(x, y)$, to find the critical points of f you have to solve the equations $\frac{\partial f}{\partial x}(x, y) = 0$ and $\frac{\partial f}{\partial y}(x, y) = 0$ simultaneously for (x, y) . Similar to the single-variable case, the *necessary* condition that $\nabla f(a, b) = 0$ is not always *sufficient* to guarantee that a critical point is a local maximum or minimum.

width height 0.5pt

Example 2.18

The function $f(x, y) = xy$ has a critical point at $(0, 0)$: $\frac{\partial f}{\partial x} = y = 0 \Rightarrow y = 0$, and $\frac{\partial f}{\partial y} = x = 0 \Rightarrow x = 0$, so $(0, 0)$ is the only critical point. But clearly f does not have a local maximum or minimum at $(0, 0)$ since any disk around $(0, 0)$ contains points (x, y) where the values of x and y have the same sign (so that $f(x, y) = xy > 0 = f(0, 0)$) and different signs (so that $f(x, y) = xy < 0 = f(0, 0)$). In fact, along the path $y = x$ in $\mathbb{R}^2$, $f(x, y) = x^2$, which has a local minimum at $(0, 0)$, while along the path $y = -x$ we have $f(x, y) = -x^2$, which has a local maximum at $(0, 0)$. So $(0, 0)$ is an example of a *saddle point*, i.e. it is a local maximum in one direction and a local minimum in another direction. The graph of $f(x, y)$ is shown in Figure 2.5.1, which is a hyperbolic paraboloid.

Figure 2.5.1 $f(x, y) = xy$, saddle point at $(0, 0)$

width height 0.5pt

The following theorem gives sufficient conditions for a critical point to be a local maximum or minimum of a *smooth function* (i.e. a function whose partial derivatives of all orders exist and are continuous), which we will not prove here.^[1]

Theorem 2.6

Let $f(x, y)$ be a smooth real-valued function, with a critical point at (a, b) (i.e. $\nabla f(a, b) = 0$). Define

$$D = \frac{\partial^2 f}{\partial x^2}(a, b) \frac{\partial^2 f}{\partial y^2}(a, b) - \left(\frac{\partial^2 f}{\partial y \partial x}(a, b) \right)^2$$

Then

1. if $D > 0$ and $\frac{\partial^2 f}{\partial x^2}(a, b) > 0$, then f has a local minimum at (a, b)
2. if $D > 0$ and $\frac{\partial^2 f}{\partial x^2}(a, b) < 0$, then f has a local maximum at (a, b)
3. if $D < 0$, then f has neither a local minimum nor a local maximum at (a, b)
4. if $D = 0$, then the test fails.

If condition (c) holds, then (a, b) is a *saddle point*. Note that the assumption that $f(x, y)$ is smooth means that

$$D = \begin{vmatrix} \frac{\partial^2 f}{\partial x^2}(a, b) & \frac{\partial^2 f}{\partial y \partial x}(a, b) \\ \frac{\partial^2 f}{\partial x \partial y}(a, b) & \frac{\partial^2 f}{\partial y^2}(a, b) \end{vmatrix}$$

since $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$. Also, if $D > 0$ then $\frac{\partial^2 f}{\partial x^2}(a, b) \frac{\partial^2 f}{\partial y^2}(a, b) = D + \left(\frac{\partial^2 f}{\partial y \partial x}(a, b) \right)^2 > 0$, and so $\frac{\partial^2 f}{\partial x^2}(a, b)$ and $\frac{\partial^2 f}{\partial y^2}(a, b)$ have the same sign. This means that in parts (a) and (b) of the theorem one can replace $\frac{\partial^2 f}{\partial x^2}(a, b)$ by $\frac{\partial^2 f}{\partial y^2}(a, b)$ if desired.

width height 0.5pt

Example 2.19

Find all local maxima and minima of $f(x, y) = x^2 + xy + y^2 - 3x$.

Solution: First find the critical points, i.e. where $\nabla f = 0$. Since

$$\frac{\partial f}{\partial x} = 2x + y - 3 \quad \text{and} \quad \frac{\partial f}{\partial y} = x + 2y$$

then the critical points (x, y) are the common solutions of the equations

$$\begin{aligned} 2x + y - 3 &= 0 \\ x + 2y &= 0 \end{aligned}$$

which has the unique solution $(x, y) = (2, -1)$. So $(2, -1)$ is the only critical point.

To use Theorem 2.6, we need the second-order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial y^2} = 2, \quad \frac{\partial^2 f}{\partial y \partial x} = 1$$

and so

$$D = \frac{\partial^2 f}{\partial x^2}(2, -1) \frac{\partial^2 f}{\partial y^2}(2, -1) - \left(\frac{\partial^2 f}{\partial y \partial x}(2, -1) \right)^2 = (2)(2) - 1^2 = 3 > 0$$

and $\frac{\partial^2 f}{\partial x^2}(2, -1) = 2 > 0$. Thus, $(2, -1)$ is a local minimum.

width height 0.5pt

Example 2.20

Find all local maxima and minima of $f(x, y) = xy - x^3 - y^2$.

Solution: First find the critical points, i.e. where $\nabla f = 0$. Since

$$\frac{\partial f}{\partial x} = y - 3x^2 \quad \text{and} \quad \frac{\partial f}{\partial y} = x - 2y$$

then the critical points (x, y) are the common solutions of the equations

$$\begin{aligned} y - 3x^2 &= 0 \\ x - 2y &= 0 \end{aligned}$$

The first equation yields $y = 3x^2$, substituting that into the second equation yields $x - 6x^2 = 0$, which has the solutions $x = 0$ and $x = \frac{1}{6}$. So $x = 0 \Rightarrow y = 3(0) = 0$ and $x = \frac{1}{6} \Rightarrow y = 3\left(\frac{1}{6}\right)^2 = \frac{1}{12}$.

So the critical points are $(x, y) = (0, 0)$ and $(x, y) = \left(\frac{1}{6}, \frac{1}{12}\right)$.

To use Theorem 2.6, we need the second-order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = -6x, \quad \frac{\partial^2 f}{\partial y^2} = -2, \quad \frac{\partial^2 f}{\partial y \partial x} = 1$$

So

$$D = \frac{\partial^2 f}{\partial x^2}(0, 0) \frac{\partial^2 f}{\partial y^2}(0, 0) - \left(\frac{\partial^2 f}{\partial y \partial x}(0, 0) \right)^2 = (-6(0))(-2) - 1^2 = -1 < 0$$

and thus $(0, 0)$ is a saddle point. Also,

$$D = \frac{\partial^2 f}{\partial x^2}\left(\frac{1}{6}, \frac{1}{12}\right) \frac{\partial^2 f}{\partial y^2}\left(\frac{1}{6}, \frac{1}{12}\right) - \left(\frac{\partial^2 f}{\partial y \partial x}\left(\frac{1}{6}, \frac{1}{12}\right) \right)^2 = \left(-6\left(\frac{1}{6}\right)\right)(-2) - 1^2 = 1 > 0$$

and $\frac{\partial^2 f}{\partial x^2}\left(\frac{1}{6}, \frac{1}{12}\right) = -1 < 0$. Thus, $\left(\frac{1}{6}, \frac{1}{12}\right)$ is a local maximum.

width height 0.5pt

Example 2.21

Find all local maxima and minima of $f(x, y) = (x - 2)^4 + (x - 2y)^2$.

Solution: First find the critical points, i.e. where $\nabla f = 0$. Since

$$\frac{\partial f}{\partial x} = 4(x - 2)^3 + 2(x - 2y) \quad \text{and} \quad \frac{\partial f}{\partial y} = -4(x - 2y)$$

then the critical points (x, y) are the common solutions of the equations

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

The second equation yields $x = 2y$, substituting that into the first equation yields $4(2y - 2)^3 = 0$, which has the solution $y = 1$, and so $x = 2(1) = 2$. Thus, $(2, 1)$ is the only critical point.

To use Theorem [2.6](#), we need the second-order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2} = 12(x - 2)^2 + 2, \quad \frac{\partial^2 f}{\partial y^2} = 8, \quad \frac{\partial^2 f}{\partial y \partial x} = -4$$

So

$$D = \frac{\partial^2 f}{\partial x^2}(2, 1) \frac{\partial^2 f}{\partial y^2}(2, 1) - \left(\frac{\partial^2 f}{\partial y \partial x}(2, 1) \right)^2 = (2)(8) - (-4)^2 = 0$$

and so the test fails. What can be done in this situation? Sometimes it is possible to examine the function to see directly the nature of a critical point. In our case, we see that $f(x, y) \geq 0$ for all (x, y) , since $f(x, y)$ is the sum of fourth and second powers of numbers and hence must be nonnegative. But we also see that $f(2, 1) = 0$. Thus $f(x, y) \geq 0 = f(2, 1)$ for all (x, y) , and hence $(2, 1)$ is in fact a *global* minimum for f .

width height 0.5pt

Example 2.22

Find all local maxima and minima of $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$.

Solution: First find the critical points, i.e. where $\nabla f = 0$. Since

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2x(1 - (x^2 + y^2))e^{-(x^2+y^2)} \\ \frac{\partial f}{\partial y} &= 2y(1 - (x^2 + y^2))e^{-(x^2+y^2)}\end{aligned}$$

then the critical points are $(0, 0)$ and all points (x, y) on the unit circle $x^2 + y^2 = 1$.

To use Theorem 2.6, we need the second-order partial derivatives:

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= 2[1 - (x^2 + y^2) - 2x^2 - 2x^2(1 - (x^2 + y^2))]e^{-(x^2+y^2)} \\ \frac{\partial^2 f}{\partial y^2} &= 2[1 - (x^2 + y^2) - 2y^2 - 2y^2(1 - (x^2 + y^2))]e^{-(x^2+y^2)} \\ \frac{\partial^2 f}{\partial y \partial x} &= -4xy[2 - (x^2 + y^2)]e^{-(x^2+y^2)}\end{aligned}$$

At $(0, 0)$, we have $D = 4 > 0$ and $\frac{\partial^2 f}{\partial x^2}(0, 0) = 2 > 0$, so $(0, 0)$ is a local minimum. However, for points (x, y) on the unit circle $x^2 + y^2 = 1$, we have

$$D = (-4x^2e^{-1})(-4y^2e^{-1}) - (-4xye^{-1})^2 = 0$$

and so the test fails. If we look at the graph of $f(x, y)$, as shown in Figure 2.5.2, it looks like we might have a local maximum for (x, y) on the unit circle $x^2 + y^2 = 1$. If we switch to using polar coordinates (r, θ) instead of (x, y) in $\mathbb{R}^2$, where $r^2 = x^2 + y^2$, then we see that we can write $f(x, y)$ as a function $g(r)$ of the variable r alone: $g(r) = r^2e^{-r^2}$. Then $g'(r) = 2r(1 - r^2)e^{-r^2}$, so it has a critical point at $r = 1$, and we can check that $g''(1) = -4e^{-1} < 0$, so the Second Derivative Test from single-variable calculus says that $r = 1$ is a local maximum. But $r = 1$ corresponds to the unit circle $x^2 + y^2 = 1$. Thus, the points (x, y) on the unit circle $x^2 + y^2 = 1$ are local maximum points for f .

Figure 2.5.2 $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$

EXERCISES

A

For Exercises 1-10, find all local maxima and minima of the function $f(x, y)$.

2

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

2. $f(x, y) = x^3 - 12x + y^2 + 8y$

2

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

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

2

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

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

2

7. $f(x, y) = \sqrt{x^2 + y^2}$

8. $f(x, y) = x + 2y$

2

9. $f(x, y) = 4x^2 - 4xy + 2y^2 + 10x - 6y$

10. $f(x, y) = -4x^2 + 4xy - 2y^2 + 16x - 12y$

B

11. For a rectangular solid of volume 1000 cubic meters, find the dimensions that will minimize the surface area. (*Hint: Use the volume condition to write the surface area as a function of just two variables.*)
12. Prove that if (a, b) is a local maximum or local minimum point for a smooth function $f(x, y)$, then the tangent plane to the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$ is parallel to the xy -plane. (*Hint: Use Theorem 2.5.*)

C

13. Find three positive numbers x, y, z whose sum is 10 such that x^2y^2z is a maximum.

1. See [§ 7.6]tm. ↩

2.6 Unconstrained Optimization: Numerical Methods

The types of problems that we solved in the previous section were examples of *unconstrained optimization* problems. That is, we tried to find local (and perhaps even global) maximum and minimum points of real-valued functions $f(x, y)$, where the points (x, y) could be any points in the domain of f . The method we used required us to find the critical points of f , which meant having to solve the equation $\nabla f = 0$, which in general is a system of two equations in two unknowns (x and y). While this was relatively simple for the examples we did, in general this will not be the case. If the equations involve polynomials in x and y of degree three or higher, or complicated expressions involving trigonometric, exponential, or logarithmic functions, then solving even one such equation, let alone two, could be impossible by elementary means.^[1]

For example, if one of the equations that had to be solved was

$$x^3 + 9x - 2 = 0 ,$$

you may have a hard time getting the exact solutions. Trial and error would not help much, especially since the only real solution^[2] turns out to be $\sqrt[3]{\sqrt{28} + 1} - \sqrt[3]{\sqrt{28} - 1}$. In a situation such as this, the only choice may be to find a solution using some numerical method which gives a sequence of numbers which converge to the actual solution. For example, Newton's method for solving equations $f(x) = 0$, which you probably learned in single-variable calculus. In this section we will describe another method of Newton for finding critical points of real-valued functions of two variables.

Let $f(x, y)$ be a smooth real-valued function, and define

$$D(x, y) = \frac{\partial^2 f}{\partial x^2}(x, y) \frac{\partial^2 f}{\partial y^2}(x, y) - \left(\frac{\partial^2 f}{\partial y \partial x}(x, y) \right)^2 .$$

Newton's algorithm: Pick an initial point (x_0, y_0) . For $n = 0, 1, 2, 3, \dots$, define:

$$x_{n+1} = x_n - \frac{\begin{vmatrix} \frac{\partial^2 f}{\partial y^2}(x_n, y_n) & \frac{\partial^2 f}{\partial x \partial y}(x_n, y_n) \\ \frac{\partial f}{\partial y}(x_n, y_n) & \frac{\partial f}{\partial x}(x_n, y_n) \end{vmatrix}}{D(x_n, y_n)}, \quad y_{n+1} = y_n - \frac{\begin{vmatrix} \frac{\partial^2 f}{\partial x^2}(x_n, y_n) & \frac{\partial^2 f}{\partial x \partial y}(x_n, y_n) \\ \frac{\partial f}{\partial x}(x_n, y_n) & \frac{\partial f}{\partial y}(x_n, y_n) \end{vmatrix}}{D(x_n, y_n)} \quad (2.14)$$

Then the sequence of points $(x_n, y_n)_{n=1}^{\infty}$ converges to a critical point. If there are several critical points, then you will have to try different initial points to find them.

Example 2.23

Find all local maxima and minima of $f(x, y) = x^3 - xy - x + xy^3 - y^4$.

Solution: First calculate the necessary partial derivatives:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 3x^2 - y - 1 + y^3, & \frac{\partial f}{\partial y} &= -x + 3xy^2 - 4y^3 \\ \frac{\partial^2 f}{\partial x^2} &= 6x, & \frac{\partial^2 f}{\partial y^2} &= 6xy - 12y^2, & \frac{\partial^2 f}{\partial y \partial x} &= -1 + 3y^2\end{aligned}$$

Notice that solving $\nabla f = 0$ would involve solving two third-degree polynomial equations in x and y , which in this case can not be done easily.

We need to pick an initial point (x_0, y_0) for our algorithm. Looking at the graph of $z = f(x, y)$ over a large region may help (see Figure [2.6.1](#) below), though it may be hard to tell where the critical points are.

Figure 2.6.1 $f(x, y) = x^3 - xy - x + xy^3 - y^4$ for $-20 \leq x \leq 20$ and $-20 \leq y \leq 20$

Notice in the formulas ([2.14](#)) that we divide by D , so we should pick an initial point where D is not zero. And we can see that $D(0, 0) = (0)(0) - (-1)^2 = -1 \neq 0$, so take $(0, 0)$ as our initial point. Since it may take a large number of iterations of Newton's algorithm to be sure that we are close enough to the actual critical point, and since the computations are quite tedious, we will let a computer do the computing. For this, we will write a simple program, using the Java programming language, which will take a given initial point as a parameter and then perform 100 iterations of Newton's algorithm. In each iteration the new point will be printed, so that we can see if there is convergence. The full code is shown in Listing [1](#).

```

//Program to find the critical points of
f(x,y)=x^3-xy-x+xy^3-y^4
public class newton {
    public static void main(String[] args) {
        //Get the initial point (x,y) as command-line
        parameters
        double x = Double.parseDouble(args[0]);
        //Initial x value
        double y = Double.parseDouble(args[1]);
        //Initial y value
        System.out.println("Initial point: (" + x +
        "," + y + ")");
        //Go through 100 iterations of Newton's
        algorithm
        for (int n=1; n<=100; n++) {
            double D = fxx(x,y)*fyy(x,y) -
            Math.pow(fxy(x,y),2);
            double xn = x; double yn = y; //The
            current x and y values
            if (D == 0) { //We can not divide by 0
                System.out.println("Error: D = 0 at
                iteration " + n);
                System.exit(0); //End the program
            }
            x = xn - (fxy(xn,yn)*fx(xn,yn) -
            fxy(xn,yn)*fy(xn,yn))/D;
            y = yn - (fxx(xn,yn)*fy(xn,yn) -
            fxy(xn,yn)*fx(xn,yn))/D;
            System.out.println("n = " + n + ": (" +
            x + "," + y + ")");
        }
    }
    //Below are the parts specific to the function
    f
    //The first partial derivative of f wrt x:
    3x^2-y-1+y^3
    public static double fx(double x, double y) {
        return 3*Math.pow(x,2) - y - 1 +
        Math.pow(y,3);
    }
    //The first partial derivative of f wrt y: -
    x+3xy^2-4y^3

```

To use this program, you should first save the code in Listing 1 in a plain text file called `newton.java`. You will need the Java Development Kit^[3] to compile the code. In the directory where `newton.java` is saved, run this command at the command prompt to compile the code:

```
javac newton.java
```

Then run the program with the initial point $(0,0)$ with this command:

```
java newton 0 0
```

Below is the output of the program using $(0,0)$ as the initial point, truncated to show the first 10 lines and the last 5 lines:

```

n = 1: (0.0, 0.0)
n = 2: (0.0, 0.0)
n = 3: (0.0, 0.0)
n = 4: (0.0, 0.0)
n = 5: (0.0, 0.0)
n = 6: (0.0, 0.0)
n = 7: (0.0, 0.0)
n = 8: (0.0, 0.0)
n = 9: (0.0, 0.0)
n = 10: (0.0, 0.0)
n = 11: (0.0, 0.0)
n = 12: (0.0, 0.0)
n = 13: (0.0, 0.0)
n = 14: (0.0, 0.0)
n = 15: (0.0, 0.0)
n = 16: (0.0, 0.0)
n = 17: (0.0, 0.0)
n = 18: (0.0, 0.0)
n = 19: (0.0, 0.0)
n = 20: (0.0, 0.0)
n = 21: (0.0, 0.0)
n = 22: (0.0, 0.0)
n = 23: (0.0, 0.0)
n = 24: (0.0, 0.0)
n = 25: (0.0, 0.0)
n = 26: (0.0, 0.0)
n = 27: (0.0, 0.0)
n = 28: (0.0, 0.0)
n = 29: (0.0, 0.0)
n = 30: (0.0, 0.0)
n = 31: (0.0, 0.0)
n = 32: (0.0, 0.0)
n = 33: (0.0, 0.0)
n = 34: (0.0, 0.0)
n = 35: (0.0, 0.0)
n = 36: (0.0, 0.0)
n = 37: (0.0, 0.0)
n = 38: (0.0, 0.0)
n = 39: (0.0, 0.0)
n = 40: (0.0, 0.0)
n = 41: (0.0, 0.0)
n = 42: (0.0, 0.0)
n = 43: (0.0, 0.0)
n = 44: (0.0, 0.0)
n = 45: (0.0, 0.0)
n = 46: (0.0, 0.0)
n = 47: (0.0, 0.0)
n = 48: (0.0, 0.0)
n = 49: (0.0, 0.0)
n = 50: (0.0, 0.0)
n = 51: (0.0, 0.0)
n = 52: (0.0, 0.0)
n = 53: (0.0, 0.0)
n = 54: (0.0, 0.0)
n = 55: (0.0, 0.0)
n = 56: (0.0, 0.0)
n = 57: (0.0, 0.0)
n = 58: (0.0, 0.0)
n = 59: (0.0, 0.0)
n = 60: (0.0, 0.0)
n = 61: (0.0, 0.0)
n = 62: (0.0, 0.0)
n = 63: (0.0, 0.0)
n = 64: (0.0, 0.0)
n = 65: (0.0, 0.0)
n = 66: (0.0, 0.0)
n = 67: (0.0, 0.0)
n = 68: (0.0, 0.0)
n = 69: (0.0, 0.0)
n = 70: (0.0, 0.0)
n = 71: (0.0, 0.0)
n = 72: (0.0, 0.0)
n = 73: (0.0, 0.0)
n = 74: (0.0, 0.0)
n = 75: (0.0, 0.0)
n = 76: (0.0, 0.0)
n = 77: (0.0, 0.0)
n = 78: (0.0, 0.0)
n = 79: (0.0, 0.0)
n = 80: (0.0, 0.0)
n = 81: (0.0, 0.0)
n = 82: (0.0, 0.0)
n = 83: (0.0, 0.0)
n = 84: (0.0, 0.0)
n = 85: (0.0, 0.0)
n = 86: (0.0, 0.0)
n = 87: (0.0, 0.0)
n = 88: (0.0, 0.0)
n = 89: (0.0, 0.0)
n = 90: (0.0, 0.0)
n = 91: (0.0, 0.0)
n = 92: (0.0, 0.0)
n = 93: (0.0, 0.0)
n = 94: (0.0, 0.0)
n = 95: (0.0, 0.0)
n = 96: (0.0, 0.0)
n = 97: (0.0, 0.0)
n = 98: (0.0, 0.0)
n = 99: (0.0, 0.0)
n = 100: (0.0, 0.0)

```

```

public static double fy(double x, double y) {
    return -x + 3*x*Math.pow(y,2) -
    java newton 0 0
    4^main now v x:
    Initial point: (0.0,0.0)
    n = 1: (0.0,-1.0)
    //the second partial derivative of f wrt x: 6x
    n = 2: (1.0,-0.5)
    public static double fxx(double x, double y) {
    n = 3: (0.6065857885615251,-0.44194107452339687)
    return h^x;
    n = 4: (0.484506572966545,-0.405341511995805)
    n = 5: (0.47123972682634485,-0.3966334583092305)
    //the second partial derivative of f wrt y:
    n = 6:
    h^xv-1+2v^2
    (0.47113558510349535,-0.39636450001936047)
    public static double fyy(double x, double y) {
    n = 7: (0.4711356343449705,-0.3963643379632247)
    return h^x^v - 1+2^Math.pow(y,2);
    n = 8: (0.4711356343449874,-0.39636433796318005)
    n = 9: (0.4711356343449874,-0.39636433796318005)
    //the mixed second partial derivative of f wrt
    n = 10:
    x and y: -1+2v^2
    (0.4711356343449874,-0.39636433796318005)
    public static double fxy(double x, double y) {
    ...return -1 + 3*Math.pow(y,2);
    n = 96:
    (0.4711356343449874,-0.39636433796318005)
    n = 97:
    (0.4711356343449874,-0.39636433796318005)
    n = 98: Listing Program listing for newton.java
    (0.4711356343449874,-0.39636433796318005)
    n = 99:
    (0.4711356343449874,-0.39636433796318005)

```

As you can see, we appear to have converged fairly quickly (after only 8 iterations) to what appears to be an actual critical point (up to Java's level of precision), namely the point $(0.4711356343449874, -0.39636433796318005)$. It is easy to confirm that $\nabla f = 0$ at this point, either by evaluating $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at the point ourselves, or by modifying our program to also print the values of the partial derivatives at the point. It turns out that both partial derivatives are indeed close enough to zero to be considered zero:

$$\begin{aligned}\frac{\partial f}{\partial x}(0.4711356343449874, -0.39636433796318005) &= 4.85722573273506 \times 10^{-17} \\ \frac{\partial f}{\partial y}(0.4711356343449874, -0.39636433796318005) &= -8.326672684688674 \times 10^{-17}\end{aligned}$$

We also have $D(0.4711356343449874, -0.39636433796318005) = -8.776075636032301 < 0$, so by Theorem 2.6 we know that $(0.4711356343449874, -0.39636433796318005)$ is a saddle point.

Since ∇f consists of cubic polynomials, it seems likely that there may be three critical points. The computer program makes experimenting with other initial points easy, and trying different values does indeed lead to different sequences which converge:

```

java newton -1 -1
Initial point: (-1.0, -1.0)
n = 1: (-0.5, -0.5)
n = 2:
(-0.49295774647887325, -0.08450704225352113)
n = 3: (-0.1855674752461383, -1.2047647348546167)
n = 4: (-0.4540060574531383, -0.8643989895639324)
n = 5: (-0.3672160534444, -0.5426077421319053)
n = 6:
(-0.4794622222856417, -0.24529117721011612)
n = 7: (0.11570743992954591, -2.4319791238981274)
n = 8:
(-0.05837851765533317, -1.6536079835854451)
n = 9: (-0.129841298650007, -1.121516233310142)
n = 10: (-1.004453014967208, -0.9206128022529645)
n = 11:
(-0.5161209914612475, -0.4176293491131443)
n = 12: (-0.5788664043863884, 0.2918236503332734)
n = 13:
(-0.6985177124230715, 0.49848120123515316)
n = 14: (-0.6733618916578702, 0.4345777963475479)
n = 15: (-0.6703832459238667, 0.42501465652420045)

```

Again, it is easy to confirm that both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ vanish at the point $(-0.6703832459238667, 0.42501465652420045)$, which means it is a critical point. And

$$D^2f(-0.6703832459238667, 0.42501465652420045) = \begin{pmatrix} -15.3853978526055 & 0 \\ 0 & -4.0222994755432 \end{pmatrix} < 0$$

so we know that $(-0.6703832459238667, 0.42501465652420045)$ is a local maximum. An idea of what the graph of f looks like near that point is shown in Figure 2.6.2, which does suggest a local maximum around that point.

```

...
n = 98:
(-0.6703832459238667, 0.42501465652420045)
Figure 2.6.2  $f(x, y) = x^3 - xy - x + xy^3 - y^4$  for  $-1 \leq x \leq 0$  and  $0 \leq y$ 
(-0.6703832459238667, 0.42501465652420045)
n = 100:
(-7.540962756992551, -5.595509445899435)

```

Finally, running the computer program with the initial point $(-7.540962756992551, -5.595509445899435)$ yields the critical point $(-7.540962756992551, -5.595509445899435)$, with $D < 0$ at that point, which makes it a saddle point.

We can summarize our findings for the function $f(x, y) = x^3 - xy - x + xy^3 - y^4$:

$(0.4711356343449874, -0.39636433796318005)$: saddle point
 $(-0.6703832459238667, 0.42501465652420045)$: local maximum
 $(-7.540962756992551, -5.595509445899435)$: saddle point

width height 0.5pt

The derivation of Newton's algorithm, and the proof that it converges (given a "reasonable" choice for the initial point) requires techniques beyond the scope of this text. See [rr](#) for more detail and for discussion of other numerical methods. Our description of Newton's algorithm is the special two-variable case of a more general algorithm that can be applied to functions of $n \geq 2$ variables.

In the case of functions which have a global maximum or minimum, Newton's algorithm can be used to find those points. In general, global maxima and minima tend to be more interesting than local versions, at least in practical applications. A maximization problem can always be turned into a minimization problem (why?), so a large number of methods have been developed to find the global minimum of functions of any number of variables. This field of study is called *nonlinear programming*. Many of these methods are based on the *steepest descent* technique, which is based on an idea that we discussed in Section 2.4. Recall that the negative gradient $-\nabla f$ gives the direction of the fastest rate of decrease of a function f . The crux of the steepest descent idea, then, is that starting from some initial point, you move a certain amount in the direction of $-\nabla f$ at that point. Wherever that takes you becomes your new point, and you then just keep repeating that procedure until eventually (hopefully) you reach the point where f has its smallest value. There is a "pure" steepest descent method, and a multitude of variations on it that improve the rate of convergence, ease of calculation, etc. In fact, Newton's algorithm can be interpreted as a modified steepest descent method. For more discussion of this, and of nonlinear programming in general, see [bss](#).

EXERCISES

C

1. Recall Example 2.21 from the previous section, where we showed that the point $(2, 1)$ was a global minimum for the function $f(x, y) = (x - 2)^4 + (x - 2y)^2$. Notice that our computer program can be modified fairly easily to use this function (just change the return values in the `fx`, `fy`, `fxx`, `fyy` and `fxxy` function definitions to use the appropriate partial derivative). Either modify that program or write one of your own in a programming language of your choice to show that Newton's algorithm does lead to the point $(2, 1)$. First use the initial point $(0, 3)$, then use the initial point $(3, 2)$, and compare the results. Make sure that your program attempts to do 100 iterations of the algorithm. Did anything strange happen when your program ran? If so, how do you explain it? (Hint: *Something strange should happen.*)
2. There is a version of Newton's algorithm for solving a system of two equations

$$f_1(x, y) = 0 \quad \text{and} \quad f_2(x, y) = 0 ,$$

where $f_1(x, y)$ and $f_2(x, y)$ are smooth real-valued functions:

Pick an initial point (x_0, y_0) . For $n = 0, 1, 2, 3, \dots$, define:

$$x_{n+1} = x_n - \frac{\begin{vmatrix} f_1(x_n, y_n) & f_2(x_n, y_n) \\ \frac{\partial f_1}{\partial y}(x_n, y_n) & \frac{\partial f_2}{\partial y}(x_n, y_n) \end{vmatrix}}{D(x_n, y_n)} , \quad y_{n+1} = y_n + \frac{\begin{vmatrix} f_1(x_n, y_n) & f_2(x_n, y_n) \\ \frac{\partial f_1}{\partial x}(x_n, y_n) & \frac{\partial f_2}{\partial x}(x_n, y_n) \end{vmatrix}}{D(x_n, y_n)} , \text{ where}$$

$$D(x_n, y_n) = \frac{\partial f_1}{\partial x}(x_n, y_n) \frac{\partial f_2}{\partial y}(x_n, y_n) - \frac{\partial f_1}{\partial y}(x_n, y_n) \frac{\partial f_2}{\partial x}(x_n, y_n) .$$

Then the sequence of points $(x_n, y_n)_{n=1}^{\infty}$ converges to a solution. Write a computer program that uses this algorithm to find approximate solutions to the system of equations

$$f_1(x, y) = \sin(xy) - x - y = 0 \quad \text{and} \quad f_2(x, y) = e^{2x} - 2x + 3y = 0 .$$

Show that you get two different solutions when using $(0, 0)$ and $(1, 1)$ for the initial point (x_0, y_0) .

1. This is also a problem for the equivalent method (the Second Derivative Test) in single-variable calculus, though one that is not usually emphasized. ↩
2. There are also two nonreal, complex number solutions. Cubic polynomial equations in one variable can be solved using *Cardan's formulas*, which are not quite as simple as the familiar quadratic formula. See usp for more details. There are formulas for solving polynomial equations of degree 4, but it can be proved that there is no general formula for solving equations for polynomials of degree five or higher. ↩
3. Available for free at <http://www.oracle.com/technetwork/java/javase/downloads/> ↩

2.7 Constrained Optimization: Lagrange Multipliers

In Sections 2.5 and 2.6 we were concerned with finding maxima and minima of functions without any constraints on the variables (other than being in the domain of the function). What would we do if there were constraints on the variables? The following example illustrates a simple case of this type of problem.

width height 0.5pt

Example 2.24

For a rectangle whose perimeter is 20 m, find the dimensions that will maximize the area.

Solution: The area A of a rectangle with width x and height y is $A = xy$. The perimeter P of the rectangle is then given by the formula $P = 2x + 2y$. Since we are given that the perimeter $P = 20$, this problem can be stated as:

$$\begin{aligned} \text{Maximize : } & f(x, y) = xy \\ \text{given : } & 2x + 2y = 20 \end{aligned}$$

The reader is probably familiar with a simple method, using single-variable calculus, for solving this problem. Since we must have $2x + 2y = 20$, then we can solve for, say, y in terms of x using that equation. This gives $y = 10 - x$, which we then substitute into f to get $f(x, y) = xy = x(10 - x) = 10x - x^2$. This is now a function of x alone, so we now just have to maximize the function $f(x) = 10x - x^2$ on the interval $[0, 10]$. Since $f'(x) = 10 - 2x = 0 \Rightarrow x = 5$ and $f''(5) = -2 < 0$, then the Second Derivative Test tells us that $x = 5$ is a local maximum for f , and hence $x = 5$ must be the global maximum on the interval $[0, 10]$ (since $f = 0$ at the endpoints of the interval). So since $y = 10 - x = 5$, then the maximum area occurs for a rectangle whose width and height both are 5 m.

width height 0.5pt

Notice in the above example that the ease of the solution depended on being able to solve for one variable in terms of the other in the equation $2x + 2y = 20$. But what if that were not possible (which is often the case)? In this section we will use a general method, called the *Lagrange multiplier method*^[1], for solving *constrained optimization* problems:

$$\begin{aligned} \text{Maximize (or minimize) : } & f(x, y) \text{ (or } f(x, y, z)) \\ \text{given : } & g(x, y) = c \text{ (or } g(x, y, z) = c) \text{ for some constant } c \end{aligned}$$

The equation $g(x, y) = c$ is called the *constraint equation*, and we say that x and y are *constrained* by $g(x, y) = c$. Points (x, y) which are maxima or minima of $f(x, y)$ with the condition that they satisfy the constraint equation $g(x, y) = c$ are called *constrained maximum* or *constrained minimum* points, respectively. Similar definitions hold for functions of three variables.

The Lagrange multiplier method for solving such problems can now be stated:

Theorem 2.7

Let $f(x, y)$ and $g(x, y)$ be smooth functions, and suppose that c is a scalar constant such that $\nabla g(x, y) \neq \mathbf{0}$ for all (x, y) that satisfy the equation $g(x, y) = c$. Then to solve the constrained optimization problem

$$\begin{aligned} &\text{Maximize (or minimize) : } f(x, y) \\ &\text{given : } g(x, y) = c, \end{aligned}$$

find the points (x, y) that solve the equation $\nabla f(x, y) = \lambda \nabla g(x, y)$ for some constant λ (the number λ is called the *Lagrange multiplier*). If there is a constrained maximum or minimum, then it must be such a point.

A rigorous proof of the above theorem requires use of the Implicit Function Theorem, which is beyond the scope of this text.^[2] Note that the theorem only gives a *necessary* condition for a point to be a constrained maximum or minimum. Whether a point (x, y) that satisfies $\nabla f(x, y) = \lambda \nabla g(x, y)$ for some λ actually *is* a constrained maximum or minimum can *sometimes* be determined by the nature of the problem itself. For instance, in Example 2.24 it was clear that there had to be a global maximum.

So how can you tell when a point that satisfies the condition in Theorem 2.7 really is a constrained maximum or minimum? The answer is that it depends on the constraint function $g(x, y)$, together with any implicit constraints. It can be shown^[3] that if the constraint equation $g(x, y) = c$ (plus any hidden constraints) describes a *bounded* set B in $\mathbb{R}^2$, then the constrained maximum or minimum of $f(x, y)$ will occur either at a point (x, y) satisfying $\nabla f(x, y) = \lambda \nabla g(x, y)$ or at a “boundary” point of the set B .

In Example 2.24 the constraint equation $2x + 2y = 20$ describes a line in $\mathbb{R}^2$, which by itself is not bounded. However, there are “hidden” constraints, due to the nature of the problem, namely $0 \leq x, y \leq 10$, which cause that line to be restricted to a *line segment* in $\mathbb{R}^2$ (including the endpoints of that line segment), which *is* bounded.

width height 0.5pt

Example 2.25

For a rectangle whose perimeter is 20 m, use the Lagrange multiplier method to find the dimensions that will maximize the area.

Solution: As we saw in Example 2.24, with x and y representing the width and height, respectively, of the rectangle, this problem can be stated as:

$$\begin{aligned} \text{Maximize : } & f(x, y) = xy \\ \text{given : } & g(x, y) = 2x + 2y = 20 \end{aligned}$$

Then solving the equation $\nabla f(x, y) = \lambda \nabla g(x, y)$ for some λ means solving the equations $\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}$ and $\frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}$, namely:

$$\begin{aligned} y &= 2\lambda, \\ x &= 2\lambda \end{aligned}$$

The general idea is to solve for λ in both equations, then set those expressions equal (since they both equal λ) to solve for x and y . Doing this we get

$$\frac{y}{2} = \lambda = \frac{x}{2} \Rightarrow x = y,$$

so now substitute either of the expressions for x or y into the constraint equation to solve for x and y :

$$20 = g(x, y) = 2x + 2y = 2x + 2x = 4x \Rightarrow x = 5 \Rightarrow y = 5$$

There must be a maximum area, since the minimum area is 0 and $f(5, 5) = 25 > 0$, so the point $(5, 5)$ that we found (called a *constrained critical point*) must be the constrained maximum.

∴ The maximum area occurs for a rectangle whose width and height both are 5 m.

width height 0.5pt

Example 2.26

Find the points on the circle $x^2 + y^2 = 80$ which are closest to and farthest from the point $(1, 2)$.

Solution: The distance d from any point (x, y) to the point $(1, 2)$ is

$$d = \sqrt{(x-1)^2 + (y-2)^2},$$

and minimizing the distance is equivalent to minimizing the square of the distance. Thus the problem can be stated as:

$$\begin{aligned} \text{Maximize (and minimize) : } f(x, y) &= (x-1)^2 + (y-2)^2 \\ \text{given : } g(x, y) &= x^2 + y^2 = 80 \end{aligned}$$

Solving $\nabla f(x, y) = \lambda \nabla g(x, y)$ means solving the following equations:

$$\begin{aligned} 2(x-1) &= 2\lambda x, \\ 2(y-2) &= 2\lambda y \end{aligned}$$

Note that $x \neq 0$ since otherwise we would get $-2 = 0$ in the first equation. Similarly, $y \neq 0$. So we can solve both equations for λ as follows:

$$\frac{x-1}{x} = \lambda = \frac{y-2}{y} \Rightarrow xy - y = xy - 2x \Rightarrow y = 2x$$

Substituting this into $g(x, y) = x^2 + y^2 = 80$ yields $5x^2 = 80$, so $x = \pm 4$. So the two constrained critical points are $(4, 8)$ and $(-4, -8)$. Since $f(4, 8) = 45$ and $f(-4, -8) = 125$, and since there must be points on the circle closest to and farthest from $(1, 2)$, then it must be the case that $(4, 8)$ is the point on the circle closest to $(1, 2)$ and $(-4, -8)$ is the farthest from $(1, 2)$ (see Figure 2.7.1).

Notice that since the constraint equation $x^2 + y^2 = 80$ describes a circle, which is a bounded set in $\mathbb{R}^2$, then we were guaranteed that the constrained critical points we found were indeed the constrained maximum and minimum.

width height 0.5pt

The Lagrange multiplier method can be extended to functions of three variables.

width height 0.5pt

Example 2.27

$$\begin{aligned} \text{Maximize (and minimize) : } & f(x, y, z) = x + z \\ \text{given : } & g(x, y, z) = x^2 + y^2 + z^2 = 1 \end{aligned}$$

Solution: Solve the equation $\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$:

$$\begin{aligned} 1 &= 2\lambda x \\ 0 &= 2\lambda y \\ 1 &= 2\lambda z \end{aligned}$$

The first equation implies $\lambda \neq 0$ (otherwise we would have $1 = 0$), so we can divide by λ in the second equation to get $y = 0$ and we can divide by λ in the first and third equations to get $x = \frac{1}{2\lambda} = z$. Substituting these expressions into the constraint equation $g(x, y, z) = x^2 + y^2 + z^2 = 1$ yields the constrained critical points $\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$ and $\left(\frac{-1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right)$. Since $f\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) > f\left(\frac{-1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right)$, and since the constraint equation $x^2 + y^2 + z^2 = 1$ describes a sphere (which is bounded) in $\mathbb{R}^3$, then $\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$ is the constrained maximum point and $\left(\frac{-1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right)$ is the constrained minimum point.

width height 0.5pt

So far we have not attached any significance to the value of the Lagrange multiplier λ . We needed λ only to find the constrained critical points, but made no use of its value. It turns out that λ gives an approximation of the change in the value of the function $f(x, y)$ that we wish to maximize or minimize, when the constant c in the constraint equation $g(x, y) = c$ is changed by 1.

For example, in Example [2.25](#) we showed that the constrained optimization problem

$$\begin{aligned} \text{Maximize : } & f(x, y) = xy \\ \text{given : } & g(x, y) = 2x + 2y = 20 \end{aligned}$$

had the solution $(x, y) = (5, 5)$, and that $\lambda = x/2 = y/2$. Thus, $\lambda = 2.5$. In a similar fashion we could show that the constrained optimization problem

$$\begin{aligned} \text{Maximize : } & f(x, y) = xy \\ \text{given : } & g(x, y) = 2x + 2y = 21 \end{aligned}$$

has the solution $(x, y) = (5.25, 5.25)$. So we see that the value of $f(x, y)$ at the constrained maximum increased from $f(5, 5) = 25$ to $f(5.25, 5.25) = 27.5625$, i.e. it increased by 2.5625 when we increased the value of c in the constraint equation $g(x, y) = c$ from $c = 20$ to $c = 21$. Notice that $\lambda = 2.5$ is close to 2.5625, that is,

$$\lambda \approx \Delta f = f(\text{new max. pt}) - f(\text{old max. pt}) .$$

Finally, note that solving the equation $\nabla f(x, y) = \lambda \nabla g(x, y)$ means having to solve a system of two (possibly nonlinear) equations in three unknowns, which as we have seen before, may not be possible to do. And the 3-variable case can get even more complicated. All of this somewhat restricts the usefulness of Lagrange's method to relatively simple functions. Luckily there are many numerical methods for solving constrained optimization problems, though we will not discuss them here.^[4]

EXERCISES

A

1. Find the constrained maxima and minima of $f(x, y) = 2x + y$ given that $x^2 + y^2 = 4$.
2. Find the constrained maxima and minima of $f(x, y) = xy$ given that $x^2 + 3y^2 = 6$.
3. Find the points on the circle $x^2 + y^2 = 100$ which are closest to and farthest from the point $(2, 3)$.

B

4. Find the constrained maxima and minima of $f(x, y, z) = x + y^2 + 2z$ given that $4x^2 + 9y^2 - 36z^2 = 36$.
5. Find the volume of the largest rectangular parallelepiped that can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 .$$

-
1. Named after the French mathematician Joseph Louis Lagrange (1736-1813). ↩
 2. See [§ 6.8]tm for more detail. ↩
 3. Again, see tm. ↩
 4. See bss. ↩