Part II: Functions of Several Variables

Week 4: Differentiation for Functions of Several Variables

Introduction

A functions of several variables

$$f : U \subseteq \mathbb{R}^n \to \mathbb{R}$$

is a rule that assigns a real number to each point in U, a subset of $\mathbb{R}^n$,

For the next four weeks we are going to study the differential and integral calculus of such functions.

Letting $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ denote a point in U and w the corresponding value in $\mathbb{R}$, we write

$$w = f(x_1, x_2, \cdots, x_n) \qquad \text{or} \qquad w = f(\mathbf{x})$$

We sometimes call f a function of n variables, or say f is a function on $\mathbb{R}^n$ (meaning perhaps a subset of $\mathbb{R}^n$).

When we focus specifically on $n = 2$ or $n = 3$ we commonly write

$$w = f(x, y) \qquad \text{or} \qquad w = f(x, y, z)$$

In fact when we consider graphs (see below) for $n = 2$ we frequently use z for the dependent variable, e.g. $z = f(x, y)$.

Many physical systems are expressed as functions of several variables and the governing laws are expressed in the calculus of such functions. Consider for example the temperature in a room. Temperature is a real number that will be a function of both position and time. Call this function T, so $T(x, y, z, t)$ is the temperature at position (x, y, z) and time t. Under certain assumptions the physical law governing the evolution of temperature is:

$$\frac{\partial T}{\partial t}(x,y,z,t) = \frac{\partial^2 T}{\partial x^2}(x,y,z,t) + \frac{\partial^2 T}{\partial y^2}(x,y,z,t) + \frac{\partial^2 T}{\partial z^2}(x,y,z,t)$$

You are familiar with ordinary differential equations. This is a **partial differential equation**. By the end of this week you will understand what these symbols mean, and given a function $T(x, y, z, t)$, you will be able to verify whether it satisfies the equation. Finding solutions to partial differential equations will come in later years.

Visualising functions on $\mathbb{R}^n$

There are two primary ways to visualise functions of several variables: graphs for $n = 2$ and level set for $n = 2$ and $n = 3$. One can also make movies of graphs or level sets, and thereby visualise functions of up to four variables. For larger n visualisation is very difficult.

Graphs

For $n = 2$, f can be visualised as the graph

$$G_f = \{(x, y, z) | \, (x, y) \in U, z = f(x, y)\}$$

The function is seen as a sheet of height $f(x, y)$ above or below each point (x, y).

Level sets

Level sets, also called contours in $\mathbb{R}^2$ or isosurfaces in R^3, are subsets of U which are all mapped to the same value by f. Formally, the definition is

> The **level sets** of a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ are sets of points
>
> $$\mathcal{L}_k = \{\mathbf{x} \in U | \, f(\mathbf{x}) = k\}$$
>
> for each constant k in the range of f.

The intuition is easy for function on $\mathbb{R}^2$. Plot the graph $z = f(x, y)$ then intersect the graph with plane $z = k$ for some constant k. Project the intersection points down onto $\mathbb{R}^2$ and these will make up the contour for this value of k. Typically contours will be curves in $\mathbb{R}^2$. To represent a function using contours one typically plots several contours with the corresponding values of k labelled in some way. These are call **contour maps** or **contour plots**. The concept is familiar from topographic maps, weather maps, and such.

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$: $f(x, y) = x^2 + y^2$. Then for any $r \in \mathbb{R}$, $\mathcal{L}_{r^2} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = r^2\}$ - circle of radius $|r|$ centred on the origin.

The level sets of a function of three variables $f(x, y, z)$ are typically surfaces in $\mathbb{R}^3$ called **isosurfaces** (iso meaning equal, so a surface of equal, i.e. constant, value of f). Using transparency or clipping, computers can often make several isosurfaces visible simultaneously, allowing for good understanding of the underlying function f.

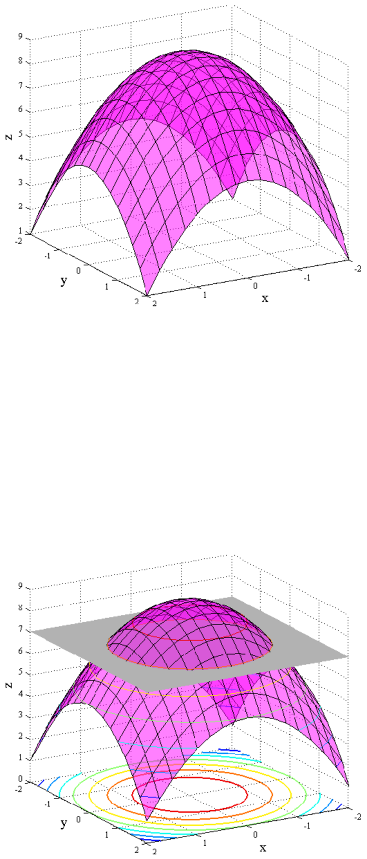


Figure 1: The function $f(x,y) = 9 - x^2 - y^2$ visualised as a graph (top) and as a contour plot (bottom) by slicing the graph at constant heights.

4.1 Caution on the extension to $\mathbb{R}^n$

Extending analysis from function $f : \mathbb{R} \to \mathbb{R}$ to functions $f : \mathbb{R}^n \to \mathbb{R}$ is more involved than it might at first appear. One can see this from the example function

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

What is the limit of $f(x,y)$ as $(x,y) \to (0,0)$? Since $f = 0$ when either $x = 0$ or $y = 0$ the limit $f(x,y)$ approaching the origin along either the x or y axis is 0. This is also the limit approaching the origin along any line $y = mx$. Hence is might seem that the $\lim_{(x,y)\to(0,0)} f(x,y) = 0$. However, along the curve $x = y^2$, we have $f(x = y^2, y) = y^4/2y^4 = 1/2$.

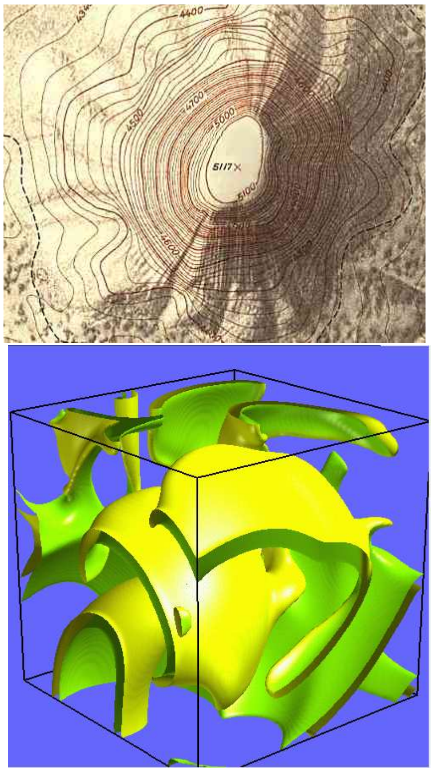


Figure 2: Example of a topographic map (contour plot) and an isosurface. (Map reproduced from http://mail.colonial.net/∼hkaiter/topographic_maps)

Hence, if the origin is approached along this curve the limit is 1/2. Since one obtains different values depending on how $(0, 0)$ is approached, the limit does not, in fact, exist.

This illustrates that limits and continuity for functions on $\mathbb{R}^n$ cannot be view from a one-dimensional perspective, but must be properly generalised using regions (called neighbourhoods) in $\mathbb{R}^n$. This will be covered in later Analysis modules and in Differentiation. While we will not define these things here, we will sometimes state properties that hold for continuous functions. You will just have to take this on faith for the present.

Fortunately, several of the most important aspects of mulivariable calculus are "one dimensional" and follow easily from things you know. Nothing stops us from being able to define and do calculations using these quantities.

4.2 Partial Derivatives

Partial derivatives are easy. For simplicity we initially restrict to the case $f : \mathbb{R}^2 \to \mathbb{R}$ and define

The **partial derivatives** of function f with respect to x and y at the point (a, b) are

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

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

(Assuming the limits exist.)

Sometimes partial derivatives are indicated by subscripts, e.g. $f_x(a, b)$ and $f_y(a, b)$, or $f_1(a, b)$ and $f_2(a, b)$. Sometimes upper case D is used, e.g. $D_x(a, b)$ and $D_y(a, b)$. We will not use any of these notations. We will on occasion use the following. Letting $z = f(x, y)$ we will denote partial derivatives of f by $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

Interpretation

What this definition says is that *the partial derivative of f with respect to x is just the ordinary one-dimensional derivative treating y as a fixed constant.* Concretely, define $g(x) = f(x, b)$ and then compute the ordinary derivative dg/dx at a. This is the partial derivative of f with respect to x. The partial derivative with respect to y is analogous. So the pedantic view of partial differentiation is:

Given $f(x, y)$,

$$\begin{cases} \text{let } g(x) = f(x,b), \text{ then } \dfrac{\partial f}{\partial x}(a,b) = \dfrac{dg}{dx}(a) \\ \\ \text{let } h(y) = f(a,y), \text{ then } \dfrac{\partial f}{\partial y}(a,b) = \dfrac{dh}{dy}(b) \end{cases}$$

This is illustrated in the following pictures. The function $g(x)$ is obtained by slicing f with a plane $y = b$, and similarly for $h(y)$.

While defining the auxiliary functions $g(x)$ and $h(y)$ is pedagogically useful for explaining partial derivatives, in practice it is unnecessary to explicitly form these functions. You will quickly master computing partial derivatives by doing examples.

Partial derivatives are functions

In the above definition we defined the partial derivatives $\frac{\partial f}{\partial x}(a, b)$ and $\frac{\partial f}{\partial y}(a, b)$ at a point (a, b). If

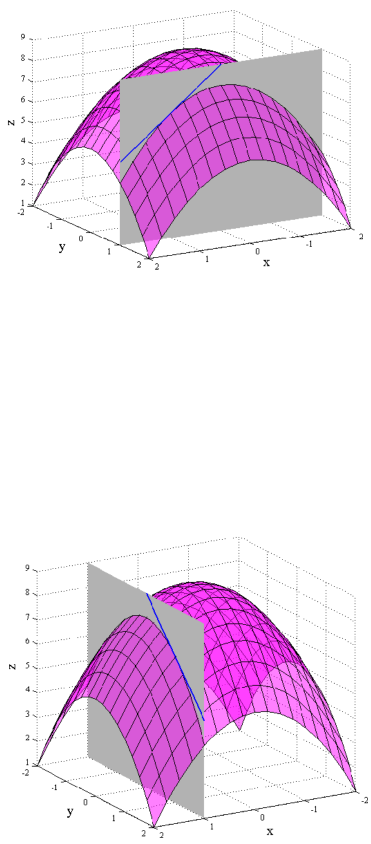


Figure 3: The x partial derivative (top) and y partial derivative (bottom) of the function $f(x, y) = 9 - x^2 - y^2$.

we allow this point to vary, then each partial derivatives will itself be a function of (x, y). In which case we would write

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}$$

to denote the functions and

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

to denote the values of these functions at the point (x, y). Sometimes vertical bars are used to indicate this evaluation, e.g.

$$\frac{\partial f}{\partial x}(a, b) = \left.\frac{\partial f}{\partial x}\right|_{(a,b)}$$

You are of course already familiar with everything just stated from functions of one variable. The derivative, f', is itself a function of x. One often suppresses the argument x by writing just $\frac{df}{dx}$ to denote $f'(x)$. To compute the derivative at a point one differentiates and then evaluates the derivative function at the required point, e.g. $f(x) = \sin(x)$, gives $f'(x) = \cos(x)$, from which $f'(0) = 1$.

Functions of n variables

The definition of partial derivative generalises to functions of n variables

The **partial derivative** of $f(x_1, x_2, \cdots, x_n)$ with respect to x_i, $1 \leq i \leq n$, is

$$\frac{\partial f}{\partial x_i}(x_1, \cdots, x_n) = \lim_{h \to 0} \frac{f(x_1, \cdots, x_i + h, \cdots, x_n) - f(x_1, \cdots, x_i, \cdots, x_n)}{h}$$

The most common cases in this course will be functions of two and three variables: $f(x, y)$ and $f(x, y, z)$.

4.3 Gradient

The gradient plays a fundamental role in the differential calculus of functions of several variables. This week and next week we will discuss different uses and interpretations of the gradient. It will appear in many subsequent courses.

Let f be a functions of n variables. The **gradient vector**, denoted by ∇f, is the vector-valued function formed from the n partial derivatives

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_i}, \cdots, \frac{\partial f}{\partial x_n} \right)$$

The gradient is a vector quantity, it has n components, and it is a function of coordinates $(x_1, \cdots, x_n)$. We are particularly interested in functions of two and three variable, for which we can write explicitly

$$\begin{aligned} \nabla f(x, y) &= \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y) \right) \\ &= \frac{\partial f}{\partial x}(x, y)\mathbf{i} + \frac{\partial f}{\partial y}(x, y)\mathbf{j} \end{aligned}$$

and

$$\nabla f(x,y,z) = \left(\frac{\partial f}{\partial x}(x,y,z), \frac{\partial f}{\partial y}(x,y,z), \frac{\partial f}{\partial z}(x,y,z)\right)$$
$$= \frac{\partial f}{\partial x}(x,y,z)\mathbf{i} + \frac{\partial f}{\partial y}(x,y,z)\mathbf{j} + \frac{\partial f}{\partial z}(x,y,z)\mathbf{k}$$

4.4 Chain Rule

Almost all of the differentiation rules you know for functions of one variable go over to rules for partial derivative exactly as you expect. In fact, one usually does not even state them as rules for partial differentiation. For example, given $f(x,y)$ and $g(x,y)$ a partial derivative of their product is

$$\frac{\partial fg}{\partial x} = g\frac{\partial f}{\partial x} + f\frac{\partial g}{\partial x}$$

but this is obvious (or soon will be to you) since taking the x partial derivative means treating y as a constant and so the product rule really is just the product rule from ordinary differentiation.

The Chain Rule is different. It is also pervasive in the treatment of functions of several variables. Recall the Chain Rule for functions of one variable. It tells us how to differentiate functions of functions. Let $g(t) = f(h(t))$ then we have

$$g'(t) = f'(h(t))\, h'(t)$$

In this chapter we consider the basic case of the mulivariable Chain Rule where we have a real valued function of several variables, and each of these variables is a function of a single other variable. In later chapters this will be generalised.

For simplicity, consider a function of just two variables f depending on (x,y). Let both x and y be functions of a third variable t. We name these functions with the variable names and write $x(t)$ and $y(t)$. Using composition we can construct a function $g : \mathbb{R} \to \mathbb{R}$, $g(t) = f\big(x(t), y(t)\big)$.

The chain Rule for this case is

$$\frac{dg}{dt}(t) = \frac{\partial f}{\partial x}\left(x(t), y(t)\right)\frac{dx}{dt}(t) + \frac{\partial f}{\partial y}\left(x(t), y(t)\right)\frac{dy}{dt}(t)$$

which is often written simply as

$$\frac{dg}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

In the general case of $f : \mathbb{R}^n \to \mathbb{R}$ where $f(x_1, \cdots, x_n)$ and where each x_i is itself a function of a single variable t, we have

The **Chain Rule**. Let $g(t) = f(x_1(t), \cdots, x_n(t))$, then

$$\frac{dg}{dt}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x_1(t), \cdots, x_n(t)) \frac{dx_i}{dt}(t) \qquad (7)$$

or

$$\frac{dg}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \qquad (8)$$

In words, the derivative is computed are follows: starting at the left, compute the partial derivative of f with respect to its first argument and multiply by the ordinary derivative of that argument function. Now do the same for the next argument of f and add that on. Continue until you get to the last component of argument of f.

Warning: *Most aspects of partial differentiation are straightforward, almost trivial extensions of what you know from functions of one variable. However, the Chain Rule has a tendency to cause trouble. The reason is compact notation that is used, as in Eq.* (8). *It is assumed you understand where the functions are being evaluated, so be sure you do understand. If necessary write out the arguments in full as in Eq.* (7).

4.5 Chain Rule (again)

Let us now re-approach the Chain Rule using vector notation. Given a function of n variables $f : \mathbb{R}^n \to \mathbb{R}$ and a vector function $\mathbf{r} : \mathbb{R} \to \mathbb{R}^n$, into same n-dimensions, we can compose these to obtain

$$g = f \circ \mathbf{r} : \mathbb{R} \to \mathbb{R}, \quad \text{with} \quad g(t) = f(\mathbf{r}(t))$$

This is the same composition considered in previous section. We have just notationally replaced all of the component functions $x_i(t)$ with a single vector function $\mathbf{r}(t)$ and used the $\circ$ notation for function composition.

Now re-write the Chain Rule using the gradient vector and the fact that $\dfrac{dx_i}{dt}$ are components of $\dfrac{d\mathbf{r}}{dt}$. Then

$$\frac{dg}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} = \nabla f \cdot \frac{d\mathbf{r}}{dt} = \nabla f \cdot \mathbf{r}'$$

The Chain Rule reduces to the dot product between the gradient vector and the derivative vector $\mathbf{r}'$.

The **Chain Rule** (again). Let $g(t) = f(\mathbf{r}(t))$, then

$$\frac{dg}{dt}(t) = \nabla f(\mathbf{r}(t)) \cdot \frac{d\mathbf{r}}{dt}(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

You should see clearly that this is simply the previous Chain Rule written using different notation.

4.6 Directional Derivative

For simplicity we again initially restrict to the case $f : \mathbb{R}^2 \to \mathbb{R}$. Using vector notation, we can write the definitions of partial derivatives as

$$\frac{\partial f}{\partial x}(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{i}) - f(\mathbf{x})}{h}$$
$$\frac{\partial f}{\partial y}(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{j}) - f(\mathbf{x})}{h}$$

where $\mathbf{x} = (x, y)$.

As you may have guessed, there is nothing special about the unit vectors $\mathbf{i}$ and $\mathbf{j}$ and the derivative can be generalised to any direction $\mathbf{u}$, where $\mathbf{u} \in \mathbb{R}^2$ is a unit vector. This is called the **directional derivative** of $f(x, y)$ in the direction $\mathbf{u}$ and is denoted by $D_{\mathbf{u}}f$. Specifically,

$$D_{\mathbf{u}}f(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{u}) - f(\mathbf{x})}{h}$$

In the general case we have

The **directional derivative** of $f : \mathbb{R}^n \to \mathbb{R}$ in the direction $\mathbf{u}$ is

$$D_{\mathbf{u}}f(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{u}) - f(\mathbf{x})}{h}$$

where $\mathbf{u} \in \mathbb{R}^n$ is a unit vector.

While one can compute the directional derivative from the definition, it is more common to use the gradient vector as follows. Let

$$g(t) = f(\mathbf{r}(t)), \quad \text{where} \quad \mathbf{r}(t) = \mathbf{x} + t\mathbf{u}$$

with $\mathbf{x}$ and $\mathbf{u}$ fixed with $\mathbf{u}$ a unit vector. These have the same meaning as above: $\mathbf{x}$ will be the point where we evaluate the directional derivative and $\mathbf{u}$ is the direction. Note

$$\mathbf{r}(0) = \mathbf{x} \qquad \mathbf{r}'(t) = \mathbf{u}$$

Now we compute $\frac{dg}{dt}(0)$ two ways. By definition:

$$\begin{aligned}\frac{dg}{dt}(0) &= \lim_{h\to 0}\frac{g(h)-g(0)}{h} = \lim_{h\to 0}\frac{f(\mathbf{r}(h))-f(\mathbf{r}(0))}{h}\\ &= \lim_{h\to 0}\frac{f(\mathbf{x}+h\mathbf{u})-f(\mathbf{x})}{h} = D_{\mathbf{u}}f(\mathbf{x})\end{aligned}$$

By the Chain Rule:

$$\frac{dg}{dt}(0) = \nabla f(\mathbf{r}(0))\cdot\mathbf{r}'(0) = \nabla f(\mathbf{x})\cdot\mathbf{u}$$

Equate the two we obtain

The directional derivative of $f:\mathbb{R}^n\to\mathbb{R}$ in the direction $\mathbf{u}$ can be obtained as the dot product of the gradient vector ∇f and the direction vector $\mathbf{u}$:

$$D_{\mathbf{u}}f(\mathbf{x}) = \nabla f(\mathbf{x})\cdot\mathbf{u}$$

Caution: there is variation in the definition of directional derivative. Some authors do not require $\mathbf{u}$ to be a unit vector, and then there is variation in how the case of non-unit vectors is treated. However, this is not an issue when $\mathbf{u}$ is a unit vector and we will always work with unit vectors when taking directional derivatives.

4.7 Higher-Order Derivatives

Just as for functions of a single variable, it is generally possible to differentiate the derivative to obtain the second and higher derivatives. In the case of functions of several variables, there a potentially many second derivatives. For example, $f(x,y)$ has the following second derivatives:

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\\ \frac{\partial^2 f}{\partial x\partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\\ \frac{\partial^2 f}{\partial y\partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\\ \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)\end{aligned}$$

You can seen that there are many possibilities for high derivatives of functions of several variables. One thing that you will learn is that the order of differentiation does not matter for **mixed partial derivatives**, e.g.

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

in the case where $\dfrac{\partial^2 f}{\partial x \partial y}$ and $\dfrac{\partial^2 f}{\partial y \partial x}$ are themselves continuous functions. This will normally be the case for functions we consider in this course, but be careful about always assuming it to be true.

Additional Material

Quadric surfaces

A **quadric surface** is the set of points in $\mathbb{R}^3$ that satisfy a second-degree equation three variables x, y, z. The most general form of such an equation is:

$$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0$$

for constants $A, \cdots, J$. In most cases (non-degenerate cases), by translation and rotation of coordinates it is possible to bring the equation into standard form of

$$Ax^2 + By^2 + Cz^2 + J = 0 \qquad \text{or} \qquad Ax^2 + By^2 + Iz = 0$$

Quadric surfaces are the generalisation to three dimensions of conic sections in two dimensions.

You should see that points satisfying an equation in three variables, e.g. $f(x, y, z) = 0$, is no different than the zero level set, or isosurface, of a function of three variables $f(x, y, z)$.

We will potentially be interested in the following surfaces and will use them as examples throughout the remainder of the module.

Ellipsoid:

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

Horizontal and vertical cuts are ellipses. For $a = b \neq c$ this is a **spheroid**. For $a = b = c$ this is a **sphere**.

Note that here, and below, we follow common practice and write the equation with the constant, or lower-order terms, on the right hand side of the equal sign. In this form it is evident that the ellipsoid is the $k = 1$ isosurface of the second-degree polynomial

$$f(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$$

Elliptic Paraboloid:

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

Horizontal cuts are ellipses and vertical cuts are parabolas.

Hyperbolic Paraboloid:

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

Horizontal cuts are hyperbolas and vertical cuts are parabolas.

Hyperboloid of One Sheet:

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

Horizontal cuts are ellipses. Vertical cuts are hyperbolas.

Hyperboloid of Two Sheets:

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

Horizontal cuts are ellipses (if they intersect the surface). Vertical cuts are hyperbolas.

We also will want to consider the degenerate case of circular cylinders.

Circular Cylinder:

$$x^2 + y^2 = a^2$$

In practice we will consider cylinders whose axis is parallel to the x or y coordinate axes.

Making contour plots

For simple functions you should be able to sketch contours and thus produce an approximate contour map. You should understand the relationship between graph of f and its contour map and you should be able to describe a function given a contour map.

In practice one often uses software to generate of contours. Algorithms for generating contours are non-trivial. Think a little about what you might do to numerically generate all *curves* at a given level for a function f. While commonly the contour levels correspond to an equal spacing in k, at times it might be more appropriate to choose a different spacing, e.g. powers of 10, $k = 1, 10, 100, \ldots$. It is also common to plot contours using colour or grey scale values. Here each set $\mathcal{L}_k$ is assigned a specific colour or grey value depending on k.

Another approach (far easier algorithmically) is to generate colour contour plots where one only need consider a grid of values (x_i, y_j) covering the region of interest. For each grid point one computes $f(x_i, y_j)$ and assigns a corresponding colour or grey level. It is not necessary to generate any curves in the plane – your eye will do that for you.

Optional: linear approximation and the derivative

The following is a brief introduction to differentiation for functions of several variables. It is the subject of the second-year module Differentiation. Read it or not as you wish.

Recall from functions of one variable $f : \mathbb{R} \to \mathbb{R}$ that the derivative provides a linear approximation to a function near any point (assuming the derivative f' exists). There are different ways of writing this approximation, for example

$$\begin{aligned} (A) \quad & f(x) \approx f(a) + f'(a)(x - a) \\ (B) \quad & f(x + h) \approx f(x) + f'(x)h \end{aligned}$$

It is essential that you understand the difference in notation for these two ways for writing the same thing. In (A), a is the fixed (but arbitrary) point at which the derivative is evaluated and x is varying. In (B), x is the fixed (but arbitrary) point at which the derivative is evaluated and h is varying. Depending of the context, one expression is more convenient than the other.

Let us focus on the (B) form where x is fixed (but arbitrary) and h is the variable. The essential issue is that while the left hand side is a general function of h, the right hand sides is linear in h. The graphical view is that the function f, (generally not linear), is approximated by the tangent line to the graph at any point where f is differentiable. You should think of it as

$$f(x + h) \approx f(x) + T(h)$$

where T is a linear map from $T : \mathbb{R} \to \mathbb{R}$. *The derivative of a function is a linear map.* The linear map, $T(h) = f'(x)h$ will depend on which x we are considering, but for each x it is a linear map on h.

Now let us generalise this to functions of several variables by letting $x \to \mathbf{x} \in \mathbb{R}^n$ and $h \to \mathbf{h} \in \mathbb{R}^n$.

$$f(\mathbf{x}+\mathbf{h}) \approx f(\mathbf{x}) + T(\mathbf{h})$$

where T will be a linear map $T : \mathbb{R}^n \to \mathbb{R}$. If it exists, this linear approximation to f will be the derivative of f. The derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $\mathbf{x} \in \mathbb{R}^n$ is a linear map $T : \mathbb{R}^n \to \mathbb{R}$. The linear map will depend on the point $\mathbf{x}$, but for each $\mathbf{x}$, T is a linear map.

We need to work out what this linear map is and we need to say briefly what $\approx$ means in this equation. It is a short calculation to deduce

$$f(\mathbf{x}+\mathbf{h}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h}$$

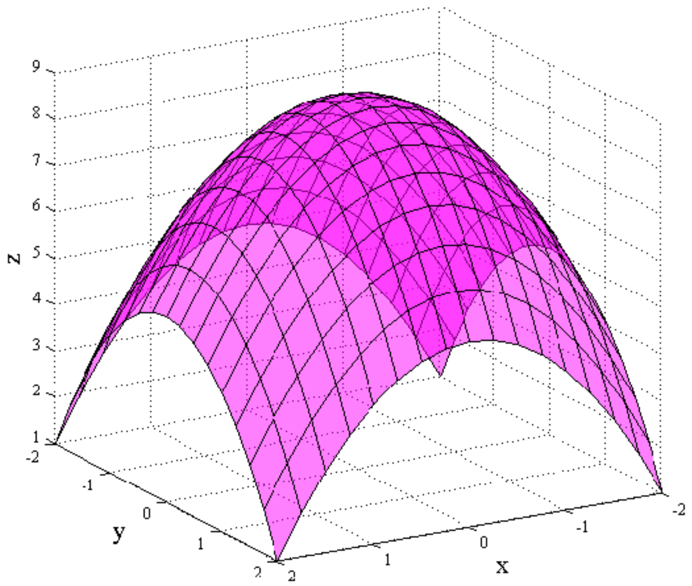
Hence $T(\mathbf{h}) = \nabla f(\mathbf{x}) \cdot \mathbf{h}$. This is a mapping between $\mathbb{R}^n$ and $\mathbb{R}$, and it is linear: $\nabla f(\mathbf{x}) \cdot (\alpha \mathbf{h}) = \alpha \nabla f(\mathbf{x}) \cdot \mathbf{h}$ and $\nabla f(\mathbf{x}) \cdot (\mathbf{h}_1 + \mathbf{h}_2) = \nabla f(\mathbf{x}) \cdot \mathbf{h}_1 + \nabla f(\mathbf{x}) \cdot \mathbf{h}_2$. Dotting $\nabla f(\mathbf{x})$ into $\mathbf{h}$ is the same as multiplying $\mathbf{h}$ as a column vector by a $1 \times n$ matrix. You should readily understand these things from Linear Algebra.

The derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $\mathbf{x}$ is a linear map $T : \mathbb{R}^n \to \mathbb{R}$ given by the gradient vector $\nabla f(\mathbf{x})$.

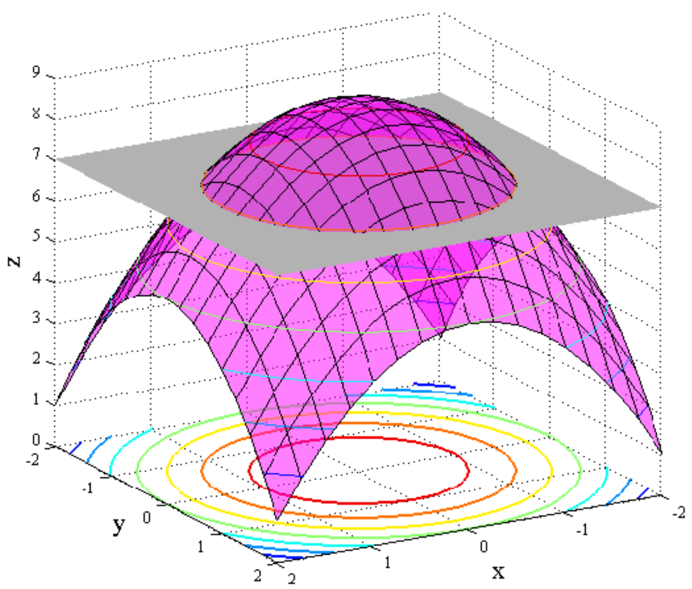
Finally, we give a meaning to $\approx$ in the above expressions. $\approx$ represents the following equality

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + \epsilon(\mathbf{h})\|\mathbf{h}\|$$

where ϵ is a function such that $\epsilon(\mathbf{h}) \to 0$ as $\|\mathbf{h}\| \to 0$. Think of $\epsilon(\mathbf{h})\|\mathbf{h}\|$ as the error in the linear approximation. Since $\epsilon(\mathbf{h}) \to 0$ as $\|\mathbf{h}\| \to 0$, the error $\epsilon(\mathbf{h})\|\mathbf{h}\| \to 0$ faster than $\|\mathbf{h}\|$. You will later learn that this is the definition of differentiability for functions of several variables. You should understand this key idea: *a function is differentiable at a point if it can be approximated by a linear map to within an error that goes to zero faster than $\|\mathbf{h}\|$.*



9
8
7
6
5
4
3
2
1
z
-2
-1
0
1
2
y
2
1
0
-1
-2
x



9
8
7
6
5
4
3
2
1
0
z
-2
-1
0
1
2
y
2
1
0
-1
-2
x