MULTIVARIBLE CALCULUS

Eric Carlen

Departments of Mathematics,
Rutgers University

and

Michael Loss

School of Mathematics,
Georgia Institute of Technology

February 20, 2010

Contents

Chapter 4

DIFFERENTIABLE FUNCTIONS OF A VECTOR VARIABLE

4.1 Vertical slices and directional derivatives

4.1.1 Directional derivatives and partial derivatives

We now pick up with the "slicing" idea introduced in Chapter 3, and try to take derivatives along slices of a function $f : \mathbb{R}^n \to \mathbb{R}$ in order to understand the nature of the graph of this function. Of course, we can only actually plot the graph for $n \leq 2$, so we will be especially interested in the case $n = 2$ at the beginning.

Definition 1 (directional derivatives)**.** *Given a function $f(\mathbf{x})$ defined in a neighborhood of some point $\mathbf{x}_0$ in $\mathbb{R}^n$, and also a non zero vector $\mathbf{v}$ in $\mathbb{R}^n$, the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{v}$ is defined by*

$$\lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} \ , \tag{4.1.1}$$

provided this limit exists. If the limit does not exist, the directional derivative does not exist.

Given f, $\mathbf{x}_0$ and $\mathbf{v}$, the function

$$g(t) = f(\mathbf{x}_0 + t\mathbf{v}) \tag{4.1.2}$$

represents the "slice" of f over the line $\mathbf{x}_0 + t\mathbf{v}$ in $\mathbb{R}^n$. Then the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{v}$ is just $g'(0)$. This means that directional derivatives can be computed by single familiar variable methods.

Example 1 (Slicing a function along a line)**.** *For example, if* $f(x,y) = \dfrac{xy^2}{1+x^2+y^2}$*,* $\mathbf{x}_0 = (1,1)$ *and* $\mathbf{v} = (1,2)$*, then*

$$g(t) = f(1+t, 1+2t) = \frac{(1+t)(1+2t)^2}{1+(1+t)^2+(1+2t)^2} = \frac{1+5t+8t^2+4t^3}{3+6t+5t^2} \ .$$

The result is a familiar garden variety function of a single variable t*. It is a laborious but straightforward task to now compute that* $g'(0) = 1$*. Please do the calculation; you will then appreciate the better way of computing directional derivatives that we shall soon explain!*

Directional derivatives may exist for some directions, but not others:

Example 2 (Sometimes there are directional derivatives only in special directions)**.** *Let* f *be the function defined by*

$$f(x,y) = \begin{cases} \dfrac{x^2-y^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \ . \tag{4.1.3}$$

Let $\mathbf{x}_0 = (0,0)$ *and let* $\mathbf{v} = (a,b)$ *for some numbers* a *and* b*. The question we now ask is: for which values of* a *and* b *does there exists the directional derivative of* f *at* $\mathbf{x}_0$ *in direction* $\mathbf{v}$*?*

To answer this, let's compute $f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)$*, divide by* h*, and try to take the limit* $h \to 0$*. We find that*

$$f(\mathbf{x}_0) = 0 \qquad \text{and} \qquad f(\mathbf{x}_0 + h\mathbf{v}) = \frac{a^2-b^2}{a^2+b^2} \ .$$

(Since by definition, the direction vector $\mathbf{v} \neq 0$*, we do not divide by zero on the right.) Therefore*

$$\frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \frac{1}{h}\left(\frac{a^2-b^2}{a^2+b^2}\right) \ .$$

As $h \to 0$*, this "blows up", unless* $a = \pm b$*, in which case the the right hand side is zero for every* $h \neq 0$*, and so the limit does exist, and is zero. Therefore, for this "bad" function, the directional derivative exists if and only if the direction vector* $\mathbf{v}$ *is is a non-zero multiple of either* $(1,1)$ *or* $(-1,1)$*.*

There are two "special" direction to consider – the direction along the coordinate axes. These special directional derivatives are called *partial derivatives*:

Definition 2 (Partial derivatives)**.** *Given a function* $f : \mathbb{R}^n \to \mathbb{R}$ *defined in a neighborhood of* $\mathbf{x}_0$*, for each* $1 \le j \le n$*,* the partial derivative of f with respect to x_j at $\mathbf{x}_0$ *is denoted by* $\frac{\partial}{\partial x_j} f(\mathbf{x}_0)$ *and is defined by*

$$\frac{\partial}{\partial x_j} f(\mathbf{x}_0) = \lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{e}_j) - f(\mathbf{x}_0)}{h} \tag{4.1.4}$$

provided that the limit exists.

•*Partial derivatives are special cases of directional derivatives – the case in which the direction vector* $\mathbf{v}$ *is one of the standard basis vectors* $\mathbf{e}_i$*.*

Now, how to compute partial derivatives and directional derivatives: This turns out to be easy! If $g(x)$ is related to $f(x, y)$ through

$$g(x) = f(x, y_0) \ ,$$

then

$$\frac{\partial}{\partial x} f(x_0, y_0) = \lim_{h \to 0} \frac{g(x_0 + h) - g(x_0)}{h} = g'(x_0) \ . \tag{4.1.5}$$

This is great news: We will not need to make explicit use of the definition of partial derivatives very often to compute them. When computing a partial derivative, just treat all of the other variables as constants, and differentiate in the single "active" variable in the usual way. Thus, we can use everything we know about computing derivatives for functions of a single variable when we are computing partial derivatives.

Example 3 (Differentiating in one variable)**.** *Let* $f(x, y) = \sqrt{1 + x^2 + y^2}$ *and* $(x_0, y_0) = (1, 2)$*. Then with* $g(x)$ *defined by* $g(x) = f(x, y_0)$*,* $g(x) = \sqrt{5 + x^2}$*. By a simple computation,*

$$g'(x) = \frac{x}{\sqrt{5 + x^2}}$$

and in particular

$$\frac{\partial}{\partial x} f(1, 2) = \frac{1}{\sqrt{9}} = \frac{1}{3} \ .$$

In single variable calculus, the derivative function $g'(x)$ is the function assigning the "output " $g'(x_0)$ to the "input" x_0. In the same way, we let $\dfrac{\partial}{\partial x}f(x,y)$ denote the function of the two variables x and y that assigns the "output" $\dfrac{\partial}{\partial x}f(x_0,y_0)$ to the "input" (x_0,y_0). The same conventions apply to other other functions and more variables.

Example 4 (Computing partial derivatives)**.** *Let $f(x,y)=\sqrt{1+x^2+y^2}$. Holding y fixed – as a parameter instead of a variable – we differentiate with respect to x as in the single variable calculus, and find*

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

Likewise,

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

Once more, because computing partial derivatives is just a matter of differentiating with respect to one chosen variable, everything we know about differentiating with respect to one variable can be applied – in particular the chain rule and the product rule.

Example 5 (Using the single variable chain rule)**.** *The function $f(x,y)=\sqrt{1+x^2+y^2}$ that we considered in Example 4 can be written as a composition $f(x,y)=g(h(x,y))$ where*

$$g(z)=\sqrt{z+1}\qquad\text{and}\qquad h(x,y)=x^2+y^2\ .$$

Since

$$g'(z)=\frac{1}{2\sqrt{1+z}}\qquad\text{and}\qquad\frac{\partial}{\partial x}h(x,y)=2x\ ,$$

we have

$$\frac{\partial}{\partial x}f(x,y)=\frac{\partial}{\partial x}g(h(x,y))=g'(h(x,y))\frac{\partial}{\partial x}h(x,y)=\frac{1}{2\sqrt{1+h(x,y)}}2x=\frac{x}{\sqrt{x^2+y^2}}\ ,$$

as before.

What we saw in Example 5 is a generally useful fact about partial derivatives: If g is a differentiable function of a single variable, and h is a function of two (or more) variables with $\partial h/\partial x$ defined, then

$$\frac{\partial}{\partial x}g(h(x,y))=g'(h(x,y))\frac{\partial}{\partial x}h(x,y)\ ,$$

and similarly with y and any other variables. The validity of this identity need not be formulated as a theorem, and does not need a new proof: It is true because of the single variable chain rule, and because we are differentiating in the single variable x.

• *In short, as far as computing partial derivatives goes, there is nothing much new: Just pay attention to one variable at a time, and differentiate with respect to it as usual.*

Now let us see how to compute directional derivatives in terms of partial derivatives. The key is another chain rule, which is a genuinely multivariable chain rule.

4.1.2 The gradient and a chain rule for functions of a vector variable

We have already seen in our study of continuity that knowing how "slices" along lines of a function f behave does not tells us the whole story about the behavior of the f: We need to look at the behavior along more general families of curves. It is the same with differentiability.

In this subsection we prove a chain rule for functions of a vector variable that is useful for understanding the behavior of f over slices along differentiable curves. That is, let $\mathbf{x}(t)$ be a differentiable vector valued function in $\mathbb{R}^n$. Let f be a function from $\mathbb{R}^n$ to $\mathbb{R}$. Consider the composite function $g(t)$ defined by

$$g(t) = f(\mathbf{x}(t)) \ .$$

Here we ask the question:

• *Under what conditions on f is g differentiable, and can we compute $g'(t)$ in terms of $\mathbf{x}'(t)$ and the partial derivatives of f?*

Before answering this question, we make a useful definition. We organize the partial derivatives of f into a vector. This definition will figure in most of what we do in the rest of this chapter.

Definition 3 (Gradient)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ have each of its partial derivatives well defined at $\mathbf{x}_0$. Then the gradient of f at $\mathbf{x}_0$ is the vector $\nabla f(\mathbf{x}_0) \in \mathbb{R}^n$ given by*

$$\nabla f(\mathbf{x}_0) = \left(\frac{\partial}{\partial x_1} f(\mathbf{x}_0) \ , \dots \ , \frac{\partial}{\partial x_n} f(\mathbf{x}_0)\right) \ .$$

Since you know how to compute partial derivatives, you know how to compute gradients: It is just a matter of organizing the partial derivatives, once you have computed them, into a vector. Here is an example:

Example 6 (Computing a gradient)**.** *With* $f(x) = \dfrac{xy^2}{1+x^2+y^2}$, *we compute that*

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

and

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

Therefore,

$$\nabla f(\mathbf{x}) = \frac{1}{(1+x^2+y^2)^2}\big(y^2(1+y^2-x^2)\ ,\ 2xy(1+x^2)\ \big)\ .$$

We are now ready to state our multivariable chain rule:

Theorem 1 (The chain rule for functions from $\mathbb{R}^n$ to $\mathbb{R}$)**.** *Let f be any function defined in an open set U of $\mathbb{R}^n$ Suppose that each of the partial derivatives of f is defined and continuous at every point of U. Let $\mathbf{x}(t)$ be a differentiable function from $\mathbb{R}$ to $\mathbb{R}^n$. Then, for all values of t so that $\mathbf{x}(t)$ lies in U,*

$$\lim_{h\to 0} \frac{f(\mathbf{x}(t+h)) - f(\mathbf{x}(t))}{h} = \mathbf{x}'(t) \cdot \nabla f(\mathbf{x}(t))\ . \tag{4.1.6}$$

This is an important theorem, and before proving it, we make a some remarks. First, the chain rule in Theorem 1 applies to the composition of functions from $\mathbb{R}$ to $\mathbb{R}^n$ and then from $\mathbb{R}^n$ to $\mathbb{R}$. Both function involved in this composition are multi-variable functions on one end or the other. The chain rule in Example 5 applies to the composition of functions from $\mathbb{R}^n$ to $\mathbb{R}$ and then from $\mathbb{R}$ to $\mathbb{R}$. In the latter case, as we have seen, we we are really only using the single variable chain rule. But the chain rule of Theorem 1 describes rates of change when all of the variables $x_1, \ldots, x_n$ are changing at once. In proving it, we will make essential use of the assumption of continuity of the partial derivatives. Without this assumption, the theorem would not be true.

Second, Theorem 1 has a simple corollary that gives us a formula for computing directional derivatives in terms of partial derivatives.

Corollary 1 (Directional derivatives and gradients)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be any function defined in an open set* U *of* $\mathbb{R}^2$ *Suppose that each partial derivative of* f *is defined and continuous at every point of* U*. Then for any* $\mathbf{x}_0$ *in* U*, and any direction vector* $\mathbf{v}$ *in* $\mathbb{R}^n$*,*

$$\lim_{h\to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \mathbf{v} \cdot \nabla f(\mathbf{x}_0) \ . \tag{4.1.7}$$

Proof: Simply consider the case in which $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$, and apply Theorem 1. □

If you worked through all the calculations in Example 1, you know that computing directional derivatives "straight from the definition" as we did there can be pretty laborious. The good new is that Corollary 1 provides a *much better* way!

Example 7 (Directional derivatives via gradients)**.** *With* $f(x) = \dfrac{xy^2}{1 + x^2 + y^2}$, $\mathbf{x}_0 = (1, 1)$ *and* $\mathbf{v} = (1, 2)$ *as in Example 1. In that example, we computed (the hard way) that the corresponding directional derives is* 1*.*

But now, from Example 6, we have that

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

and hence, substituting $x = 1$ *and* $y = 1$*, we have*

$$\nabla f(\mathbf{x}_0) = \frac{1}{9}(1, 4)\ .$$

Therefore,

$$\mathbf{v} \cdot \nabla f(\mathbf{x}_0) = 1\ .$$

Hence, we get the same value, namely 1*, for the directional derivative.*

The reason that we did not already introduce a special notation for the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{v}$ is that Corollary 1 provides one, namely $\mathbf{v} \cdot \nabla f(\mathbf{x}_0)$. We couldn't use it in the last subsection because we hadn't yet defined gradients, but now that we have, this will be our standard notation for directional derivatives, at least when we are dealing with "nice" functions whose partial derivatives are continuous.

Corollary 1 provides an efficient means for computing directional derivatives, because it is easy to compute partial derivatives – even if there are many variables, just one is varying at a time. Also, one you have computed the gradient, you are done with that once and for all. You can take the dot product with lots of different direction vectors and compute lots of directional derivatives without doing any more serious work. In the approach used in Example 1, you would have to start from scratch each time you considered a new direction vector.

We are finally ready for the proof of Theorem 1. The key to the proof, in which we finally explain the importance of continuity for the partial derivatives if the *Mean Value Theorem* from single variable calculus:

The Mean Value Theorem says that if $g(s)$ has a *continuous* first derivative $g'(s)$, then for any numbers a and b, with $a < b$, there is a value of c in between; i.e., with $a < c < b$

$$\frac{g(b) - g(a)}{b - a} = g'(c) \ . \tag{4.1.8}$$

The principle expressed here is the one by which the police know that if you drove 100 miles in one hour, then at some point on your trip, you were driving at 100 miles per hour.

Proof of Theorem 1: We give the proof for $n = 2$ to keep the notation simple. Once this case is understood, the general case will be clear.

Fix some t, and some $h > 0$. To simplify the notation, define the numbers x_0, y_0, x_1 and y_1 by

$$(x_0\ ,\ y_0) = \mathbf{x}(t) \qquad \text{and} \qquad (x_1\ ,\ y_1) = \mathbf{x}(t+h) \ .$$

To enable the analysis of $f(\mathbf{x}(t+h)) - f(\mathbf{x}(t))$ by single variable methods, note that

$$\begin{aligned} f(\mathbf{x}(t+h)) - f(\mathbf{x}(t)) &= f(x_1, y_1) - f(x_0, y_0) \\ &= \big[f(x_1, y_1) - f(x_0, y_1)\big] + \big[f(x_0, y_1) - f(x_0, y_0)\big] \ . \end{aligned}$$

Notice that in going from the first line to the second, we have subtracted and added back in the quantity $f(x_0, y_1)$, and grouped the terms in brackets.

In the first group, only the x variable is varying, and in the second group, only the y variable is varying. Thus, we can use single variable methods on these groups.

To do this for the first group, define the function $g(s)$ by

$$g(s) = f(x_0 + s(x_1 - x_0), y_1) \ .$$

Notice that

$$g(1) - g(0) = f(x_1, y_1) - f(x_0, y_1) \ .$$

Then, if g is continuously differentiable, the Mean Value Theorem tells us that

$$f(x_1, y_1) - f(x_0, y_1) = \frac{g(1) - g(0)}{1 - 0} = g'(c)$$

for some c between 0 and 1.

But by the definition of $g(s)$, we can compute $g'(s)$ by taking a partial derivative of f, since as s varies, only the x component of the input to f is varied. Thus,

$$g'(s) = \frac{\partial}{\partial x} f(x_0 + s(x_1 - x_0), y_1)(x_1 - x_0) \ .$$

Therefore, for some c between 0 and 1,

$$\big[f(x_1, y_1) - f(x_0, y_1)\big] = \left[\frac{\partial}{\partial x} f(x_0 + c(x_1 - x_0), y_1)\right](x_1 - x_0) \ .$$

In the exact same way, we deduce that for some $\tilde{c}$ between 0 and 1,

$$\big[f(x_0, y_1) - f(x_0, y_0)\big] = \left[\frac{\partial}{\partial y} f(x_0, y_0 + \tilde{c}(y_1 - y_0))\right](y_1 - y_0) \ .$$

Therefore,

$$\begin{aligned}
\frac{f(\mathbf{x}(t+h)) - f(\mathbf{x}(t))}{h} &= \left[\frac{\partial}{\partial x} f(x_0 + c(x_1 - x_0), y_1)\right] \frac{x_1 - x_0}{h} \\
&+ \left[\frac{\partial}{\partial y} f(x_0, y_0 + \tilde{c}(y_1 - y_0))\right] \frac{y_1 - y_0}{h} \ .
\end{aligned}$$

Up to now, h has been fixed. But having derived this identity, it is now easy to analyze the limit $h \to 0$.

First, as $h \to 0$, $x_1 \to x_0$ and $y_1 \to y_0$. Therefore,

$$\lim_{h \to 0} \frac{\partial}{\partial x} f(x_0 + c(x_1 - x_0), y_1) = \frac{\partial}{\partial x} f(x_0, y_0) = \frac{\partial}{\partial x} f(\mathbf{x}(t)) \ ,$$

and

$$\lim_{h \to 0} \frac{\partial}{\partial y} f(x_0, y_0 + \tilde{c}(y_1 - y_0) = \frac{\partial}{\partial y} f(x_0, y_0) = \frac{\partial}{\partial y} f(\mathbf{x}(t)) \ .$$

Also, since $\mathbf{x}(t)$ is differentiable

$$\lim_{h \to 0} \frac{x_1 - x_0}{h} = \lim_{h \to 0} \frac{x(t+h) - x(t)}{h} = x'(t)$$

and

$$\lim_{h \to 0} \frac{y_1 - y_0}{h} = \lim_{h \to 0} \frac{y(t+h) - y(t)}{h} = y'(t)$$

Since the limit of a product is the product of the limits,

$$\begin{aligned}
\lim_{h \to 0} \frac{f(\mathbf{x}(t+h)) - f(\mathbf{x}(t))}{h} &= \left[\frac{\partial}{\partial x} f(\mathbf{x}(t))\right] x'(t) + \left[\frac{\partial}{\partial y} f(\mathbf{x}(t))\right] y'(t) \\
&= \nabla f(\mathbf{x}(t)) \cdot \mathbf{x}'(t) \ .
\end{aligned}$$

This is what we had to show. $\square$

4.1.3 The geometric meaning of the gradient

The gradient of a function is a vector. As such, it has a *magnitude*, and a *direction.* To understand the gradient in geometric terms, let's try to understand what the magnitude and direction are telling us.

The key to this is the formula

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos(\theta) \tag{4.1.9}$$

which says that the dot product of two vectors in $\mathbb{R}^n$ is the product of their magnitudes times the cosine of the angle between their directions.

Now pick any point $\mathbf{x}_0$ and any unit vector $\mathbf{u}$ in $\mathbb{R}^n$. Suppose $f : \mathbb{R}^n \to R$ has continuous partial derivatives at $\mathbf{x}_0$, and consider the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{u}$. By Theorem 1, this is

$$\mathbf{u} \cdot \nabla f(\mathbf{x}_0) \ .$$

By 4.1.9 and the fact that $\mathbf{u}$ is a unit vector (i.e., a pure direction vector),

$$\mathbf{u} \cdot \nabla f(\mathbf{x}_0) = \|\nabla f(\mathbf{x}_0)\| \cos(\theta)$$

where θ is the angle between $\nabla f(\mathbf{x}_0)$ and $\mathbf{u}$. (This is defined as long a $\nabla f(\mathbf{x}_0) \neq 0$, in which case the right hand side is zero.)

As $\mathbf{u}$ varies of the unit circle, $\cos(\theta)$ varies between -1 and 1, and hence

$$-\|\nabla f(\mathbf{x}_0)\| \leq \mathbf{u} \cdot \nabla f(\mathbf{x}_0) \leq \|\nabla f(\mathbf{x}_0)\|$$

Recall that by Theorem 1, $\mathbf{u} \cdot \nabla f(\mathbf{x}_0)$ is the slope at $\mathbf{x}_0$ of the slice of the graph $z = f(\mathbf{x})$ that you get when slicing along $\mathbf{x}_0 + t\mathbf{u}$. Hence we can rephrase this as

$$-\|\nabla f(\mathbf{x}_0)\| \ \leq \ \ [\text{slope of a slice at } \mathbf{x}_0] \ \ \leq \ \|\nabla f(\mathbf{x}_0)\|$$

That is,

• *The magnitude of the gradient, $\|\nabla f(\mathbf{x}_0)\|$ tells us the minimum and maximum values of the slopes of all slices of $z = f(\mathbf{x})$ through $\mathbf{x}_0$.*

The slope has the maximal value, $\|\nabla f(\mathbf{x}_0)\|$, exactly when $\theta = 0$; i.e., when $\mathbf{u}$ and $\nabla f(\mathbf{x}_0)$ point in the same direction. In other words:

• *The gradient of f at $\mathbf{x}_0$ points in the direction of steepest increase of f at $\mathbf{x}_0$*

For the same reasons, we get the steepest negative slope by taking $\mathbf{u}$ to point in the direction of $-\nabla f(\mathbf{x}_0)$.

Example 8 (Which way the water runs). *Let* $f(x) = \dfrac{xy^2}{1+x^2+y^2}$, $\mathbf{x}_0 = (1,1)$ *and let* $\mathbf{x}_0 = (0,1)$. *If* $z = f(\mathbf{x})$ *denotes the altitude at* $\mathbf{x}$, *and you stood at* $\mathbf{x}_0$, *and spilled a glass of water, which way would the water run?*

For purposes of this question, let's say that the direction of the positive x *axis is due East, and the direction of the positive* y *axis is due North.*

But now, from Example 6, we have that

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

and hence, substituting $x = 0$ *and* $y = 1$, *we have*

$$\nabla f(\mathbf{x}_0) = \frac{1}{4}(2,0)\ .$$

Thus, the gradient points due East. This is the "straight uphill" direction. The water will run in the "straight downhill" direction, which is opposite. That is the water will run due West.

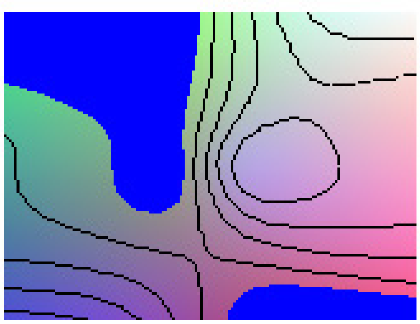
4.2 Horizontal slices and contour curves

In Section 1 of this chapter, we have considered *vertical* slices of the graph of $z = f(x,y)$. We can gain a new perspective by considering *horizontal* slices.

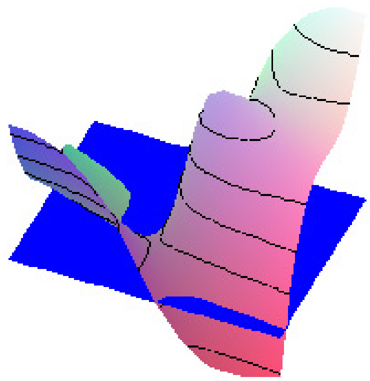
Consider once again the "mountain landscape" graphed in the first section:

$$f(x,y) = \frac{3(1+x)^2 + xy^3 + y^2}{1+x^2+y^2} \tag{4.2.1}$$

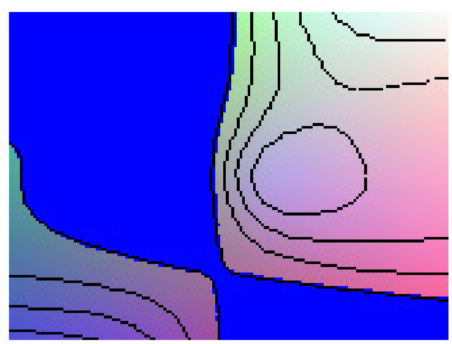
Suppose that a dam is built, and this landscape is flooded, up to an altitude 0.5 in the vertical distance units. This produces a lake that is shown below, in a top view; i.e., an aerial or satellite image:



The other lines of the land are the lines at other constant altitudes, specifically $x = 1.5$, $z = 2.5$, $z = 3.5$ and so on. On a topographic map, these curves are called *contour curves*. Here is a sort of side view showing the lake as a horizontal "slice" through the graph $z = f(x, y)$ at height $z = 1.5$:



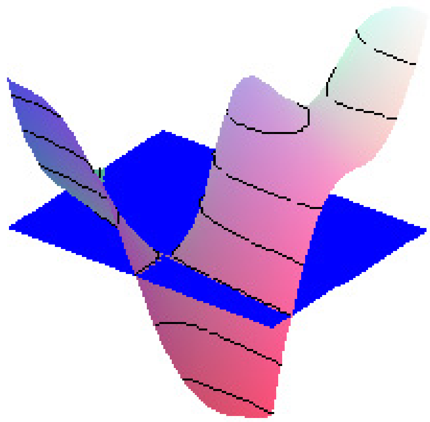
If the water level is raised further, say to the altitude $z = 1.5$, everything will be flooded up to the next contour curve:



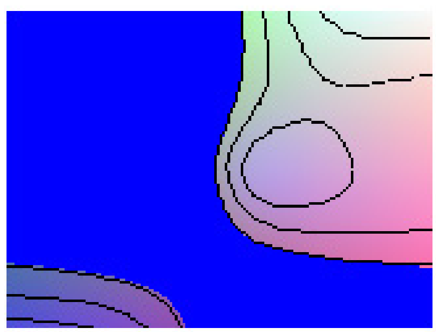
Comparing with the first picture, you clearly see that everything has been flooded up to the $z = 1.5$ contour curve. The isthmus joining the two tall hills is now submerged, and the two regions of the lake in the first graph have merged.

If you walked along the lake shore, your path would trace out the contour curve at $z = 1.5$ in the first picture.

Here is a side view showing the lake at this level. It shows it as a horizontal "slice" through the graph $z = f(x, y)$ at height $z = 1.5$:



If the water level is raised further, to the height $z = 2.5$, the shore line moves up to the next contour line. Now a walk along the shoreline would trace out the path along the $x = 2.5$ contour line in the first picture. Here is the top view showing the lake at this stage:



The contour curves, which are the results of horizontal slices of the graph of $z = f(x, y)$, tell us a lot about the function $f(x, y)$. This section is an introduction to what they tell us.

First, let us recall a definition from Chapter 3: Given a function $f : \mathbb{R}^n \to \mathbb{R}$, the set of points $\mathbf{x} \in \mathbb{R}^n$ satisfying satisfying

$$f(\mathbf{x}) = c$$

is the *level set* of f at c In other words, the level set of f at c is the solution set of the equation $f(\mathbf{x}) = c$.

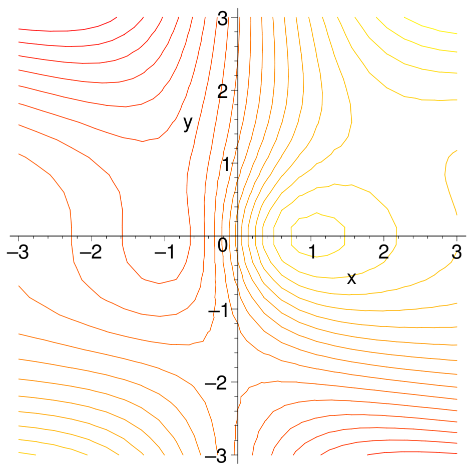
Let us consider the case $n = 2$ more closely. If we think of $f(x, y)$ as representing the altitude at the point with coordinates (x, y), then the level set of f at height c is the set of all points at which the altitude is c. The level set at height c would be the "shore line" curve if the landscape were flooded up to an altitude c.

Now, here is a very important point, whose validity you can more or less see from the pictures we have displayed:

• *Under normal circumstances, the level set of f at c will be a curve in the plane, possible with several disconnected components.*

It is for this reason that it is reasonable to refer to level sets as contour curves. We can plot a number of the level sets on a common graph. A *contour plot* of a function $f(x, y)$ is graph in which level curves of f are plotted at several different "altitudes" $c_1, c_2, c_3, \ldots$. You have probably seen these on maps for hiking.

Here is a contour plot for the function "mountain landscape" function $f(x, y)$ in (4.2.1):



4.2.1 Implicit and explicit descriptions of planar curves

How could one go about actually drawing the contour curves starting from a formula like (4.2.1)? That is not so easy in general. You can see a hint of this in the convoluted form of the contour curves plotted here. The difficulty lies here:

• *The description of contour curves given by the defining equation $f(x, y) = c$ is, alas, just an implicit description. However, plotting a curve is a simple matter only when one has an explicit description.*

To really appreciate this point, one has to understand the distinction between an implicit and and explicit description of a curve. The unit circle is a great example with which to start.

Let $f(x, y) = x^2 + y^2$ and let $c = 1$. Then the level set of f at height c is the set of points (x, y) satisfying

$$x^2 + y^2 = 1 \ . \tag{4.2.2}$$

This set, of course, is the unit circle. If we drew a contour plot of f showing the level curves at several altitudes "altitudes" $c_1, c_2, c_3, \ldots,$ you would see, several concentric circles.

The equation (4.2.2) is the *implicit* equation for the unit circle. To get an *explicit description*, just solve the equation (4.2.2) for y as a function of x. The result is

$$y(x) = \pm\sqrt{1 - x^2} \ . \tag{4.2.3}$$

There are two "branches", corresponding to the two signs.

Given y as an explicit function of x, it is easy to plot the curve. For instance, taking the "upper branch" $y = \sqrt{1 - x^2}$ from (4.2.3), and plugging in a sequence of values for x, we get the table

$$\begin{array}{rcl} x = 0 & \quad y & = & 1 \\ x = 1/8 & \quad y & = & \sqrt{63}/8 \\ x = 1/4 & \quad y & = & \sqrt{15}/4 \\ x = 3/8 & \quad y & = & \sqrt{55}/8 \\ x = 1/2 & \quad y & = & \sqrt{3}/2 \\ x = 5/8 & \quad y & = & \sqrt{39}/8 \end{array}$$

and so forth. Connecting the dots, we get a good picture of part of the curve. (To get the rest, just extend the table so that the x values range all the way from -1 to 1, and do the same for the "lower branch" $y = -\sqrt{1 - x^2}$).

As long as one variable is given as an *explicit* function of another, or several explicit functions in case there are several branches, one can produce the graph by making a table and connecting the dots. This is what is nice about "explicitly defined curves".

There is a more general kind of explicit description – *parametric*. As t varies between 0 and 2π,

$$\big(\cos(t), \sin(t)\big) \tag{4.2.4}$$

traces out the unit circle in the counterclockwise direction. This is another sort of explicit description since if you plug in any value of t, you get a point (x, y) on the unit circle, and as you vary t, you "sweep out" all such points. Again, there are no equations to solve, just computations to do.

If we take $t = x$, we can rewrite (4.2.3) as

$$x(t) = t \qquad \text{and} \qquad y = \pm\sqrt{1-t^2} \qquad \text{for} \quad -1 \le t \le 1 \ , \tag{4.2.5}$$

so (4.2.3) is just a particular parametric representation of the circle in which we use x itself as the parameter.

Definition 4 (Implicit and explicit descriptions of curves)**.** *An equation of the form $f(x,y) = c$ provides an implicit description of a curve. A parameterization $\mathbf{x}(t)$, possibly with $t = x$ and $y(x)$ given as an explicit function of x, provides an explicit description of a curve.*

Once one has an explicit description, it is easy to generate a plot, just by plugging in values for the parameter, plotting the resulting points, and "connecting the dots". Passing from an implicit description to an explicit description involves *solving the equation* $f(x,y) = c$ to find an explicit description. Generally, that is easier said than done.

Example 9 (From implicit to explicit by means of algebra)**.** *Consider the function*

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

The level curve at $c = 1$ for this function is given implicitly by the equation

$$2x^2 - 2xy + y^2 = 1 \ .$$

This can be rewritten as

$$y^2 - 2xy = 1 - 2x^2 \ .$$

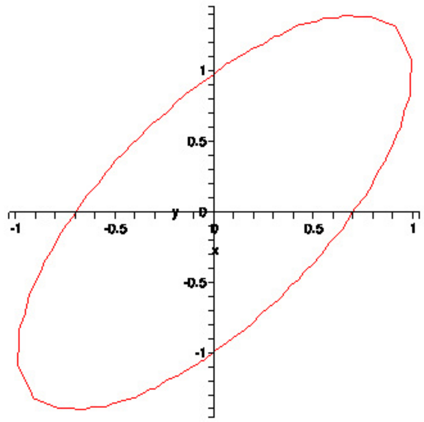
Completing the square in y, we have

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

Therefore, we can solve for y as a function of x, finding

$$y(x) = x \pm \sqrt{1-x^2} \ .$$

If we take x as the parameter, evidently y has a real value only for $-1 \le x \le 1$. It is now easy to plot the contour curve:



In this example, it was not so horrible passing from an implicit description to an explicit description; i.e., a parameterization, since the equation $f(x, y) = 1$ was quadratic in both x and y. We know how to deal with quadratic equations, so in this case, we were able to make the transition from implicit to explicit.

However, in general, this will not be possible to do. In general, we are going to need to extract information on the contour curves *directly* from the implicit description. Fortunately, what we have learned about gradients can help us to do this.

4.2.2 The tangent line to a contour curve

Let $\mathbf{x}(t)$ be a parameterization of a contour curve. As this curve passes through the point $\mathbf{x}(t_0)$, supposing it does so in a reasonably nice way, one can ask about the *direction of motion* as the contour curve passes though the point $\mathbf{x}(t_0)$. We can answer this question using Theorem 1 of the previous section.

Suppose $\mathbf{x}(t) = (x(t), y(t))$ is *some* parameterization of the contour curve of f through c. Then, by definition,

$$g(t) = f(\mathbf{x}(t)) = c$$

for all t.

Since $g(t)$ is constant, $g'(t) = 0$. But by Theorem 1 of the previous section, $g'(t) = \mathbf{x}'(t) \cdot \nabla f(\mathbf{x}(t))$, and hence

$$\mathbf{x}'(t) \cdot \nabla f(\mathbf{x}(t)) = 0$$

for all t.

In other words, the velocity vector $\mathbf{x}'(t)$ is orthogonal to the gradient $\nabla f(\mathbf{x}(t))$ at $\mathbf{x}(t)$ for each t.

• *The direction of motion along the level curve of f passing through $\mathbf{x}_0$ is orthogonal to $\nabla f(\mathbf{x}_0)$. The tangent line to the contour curve through $\mathbf{x}_0$ is given, in parametric form, by*

$$\mathbf{x}_0 + t\nabla^\perp f(\mathbf{x}_0) \ ,$$

where if $\nabla f(\mathbf{x}_0) = (a, b)$, then $\nabla^\perp f(\mathbf{x}_0)$ is defined to be $(-b, a)$.

What we have observed so far leaves one obvious question hanging in the air: What if $\nabla f(\mathbf{x}_0) = 0$, so that the gradient does not define a direction? In this case, the level set given by $f(\mathbf{x}) = c$ with $c = f(\mathbf{x}_0)$ may not be a proper curve and may not have a direction at all. Indeed, suppose $f(x, y) = x^2 + y^2$ and consider $\mathbf{x}_0 = (0, 0)$, Then the level set of f through $\mathbf{x}_0$ consists of the single point $\mathbf{x}_0$. It has no tangent line. And indeed, $\nabla f(\mathbf{x}_0) = 0$.

The question of when $f(x, y) = c$ with $c = f(x_0, y_0)$ defines a proper curve is an important one, to which we shall return. We shall see that this is the case whenever $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$, and the partial derivatives of f are both continuous at $\mathbf{x}_0$. For now, let us take this for granted, and compute some tangent lines to implicitly defined contour curves.

4.2.3 Computing tangent lines for implicitly defined curves

It is actually quite simple to compute the tangent line to an implicitly defined curve through a point *directly from the implicit description*, without first finding a parameterization. Here is how: Recall that the equation of a line in the plane has the form

$$\mathbf{a} \cdot (\mathbf{x} - \mathbf{x}_0) = 0 \ ,$$

where $\mathbf{x}_0$ is any base point on the line, and $\mathbf{a}$ is any non zero vector orthogonal to the line. Now, as we have observed, the gradient vector $\nabla f(\mathbf{x}_0)$ is orthogonal to the contour curve, and hence to the tangent line, and thus we may take $\mathbf{a} = \nabla f(\mathbf{x}_0)$. This gives us the equation of the tangent line!

Example 10 (Computing a tangent line from an implicit description)**.** *Consider the implicit description of the unit circle given by the equation $x^2 + y^2 = 1$, and the point $\mathbf{x}_0 = \dfrac{1}{\sqrt{2}}(1, 1)$, which is on this circle.*

We compute $\nabla f(\mathbf{x}_0) = \sqrt{2}(1,1)$*. Then with* $\mathbf{a} = \nabla f(\mathbf{x}_0)$*,* $\mathbf{a}\cdot(\mathbf{x}-\mathbf{x}_0) = 0$ *is*

$$\sqrt{2}(1,1)\cdot\left((x,y) - \frac{1}{\sqrt{2}}(1,1)\right) = (\sqrt{2}x - 1) + \sqrt{2}y - 1) = 0 \ ,$$

and so the equation of the tangent line is

$$x + y = \sqrt{2} \ .$$

In summary:

• *If there is a differentiable contour curve of* f*,* $\mathbf{x}(t)$*, passing through* $\mathbf{x}_0$ *at* $t = t_0$*,* $\nabla f(\mathbf{x}_0)$ *is orthogonal to* $\mathbf{x}'(t_0)$*, and therefore to the tangent line to the contour curve of* f *through* $\mathbf{x}_0$*.*

The point of this is that while you need to know an explicit parameterization to compute $\mathbf{x}'(t_0)$, you do not need this to compute $\nabla f(\mathbf{x}_0)$ – computing $\nabla f(\mathbf{x}_0)$ gives you a direction orthogonal to the tangent line, which is all we need to write down the equation of the line, as we have observed above.

Example 11 (Tangent line to a level curve)**.** *Let* $f(x,y) = (x^2-y^2)^2 - 2xy$*. We will now compute the tangent line to the level curve of* f *through* $\mathbf{x}_0 = (1,2)$*. We first work out that*

$$\nabla f(x,y) = \big(4x(x^2-y^2) - 2y, -4y(x^2-y^2) - 2x\big) \ .$$

Therefore,

$$\nabla f(1,2) = \big(-16, 22\big) \ .$$

The equation for the tangent line is

$$0 = \big(-16,22\big)\cdot\big(x-1, y-2\big) = -16(x-1) + 22(y-2) \ ,$$

or

$$-16x + 22y = 14 \ .$$

To find the parametric form, we compute

$$\mathbf{v} = \mathbf{x}' = (\nabla f(\mathbf{x}_0))^{\perp} = \big(-16,22\big)^{\perp} = -\big((22,16\big).$$

Hence the tangent line is

$$\mathbf{x}_0 + t\mathbf{v} = \big(1,2\big) - t\big(22,16\big).$$

4.2.4 Is the contour curve always a curve?

All of our conclusions so far in this discussion were predication on the assumption that there was a differentiable level curve of f through $\mathbf{x}_0$. Was this a reasonable assumption? more to the point, can we check the validity of this assumption?

It turns out that if both of the partial derivatives of f are continuous in a neighborhood of $\mathbf{x}_0$, and $\nabla f(\mathbf{x}_0) \neq 0$, then at least nearby $\mathbf{x}_0$, this level set is actually a curve that has a parameterization $\mathbf{x}(t)$, and moreover, $\mathbf{x}(t)$ is itself differentiable.

The theorem that guarantees this is called *Implicit Function Theorem*. We will prove it later on, but what it is saying can be pictured easily enough now. The point is that if the gradient is not zero, then there is a well defined "uphill" direction, given by the gradient, and perpendicular to this are the two "level" directions that are neither uphill nor downhill. The level curve through $\mathbf{x}_0$ is just the contour curve through $\mathbf{x}_0$, and it must proceed away from $\mathbf{x}_0$ in the level directions. If you were standing on the landscape described by $z = f(x, y)$, you would not need to solve anything to walk along the level curve: Just step away from $\mathbf{x}_0$, and walk along without changing your altitude. You would be walking along the the curve whose existence is asserted by the Implicit Function Theorem.

However, if the gradient were zero, in which direction would you walk? In this case there might be no preferred direction, and hence no path. The level set is not always a curve, even when it is not empty.

Example 12 (A level set that is a single point)**.** *Let $f(x, y) = x^2 + y^2$. Then the level set of f at height 0, which is the level set of f passing through $(0, 0)$ is just the single point $(0, 0)$. It is not a curve. At every other point (x, y), f takes on a strictly positive value. Indeed, for this reason, you see that the level set of f at height -1, or any other strictly negative value, is empty. Notice that $\nabla f(0, 0)$ is zero, as the Implicit Function Theorem says it must be.*

Example 13 (A level set that is not a simple curve)**.** *Let $f(x, y) = xy$. Then the level set of f at height 0, which is the level set of f passing through $(0, 0)$, consists of both the x and y axes. Here there are two level directions, and the level set of f passing through $(0, 0)$ is not a curve, but a union of two curves. Again, notice that $\nabla f(0, 0)$ is zero, as the Implicit Function Theorem says it must be.*

4.2.5 Points on a contour curve where the tangent line has a given slope

Consider the curve given implicitly by

$$x^4 + y^4 + 4xy = 0 \ .$$

This is quartic, so it is not so easy to express y as a function of x (though it can be done). A good way to graph this curve is to notice that it is the level set of $f(x,y) = x^4 + y^4 + 4xy$ at $c = 0$. We can use what we have learned to find all points (x,y) on the curve at which the tangent line through then has a given slope s. For example, when $s = 0$, the line is horizontal. When $s = \pm\infty$, it is vertical. When $s = 1$, the tangent line runs at an angle of $\pi/4$ with respect to the x–axis. To graph the contour curve, one can find all such points, and draw a small bit if a line segment through them at the corresponding slope. Connecting these up, one has a sketch of the curve.

Example 14 (Points on a level curve where the tangent has a given direction)**.** *Let $f(x,y) = x^4 + y^4 + 4xy$, and consider the level curve given in implicit form by*

$$x^4 + y^4 + 4xy = 0 \ . \tag{4.2.6}$$

We will now find all points on this level curve at which the tangent is horizontal; i.e., parallel to the x axis.

We compute that

$$\nabla f(x,y) = 4(x^3 + y, y^3 + x) \ .$$

Notice that the gradient is zero exactly when

$$x^3 + y = 0 \qquad \text{and} \qquad y^3 + x = 0 \ .$$

The first equation says that $y = -x^3$. substituting this into the second equation, we eliminate y and get $x^9 = x0$. One solution is $x = 0$. If $x \neq 0$, we can divide by x and get the equation $x^8 = 10$, which two solutions, $x = -1$ and $x = 1$. Then from $y = -x^3$, we see that the three points at which $\nabla f(\mathbf{x}) = 0$ are

$$(0,0) \qquad (-1,1) \qquad \text{and} \qquad (1,-1) \ . \tag{4.2.7}$$

At all other points, the level set specified implicitly by (4.2.6) is a differentiable curve – by the Implicit Function Theorem – though here it may be somewhat more complicated.

Now let's focus on points where $\nabla f(\mathbf{x}) \neq \mathbf{0}$*, and see at which of the the tangent line is horizontal. The tangent to the level curve is horizontal exactly when the perpendicular direction is vertical. The perpendicular direction is the direction of the gradient, so the tangent line is horizontal only in case the first component of the gradient is zero; i.e.,*

$$x^3 + y = 0 \ . \tag{4.2.8}$$

The equation (4.2.8), together with (4.2.6), gives us a nonlinear system of equations for the points we seek. To solve it, notice that while it is not at all easy to eliminate either variable in the original equation (4.2.6), it is obvious from (4.2.8) that

$$y = -x^3 \ . \tag{4.2.9}$$

Substituting this into (4.2.6), we get

$$x^{12} - 3x^4 = 0 \ .$$

This has exactly 3 *solutions:* $x = 0$*,* $x = 3^{1/8}$ *and* $x = 3^{-1/8}$*. Going back to (4.2.9), we can now easily find the corresponding points. They are:*

$$(0,0) \qquad (3^{1/8}, -3^{3/8}) \qquad \text{and} \qquad (-3^{1/8}, 3^{3/8}) \ .$$

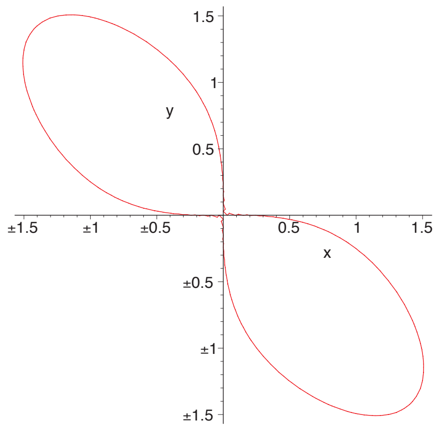
At the first of these, the gradient is zero. That is, $\nabla f(0,0)$ *is zero, and so it has no direction at all. At the other two, we can be sure that the level set is a differentiable curve, and that its tangent vector is vertical.*

Deciding what actually happens at $(0,0)$ *is more subtle. For* x *and* y *very close to zero,* $x^4 + y^4 + 4xy \approx 4xy$*, and so for such* x *and* y*, the level set of* f *at height* 0 *looks pretty much like the level set of* $4xy$ *at height zero. This is easy to find:* $4xy = 0$ *if and only if* $x = 0$ *or* $y = 0$*. Hence this level set consists of two lines, namely the* x *and* y *axes, as in Example 7. At this point, there is a branch of the level set that is vertical, but also a branch that is horizontal, so we cannot properly say that the level curve has a horizontal tangent here.*

However, using the same procedure that we used to find the points of vertical tangency, we find that the level curve of f *at height* 0 *has a horizontal tangent exactly* *at*

$$(3^{1/8}, -3^{3/8}) \qquad \text{and} \qquad (-3^{1/8}, 3^{3/8}) \ .$$

By doing this for a few more slopes, and connecting the points up, you could get a pretty good sketch of the curve. Here is such a sketch:



4.3 Differentiability and tangent planes

4.3.1 Finding the equation of the tangent plane

For a function $g(t)$ of a single variable t, the tangent line to the graph of g at t_0 is the line that "best fits" the graph $y = g(t)$ at t_0. It is the line that passes through the point $(t_0, g(t_0))$ with the same slope at the graph of $y = g(t)$ at t_0. Hence, it is given by the formula

$$y = g(t_0) + g'(t_0)(t - t_0) \ .$$

Since the graph of $y = g(t)$ and the tangent line to this graph at t_0 pass through the *exact same point* with the *exact same slope*, if you "zoom in" really close around a picture of both of them as they pass through $(t_0, g(t_0))$, you will not be able to distinguish between the two of them.

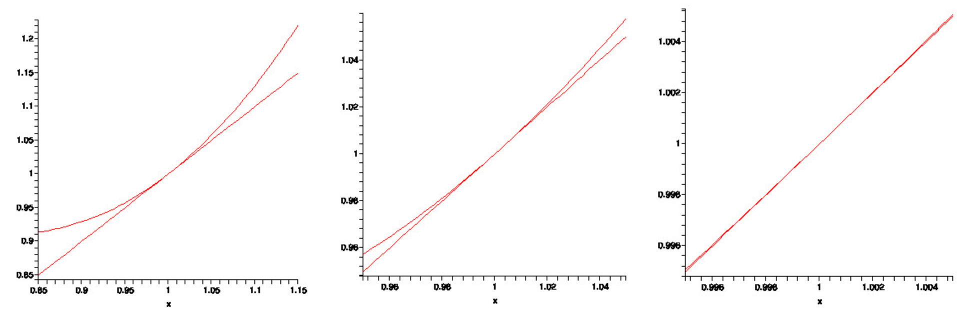
Example 15 (How well tangent lines fit). *Consider $g(t) = t^3 - 2t + 2$, and $t_0 = 1$. The tangent line to the graph of $y = g(x)$ at $t_0 = 1$ is the graph of*

$$y = g(1) + g'(1)(t - 1) \ .$$

Since $g(1) = 1$ and $g'(1) = 1$, the tangent line is the graph of

$$y = 1 + (t - 1) = t \ .$$

Here are three graphs of $y = g(t)$ that are "zoomed in" closer and closer about the point $(1,1)$*:*



In the first graph, the t values range from 0.85 *to* 1.15*. In the second, they range from* 0.95 *to* 1.05*. In the third, they range from* 0.995 *to* 1.005*. In the third graph, the curve and the tangent line are almost indistinguishable.*

Clearly, there is just one line that fits so well: If the line had any other slope, the two graphs would not even look parallel, let alone the same, when we "zoom in". Also, they clearly must both pass through the point $(x_0, g(x_0))$ to have any sort of fit at all. Since the point and the slope determine the line, there is just one line that fits this well.

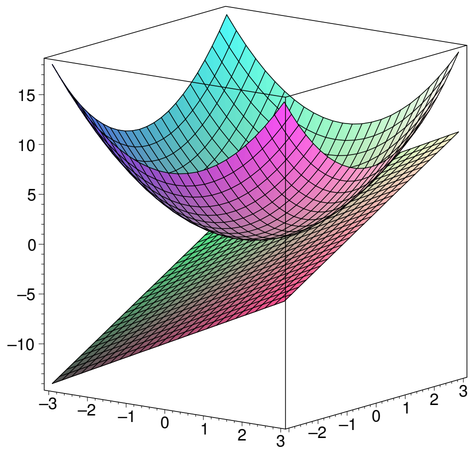
The approximation in which we approximate $g(t)$ by the affine function $g(t_0) + (t - t_0)g'(t_0)$, i.e.,

$$g(t) \approx g(t_0) + (t - t_0)g'(t_0) \ , \tag{4.3.1}$$

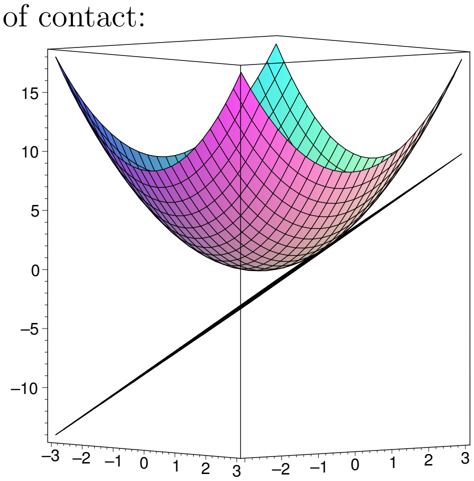
which will be a good approximation for small values of $|t - t_0|$, is called the *tangent line approximation* of g at t_0, and the function $g(t_0) + (t - t_0)g'(t_0)$ is called the *linearization of f at* t_0, even though, strictly speaking, it is an affine function whenever $g(t_0) \neq 0$.

Now, lets move on to functions of two variables. The graph of $z = f(x, y)$ is a surface. The tangent plane to the graph of $z = f(x, y)$ at (x_0, y_0) is the plane that "best fits" this graph at (x_0, y_0) in the same way that the tangent line to the graph of $y = g(x)$ is the line that "best fits" this graph at x_0.

For example here is the graph of $z = x^2 + y^2$, together with the tangent plane to this graph at the point $(1, 1)$.



Here is another picture of the same thing from a different vantage point, giving a better view of the point of contact:



We can find an equation for this plane by looking along slices, and computing tangent lines. Here is how this goes.

Pick any $\mathbf{x}_0$ and pick any $\mathbf{x}$ so that $\|\mathbf{x}-\mathbf{x}_0\|$ is "small", or at least "not too large". We will say precisely what we mean by this soon, but for now, let us use the clear intuitive meaning.

Define a unit vector $\mathbf{u}$ by

$$\mathbf{u} = \frac{1}{\|\mathbf{x}-\mathbf{x}_0\|}(\mathbf{x}-\mathbf{x}_0) \ ,$$

and define a single variable function $g(t)$ by

$$g(t) = f(\mathbf{x}_0 + t\mathbf{u}) \ .$$

Now let us apply the tangent line approximation (4.3.1) to $g(t)$ at $t_0 = 0$ with $t = \|\mathbf{x}-\mathbf{x}_0\|$, which we recall is "small", so that this should be a good approximation. We have

$$g(\|\mathbf{x}-\mathbf{x}_0\|) \approx g(0) + \|\mathbf{x}-\mathbf{x}_0\| g'(0) \ . \tag{4.3.2}$$

We compute

$$\mathbf{x}_0 + \|\mathbf{x}-\mathbf{x}_0\|\mathbf{u} = \mathbf{x}_0 + (\mathbf{x}-\mathbf{x}_0) = \mathbf{x} \ ,$$

and so

$$g(\|\mathbf{x}-\mathbf{x}_0\|) = f(\mathbf{x}_0 + \|\mathbf{x}-\mathbf{x}_0\|\mathbf{u}) = f(\mathbf{x}) \ ,$$

while of course

$$g(0) = f(\mathbf{x}_0) \ .$$

Likewise, $g'(0) = \mathbf{u}\cdot\nabla f(\mathbf{x}_0)$ so that

$$\|\mathbf{x}-\mathbf{x}_0\| g'(0) = (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0) \ .$$

Therefore, (4.3.2) yields the approximation

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0) \ . \tag{4.3.3}$$

Definition 5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ have each of its partial derivatives defined and continuous at $\mathbf{x}_0$. Then the function $h(\mathbf{x})$*

$$h(\mathbf{x}) = (\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0)$$

is the linearization *of f at $\mathbf{x}_0$, and the graph of the linearization; i.e.,*

$$z = f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0)$$

is the tangent plane *to the graph of f at $\mathbf{x}_0$.*

You may be wondering why we have required the partial derivatives to be continuous, since, after all, the formulas specifying the tangent plane make sense as long as the partial derivatives merely exist. But as we shall see, if they merely exist, and are not continuous, the corresponding "tangent plane" would not at all give a good approximation, like the one we have displayed in the pictures above.

Example 16 (Direct tangent plane computations)**.** *Let $f(x,y) = x^4 + y^4 + 4xy$, as in Example 14 of the previous section, and let us take $\mathbf{x}_0 = (2,1)$. From the computations of the partial derivatives there,*

$$\nabla f(2,1) = ((36,12) \ .$$

We also compute $f(2,1) = 25$*. Thus, the linearization of* f *at* $(2,1)$ *is given by*

$$25 + (x-2, y-1)) \cdot ((36,12) = 36x + 12y - 59 \ .$$

Thus, the equation of the tangent plane is

$$z = 36x + 12y - 59 \ .$$

4.3.2 Critical points

Theorem 1 has important application to *optimization problems*, which are problems in which we look for minimum and maximum values of f, and the inputs $\mathbf{x}$ that produce them. Indeed, you see that:

- *If* $\nabla f(\mathbf{x}_0) \neq 0$*, then there is an "uphill" direction and a "downhill" direction at* $\mathbf{x}_0$*.*

If it is possible to "move uphill" from $\mathbf{x}_0$, then $f(\mathbf{x}_0)$ cannot possibly be a maximum value of f. Likewise, it is possible to "move downhill" from $\mathbf{x}_0$, then $f(\mathbf{x}_0)$ cannot possibly be a minimum value of f.

- *If we are looking for either minimum values of* f *or maximum values of* f *in some open set* U*, and* f *has continuous partial derivatives everywhere in* U*, then it suffices to look among only at those points* $\mathbf{x}$ *at which* $\nabla f(\mathbf{x}) = 0$*.*

This leads us to the definition of a critical point:

Definition 6 (Critical point)**.** *Suppose that* f *is defined and has a gradient in a neighborhood of some point* $\mathbf{x}_0$*. Then* $\mathbf{x}_0$ *is a* critical point *of* f *in case* $\nabla f(\mathbf{x}_0) = 0$*.*

Example 17 (Computing critical points)**.** *Let* $f(x,y) = x^4 + y^4 + 4xy$*, as in Example 14. We have already computed the gradient of* f*, and and found that* $\nabla f(x,y) = 0$ *if and only if* (x,y) *is one of the three points*

$$(0,0) \qquad (-1,1) \qquad \text{and} \qquad (1,-1) \ . \tag{4.3.4}$$

These are the critical points of f*.*

The three critical points found in Example 17 are the *only* points at which f can possibly take on either a maximum value or a minimum value. We will see later that f is minimized at both $(1. - 1)$ and $(-1,1)$, and that f has no maximum.

4.3.3 What differentiability means for a function on $\mathbb{R}^n$

Tangent planes are as central to the concept of differentiability in $\mathbb{R}^2$ as tangent lines are to the concept of differentiability in $\mathbb{R}$.

Roughly speaking, a function $f(x, y)$ is *differentiable* at (x_0, y_0) means that there is an affine function $h(x, y) = Ax + By + C$ such that

$$h(x, y) \approx f(x, y) \tag{4.3.5}$$

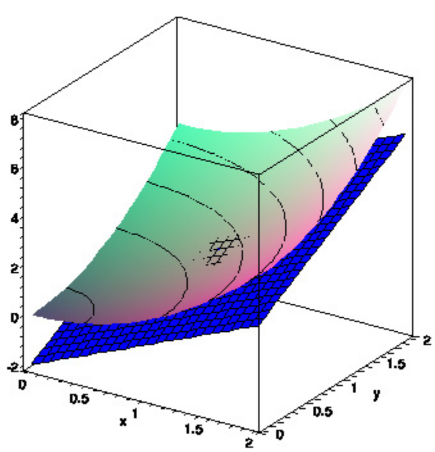
nearby (x_0, y_0), and the approximation in (4.3.5) is good enough that if you "zoom in" sufficiently closely on a joint graph of $z = f(x, y)$ and $z = h(x, y)$, you cannot see the difference.

For example, consider the function $f(x, y) = x^2 + y^2$ with $x_0 = 1$ and $y_0 = 1$. Then $f(x_0, y_0) = 2$ and $\nabla f(x_0, y_0) = (2, 2)$. Therefore, the linearization of f at $(1, 1)$ is

$$h(x, y) = 2 + 2(x - 2) + 2(y - 2) = 2x + 2y - 2\ .$$

Here is a three dimensional graph of f and h together for the region

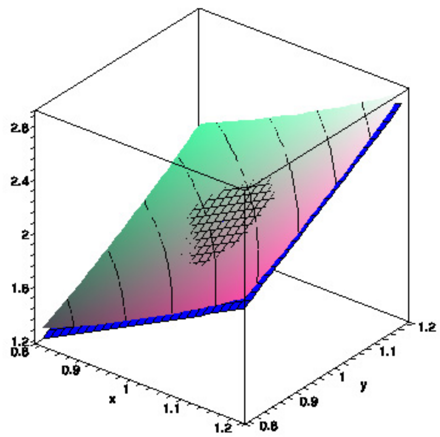
$$|x - 1| \le 1 \qquad \text{and} \qquad |y - 1| < 1\ \ :$$



As you see, the graphs are almost indistinguishable for x and y in the region

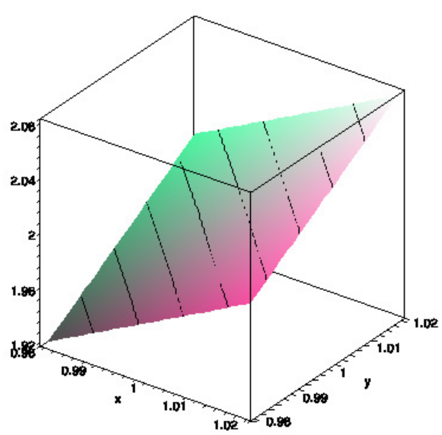
$$|x - 1| \le 0.2 \qquad \text{and} \qquad |y - 1| < 0.2\ .$$

Let's "zoom in" on this region:



The vertical separation between the graphs is getting to be a pretty small percentage of the displayed distances. The graphs are almost indistinguishable. Let's zoom in by a factor of 10, and have a look for

$$|x - 1| \leq 0.02 \qquad \text{and} \qquad |y - 1| < 0.02 \ .$$



Now, the graphs really are indistinguishable.

On the other hand, consider the function f defined by

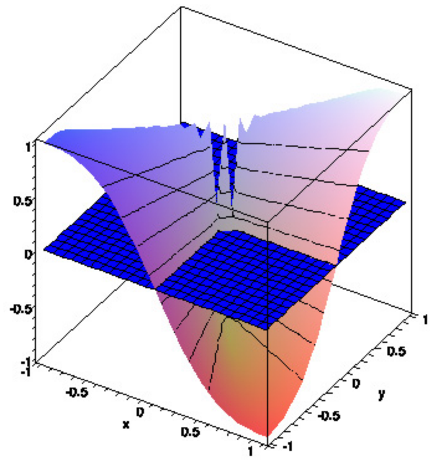
$$f(x,y) = \begin{cases} \dfrac{2xy}{x^4 + y^4} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} \tag{4.3.6}$$

As you can check, both partial derivatives are defined everywhere and

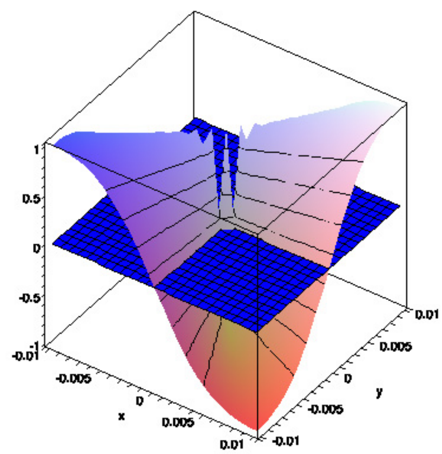
$$f(0,0) = 0 \qquad \frac{\partial f}{\partial x}(0,0) = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(0,0) = 0 \ .$$

Hence our formula for the best linear approximation gives us $h(x,y) = 0$. *This does not give a good approximation no matter how far you zoom in.* Here is a graph of f and h together for

$$|x-1| \leq 0.1 \qquad \text{and} \qquad |y-1| < 0.1 \ .$$



Anyone who calls that a "good fit" has pretty lax standards. Let's zoom in. Here is what we get when we zoom in by a factor of 10:



The graph looks just the same, except that as the scales indicate, we have zoomed in on the region

$$|x-1| \leq 0.1 \qquad \text{and} \qquad |y-1| < 0.1 \ .$$

The same thing happens no matter how much you zoom in.

• *The function* $f(x,y) = x^2 + y^2$ **is differentiable** *at* $x_0 = 1$ *and* $y_0 = 1$ *because if you zoom in enough, its graph is indistinguishable from that of its tangent plane. The function* f *defined by (4.3.6)* **is not differentiable** *at* $x_0 = 0$ *and* $y_0 = 0$ *because no matter how close you zoom, the graph never looks planar – there is no plane that really fits in the way a tangent plane should fit.*

Now that we understand in visual terms what differentiability is all about, let's phase it in mathematical terms.

In each of the graphs we drew above, we picked some number $\delta > 0$ which determined the *horizontal display size*. It determined the horizontal size simply because what we graphed was the part of the surfaces over

$$|x - x_0| \le \delta \qquad \text{and} \qquad |y - y_0| < \delta \ . \tag{4.3.7}$$

For x and y in this region, the vertical separation between the two graphs is

$$|f(x,y) - h(x,y)| \ .$$

For the graphs to be indistinguishable on this scale, we require that the vertical separation is a "small" fraction of δ at each x and y.

What passes for "small" depends on the resolution in our graph. For example, suppose we use a computer to draw the graphs and zoom in so that the *height and width* of the screen correspond to the distance 2δ. Then the two graphs will fill the screen horizontally. If the graph is 1000 pixels by 1000 pixels, and if for each x and y in the region (4.3.7), the difference

$$\frac{|f(x,y) - h(x,y)|}{2\delta} < \frac{1}{1000} \ ,$$

then the differences in the heights will be less than one pixel everywhere on the screen, and the two graphs will be indistinguishable.

So, indistinguishability has to do with the *ratio* of the vertical separation $|f(x,y) - h(x,y)|$ and the horizontal separation $|\mathbf{x} - \mathbf{x}_0|$. Therefore, define the function $r(\mathbf{x})$ by

$$r(\mathbf{x}) = \begin{cases} \dfrac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x} - \mathbf{x}_0|} & \text{if } \mathbf{x} \neq \mathbf{x}_0 \\ \\ 0 & \text{if } \mathbf{x} = \mathbf{x}_0 \ . \end{cases} \tag{4.3.8}$$

The function $r(\mathbf{x})$ is continuous at $\mathbf{x}_0$ if and only if the ratio of the vertical to horizontal separations, $\dfrac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x} - \mathbf{x}_0|}$, tends to zero as $\mathbf{x}$ approaches $\mathbf{x}_0$.

Recalling the ϵ and δ definition of continuity, this amounts to requiring that if $\epsilon > 0$ is any given threshold of "indistinguishability" that we care to set, then there is then a $\delta > 0$ so that

$$|\mathbf{x} - \mathbf{x}_0| < \delta \quad \Rightarrow \quad |h(\mathbf{x}) - f(\mathbf{x})| \le \epsilon |\mathbf{x} - \mathbf{x}_0| \ . \tag{4.3.9}$$

In this case, if you "zoom in" on a joint graph of $z = h(x,y)$ and $z = f(x,y)$ until the region with $|\mathbf{x}-\mathbf{x}_0| < \delta$ fills your screen, the vertical separation of the two graphs will be a negligible fraction – ϵ at most – of the width.

We are now ready to turn this into a precise mathematical definition of differentiability. We have so far considered only $n = 2$ because we relied on pictures to motivate the argument leading to (4.3.9). But since this makes sense for any n, we give the definition in $\mathbb{R}^n$.

Definition 7 (Differentiability of functions on $\mathbb{R}^n$)**.** *Let f be a function defined on an open set U in $\mathbb{R}^n$. then, for any $\mathbf{x}_0$ in U, f is differentiable at $\mathbf{x}_0$ in case it there is an affine function*

$$h(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a}\cdot(\mathbf{x}-\mathbf{x}_0)$$

so that

$$r(\mathbf{x}) = \begin{cases} \dfrac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x}-\mathbf{x}_0|} & \text{if } \mathbf{x} \neq \mathbf{x}_0 \\ \\ 0 & \text{if } \mathbf{x} = \mathbf{x}_0 \ . \end{cases} \tag{4.3.10}$$

is continuous at $\mathbf{x}_0$.

Note that the continuity of r at $\mathbf{x}_0$ is equivalent to the requirement that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x}-\mathbf{x}_0|} = 0 \ . \tag{4.3.11}$$

Also notice that $h(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a}\cdot(\mathbf{x}-\mathbf{x}_0)$ is the general affine function that has the same value as f at $\mathbf{x}_0$.

Theorem 2 (Differentiability of functions on $\mathbb{R}^n$ and the gradient)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{x}_0$. Then f is continuous at $\mathbf{x}_0$, and the gradient of f is well defined at $\mathbf{x}_0$, and the unique affine function $h(\mathbf{x})$ for which $r(\mathbf{x})$, defined by (4.3.10) is continuous is given by*

$$h(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)\cdot\nabla f(\mathbf{x}_0) \ .$$

Proof: Let $h(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{a}\cdot(\mathbf{x}-\mathbf{x}_0)$, and suppose that (4.3.11) is true for this h, and $r(\mathbf{x})$ defined by (4.3.10). We shall first show that f is continuous at $\mathbf{x}_0$. To do this, consider any sequence $\{\mathbf{x}_n\}$ with $\lim_{n\to\infty}\mathbf{x}_n = \mathbf{x}_0$. Then says that

$$\lim_{n\to\infty} \frac{|h(\mathbf{x}_n) - f(\mathbf{x}_n)|}{|\mathbf{x}_n-\mathbf{x}_0|} = 0 \ ,$$

and this certainly implies that $\lim_{n\to\infty} |h(\mathbf{x}_n) - f(\mathbf{x}_n)| = 0$.

But

$$|h(\mathbf{x}_n) - f(\mathbf{x}_n)| = |f(\mathbf{x}_0) + \mathbf{a}\cdot(\mathbf{x}_n - \mathbf{x}_0) - f(\mathbf{x}_n)| = |f(\mathbf{x}_n) - f(\mathbf{x}_0) - \mathbf{a}\cdot(\mathbf{x}_n - \mathbf{x}_0)| \ .$$

By the triangle inequality, and then the Cauchy-Schwarz inequality,

$$\begin{aligned} |f(\mathbf{x}_n) - f(\mathbf{x}_0)| &\le |f(\mathbf{x}_n) - f(\mathbf{x}_0) - \mathbf{a}\cdot(\mathbf{x}_n - \mathbf{x}_0)| + |\mathbf{a}\cdot(\mathbf{x}_n - \mathbf{x}_0)| \\ &\le |f(\mathbf{x}_n) - f(\mathbf{x}_0) - \mathbf{a}\cdot(\mathbf{x}_n - \mathbf{x}_0)| + \|\mathbf{a}\|\|\mathbf{x}_n - \mathbf{x}_0\| \\ &= |h(\mathbf{x}_n) - f(\mathbf{x}_n)| + \|\mathbf{a}\|\|\mathbf{x}_n - \mathbf{x}_0\| \ . \end{aligned}$$

Since $\lim_{n\to\infty} |h(\mathbf{x}_n) - f(\mathbf{x}_n)| = 0$ and $\lim_{n\to\infty} |\mathbf{x}_n - \mathbf{x}_0| = 0$, it follows that

$$\lim_{n\to\infty} |f(\mathbf{x}_n) - f(\mathbf{x}_0)| = 0 \ .$$

Thus, $f(\mathbf{x}_0) = \lim_{n\to\infty} f(\mathbf{x}_n)$, and hence f is continuous.

Next, taking $\mathbf{x} = \mathbf{x}_0 + h\mathbf{e}_j$, (4.3.11) says that

$$\lim_{t\to 0} \frac{|f(\mathbf{x}_0) + t\mathbf{a}\cdot\mathbf{e}_j - f(\mathbf{x}_0 + t\mathbf{e}_j)|}{|t|} = 0 \ .$$

But this is equivalent to

$$\lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t} = \mathbf{a}\cdot\mathbf{e}_j \ .$$

and hence

$$\mathbf{a}\cdot\mathbf{e}_j = \frac{\partial}{\partial x_j} f(\mathbf{x}_0) \ .$$

□

Now that we know what it means for a function to be differentiable, we can prove the following theorem which relates differentiability to the existence of continuous partial derivatives.

Theorem 3 (Differentiability and partial derivatives)**.** *Let f be any function defined in an open set U of $\mathbb{R}^2$ Suppose that both partial derivatives of f are defined and continuous at some point $\mathbf{x}_0$ in U. Then f is differentiable at $\mathbf{x}_0$.*

Proof of Theorem 3: The proof will be very similar to the proof of Theorem 1, and will make the same sort of use of continuity. Fix $\mathbf{x}$ and $\mathbf{x}_0$ and define numbers a and b by

$$x - x_0 = a \qquad \text{and} \qquad y - y_0 = b$$

We have see seen in Theorem 2 that if f is differentiable at $\mathbf{x}_0$, the gradient of f exists at $\mathbf{x}_0$ and the affine function $h(\mathbf{x})$ must be given by

$$h(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0) \cdot \nabla f(\mathbf{x}_0) \ .$$

We must show that for this choice of h,

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\|h(\mathbf{x}) - f(\mathbf{x})\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0 \ . \tag{4.3.12}$$

Define single variable functions ϕ and ψ by

$$\phi(s) = h(x_0, y_0 + sb) - f(x_0, y_0 + sb)$$

and

$$\psi(t) = h(x_0 + ta, y_0 + b) - f(x_0 + ta, y_0 + b) \ .$$

As you can check,

$$\begin{aligned} [\psi(1) - \psi(0)] + [\phi(1) - \phi(0)] &= \left[\Big(f(x,y) - h(x,y) \Big) - \Big(h(x_0, y) - f(x_0, y) \Big) \right] \\ &+ \left[\Big(f(x_0,y) - h(x_0,y) \Big) - \Big(f(x_0, y_0) - h(x_0, y_0) \Big) \right] \\ &= [h(\mathbf{x}) - f(\mathbf{x})] - [h(\mathbf{x}_0) - f(\mathbf{x}_0)] \ . \end{aligned}$$

Since $h(\mathbf{x}_0) = f(\mathbf{x}_0)$,

$$h(\mathbf{x}) - f(\mathbf{x}) = [\psi(1) - \psi(0)] + [\phi(1) - \phi(0)] \ . \tag{4.3.13}$$

We can now apply the Mean Value Theorem to the single variable functions $\psi(t)$ and $\phi(s)$ to find

$$[\psi(1) - \psi(0)] = \psi'(\tilde{t}) \qquad \text{and} \qquad [\phi(1) - \phi(0)] = \phi'(\tilde{s}) \tag{4.3.14}$$

for some $\tilde{s}$ and $\tilde{t}$ between 0 and 1.

Using the chain rule, and our explicit formula for h, we compute

$$\begin{aligned} \psi'(t) &= \left[\frac{\partial h}{\partial x}(x_0 + t\tilde{a}, y_0 + b) - \frac{\partial f}{\partial x}(x_0 + ta, y_0 + b) \right] a \\ &= \left[\frac{\partial f}{\partial x}(x_0, y_0) - \frac{\partial f}{\partial x}(x_0 + \tilde{t}a, y_0 + b) \right] a \ . \end{aligned}$$

Likewise,

$$\begin{aligned}\phi'(\tilde{s}) &= \left[\frac{\partial h}{\partial y}(x_0, y_0 + \tilde{s}b) - \frac{\partial f}{\partial y}(x_0, y_0 + \tilde{s}b)\right] b \\ &= \left[\frac{\partial f}{\partial y}(x_0, y_0) - \frac{\partial f}{\partial y}(x_0, y_0 + \tilde{s}b)\right] b\end{aligned}$$

Combining (4.3.13) and (4.3.14),

$$\begin{aligned}h(\mathbf{x}) - f(\mathbf{x}) &= \left[A - \frac{\partial f}{\partial x}(x_0 + \tilde{t}a, y_0 + b)\right] a \\ &+ \left[B - \frac{\partial f}{\partial y}(x_0, y_0 + \tilde{s}b)\right] b\ .\end{aligned}$$

The right hand side is the dot product of

$$\left(\frac{\partial f}{\partial x}(x_0, y_0) - \frac{\partial f}{\partial x}(x_0 + \tilde{t}a, y_0 + b)\ ,\ \frac{\partial f}{\partial y}(x_0, y_0) - \frac{\partial f}{\partial y}(x_0, y_0 + \tilde{s}b)\right) \quad \text{with} \quad (a, b) = \mathbf{x} - \mathbf{x}_0\ .$$

Therefore, by the Schwarz inequality,

$$\frac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x} - \mathbf{x}_0|} \le \sqrt{\left(\frac{\partial f}{\partial x}(x_0, y_0) - \frac{\partial f}{\partial x}(x_0 + \tilde{t}a, y_0 + b)\right)^2 + \left(\frac{\partial f}{\partial y}(x_0, y_0) - \frac{\partial f}{\partial y}(x_0, y_0 + \tilde{s}b)\right)^2}\ .$$

Now, as $\mathbf{x} \to \mathbf{x}_0$, $a \to 0$ and $b \to 0$. Therefore, *using the continuity of the partial derivatives at this point*,

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{|h(\mathbf{x}) - f(\mathbf{x})|}{|\mathbf{x} - \mathbf{x}_0|} = 0\ ,$$

as was to be shown. □