The Calculus of Functions

of

Several Variables

Section 3.3

Best Affine Approximations

Best affine approximations

Given a function $f:\mathbb{R}^n \to \mathbb{R}$ and a point $\mathbf{c}$, we wish to find the affine function $A:\mathbb{R}^n \to \mathbb{R}$ which best approximates f for points close to $\mathbf{c}$. As before, best will mean that the remainder function,

$$R(\mathbf{h}) = f(\mathbf{c}+\mathbf{h}) - A(\mathbf{c}+\mathbf{h}), \tag{3.3.1}$$

approaches 0 at a sufficiently fast rate. In this context, since $R(\mathbf{h})$ is a scalar and $\mathbf{h}$ is a vector, sufficiently fast will mean that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{3.3.2}$$

Generalizing our previous notation, we will say that a function $R:\mathbb{R}^n \to \mathbb{R}$ satisfying (3.3.2) is $o(\mathbf{h})$. Note that if $n = 1$ this extended definition of $o(\mathbf{h})$ is equivalent to the definition given in Section 2.2.

Definition Suppose $f:\mathbb{R}^n \to \mathbb{R}$ is defined on an open ball containing the point $\mathbf{c}$. We call an affine function $A:\mathbb{R}^n \to \mathbb{R}$ the *best affine approximation* to f at $\mathbf{c}$ if (1) $A(\mathbf{c}) = f(\mathbf{c})$ and (2) $R(\mathbf{h})$ is $o(\mathbf{h})$, where

$$R(\mathbf{h}) = f(\mathbf{c}+\mathbf{h}) - A(\mathbf{c}+\mathbf{h}). \tag{3.3.3}$$

Suppose $f:\mathbb{R}^n \to \mathbb{R}$ and suppose $A:\mathbb{R}^n \to \mathbb{R}$ is the best affine approximation to f at $\mathbf{c}$. Since A is affine, there exists a linear function $L:\mathbb{R}^n \to \mathbb{R}$ and a scalar b such that

$$A(\mathbf{x}) = L(\mathbf{x}) + b \tag{3.3.4}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$. Since $A(\mathbf{c}) = f(\mathbf{c})$, we have

$$f(\mathbf{c}) = L(\mathbf{c}) + b, \tag{3.3.5}$$

which implies that

$$b = f(\mathbf{c}) - L(\mathbf{c}). \tag{3.3.6}$$

Hence

$$A(\mathbf{x}) = L(\mathbf{x}) + f(\mathbf{c}) - L(\mathbf{c}) = L(\mathbf{x}-\mathbf{c}) + f(\mathbf{c}) \tag{3.3.7}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$. Moreover, if we let

$$\mathbf{a} = (L(\mathbf{e}_1), L(\mathbf{e}_2), \ldots, L(\mathbf{e}_n)), \tag{3.3.8}$$

where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are, as usual, the standard basis vectors for $\mathbb{R}^n$, then, from our results in Section 1.5,

$$L(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} \tag{3.3.9}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$. Hence

$$A(\mathbf{x}) = \mathbf{a} \cdot (\mathbf{x} - \mathbf{c}) + f(\mathbf{c}), \tag{3.3.10}$$

for all $\mathbf{x}$ in $\mathbb{R}^n$, and we see that A is completely determined by the vector $\mathbf{a}$

Definition Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is defined on an open ball containing the point $\mathbf{c}$. If f has a best affine approximation at $\mathbf{c}$, then we say f is *differentiable* at $\mathbf{c}$. Moreover, if the best affine approximation to f at $\mathbf{c}$ is given by

$$A(\mathbf{x}) = \mathbf{a} \cdot (\mathbf{x} - \mathbf{c}) + f(\mathbf{c}), \tag{3.3.11}$$

then we call $\mathbf{a}$ the *derivative* of f at $\mathbf{c}$ and write $Df(\mathbf{c}) = \mathbf{a}$.

Now suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{c}$ with best affine approximation A and let $\mathbf{a} = (a_1, a_2, \ldots, a_n) = Df(\mathbf{c})$. Since

$$R(\mathbf{h}) = f(\mathbf{c} + \mathbf{h}) - A(\mathbf{c} + \mathbf{h}) = f(\mathbf{c} + \mathbf{h}) - \mathbf{a} \cdot \mathbf{h} - f(\mathbf{c}) \tag{3.3.12}$$

is $o(\mathbf{h})$, we must have

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{3.3.13}$$

In particular, for $k = 1, 2, \ldots, n$, if we let $\mathbf{h} = t\mathbf{e}_k$, then $\mathbf{h}$ approaches $\mathbf{0}$ as t approaches 0, so

$$0 = \lim_{t \to 0} \frac{R(t\mathbf{e}_k)}{\|t\mathbf{e}_k\|} = \lim_{t \to 0} \frac{f(\mathbf{c} + t\mathbf{e}_k) - t(\mathbf{a} \cdot \mathbf{e}_k) - f(\mathbf{c})}{|t|} = \lim_{t \to 0} \frac{f(\mathbf{c} + t\mathbf{e}_k) - ta_k - f(\mathbf{c})}{|t|}$$

First considering $t > 0$, we have

$$0 = \lim_{t \to 0^+} \frac{f(\mathbf{c} + t\mathbf{e}_k) - ta_k - f(\mathbf{c})}{t} = \lim_{t \to 0^+} \left(\frac{f(\mathbf{c} + t\mathbf{e}_k) - f(\mathbf{c})}{t} - a_k \right), \tag{3.3.14}$$

implying that

$$a_k = \lim_{t \to 0^+} \frac{f(\mathbf{c} + t\mathbf{e}_k) - f(\mathbf{c})}{t}. \tag{3.3.15}$$

With $t < 0$, we have

$$0 = \lim_{t \to 0^-} \frac{f(\mathbf{c} + t\mathbf{e}_k) - ta_k - f(\mathbf{c})}{-t} = - \lim_{t \to 0^-} \left(\frac{f(\mathbf{c} + t\mathbf{e}_k) - f(\mathbf{c})}{t} - a_k \right), \tag{3.3.16}$$

implying that

$$a_k = \lim_{t \to 0^-} \frac{f(\mathbf{c} + t\mathbf{e}_k) - f(\mathbf{c})}{t}. \tag{3.3.17}$$

Hence

$$a_k = \lim_{t \to 0} \frac{f(\mathbf{c} + t\mathbf{e}_k) - f(\mathbf{c})}{t} = \frac{\partial}{\partial x_k} f(\mathbf{c}). \tag{3.3.18}$$

Thus we have shown that

$$\mathbf{a} = \left(\frac{\partial}{\partial x_1} f(\mathbf{c}), \frac{\partial}{\partial x_2} f(\mathbf{c}), \ldots, \frac{\partial}{\partial x_n} f(\mathbf{c}) \right) = \nabla f(\mathbf{c}). \tag{3.3.19}$$

Theorem If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{c}$, then

$$Df(\mathbf{c}) = \nabla f(\mathbf{c}) \tag{3.3.20}$$

It now follows that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{c}$, then the best affine approximation to f at $\mathbf{c}$ is

$$A(\mathbf{x}) = \nabla f(\mathbf{c}) \cdot (\mathbf{x} - \mathbf{c}) - f(\mathbf{c}). \tag{3.3.21}$$

However, the converse does not hold: it is possible for $\nabla f(\mathbf{c})$ to exist even when f is not differentiable at $\mathbf{c}$. Before looking at an example, note that if f is differentiable at $\mathbf{c}$ and A is the best affine approximation to f at $\mathbf{c}$, then, since $R(\mathbf{h}) = f(\mathbf{c} + \mathbf{h}) - A(\mathbf{c} + \mathbf{h})$ is $o(\mathbf{h})$,

$$\lim_{\mathbf{h} \to \mathbf{0}} (f(\mathbf{c} + \mathbf{h}) - A(\mathbf{c} + \mathbf{h})) = \lim_{\mathbf{h} \to \mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|} \|\mathbf{h}\| = 0\|\mathbf{0}\| = 0. \tag{3.3.22}$$

Now A is continuous at $\mathbf{c}$, so it follows that

$$\lim_{\mathbf{h} \to \mathbf{0}} f(\mathbf{c} + \mathbf{h}) = \lim_{\mathbf{h} \to \mathbf{0}} A(\mathbf{c} + \mathbf{h}) = A(\mathbf{c}) = f(\mathbf{c}). \tag{3.3.23}$$

In other words, f is continuous at $\mathbf{c}$.

Theorem If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{c}$, then f is continuous at $\mathbf{c}$.

Example Consider the function

$$g(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2}, & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

In Section 3.1 we showed that g is not continuous at $(0, 0)$ and in Section 3.2 we saw that $\nabla g(0, 0) = (0, 0)$. Since g is not continuous at $(0, 0)$, it now follows, from the previous theorem, that g is not differentiable at $(0, 0)$, even though the gradient exists at that point. From the graph of g in Figure 3.3.1 (originally seen in Figure 3.1.7), we can see

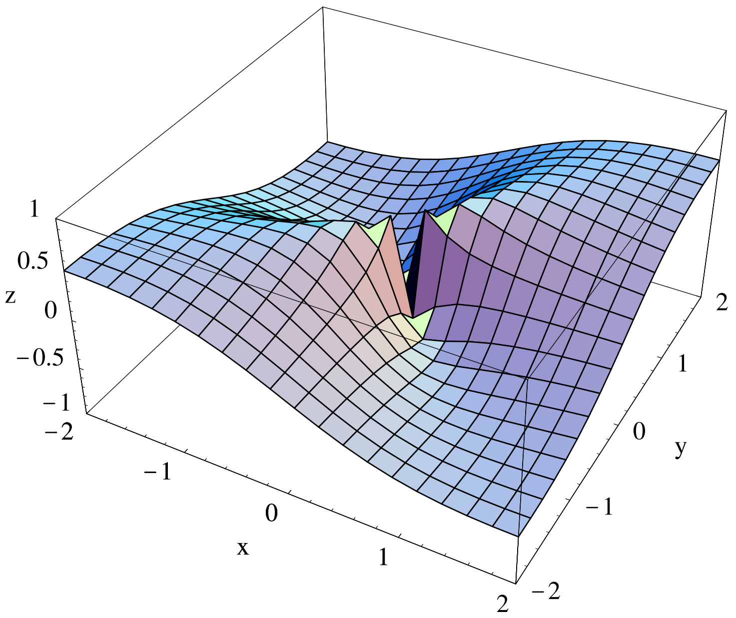

Figure 3.3.1 The graph of a nondifferentiable function

that the fact that g is not differentiable, in fact, not even continuous, at the origin shows up geometrically as a tear in the surface.

From this example we see that the differentiability of a function $f : \mathbb{R}^n \to \mathbb{R}$ at a point $\mathbf{c}$ requires more than just the existence of the gradient of f at $\mathbf{c}$. It turns out that continuity of the partial derivatives of f on an open ball containing $\mathbf{c}$ suffices to show that f is differentiable at $\mathbf{c}$. Note that the partial derivatives of g in the previous example are not continuous (see Problem 8 of Section 3.2).

So we will now assume that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 on some open ball containing $\mathbf{c}$. If we define an affine function $A : \mathbb{R}^n \to \mathbb{R}$ by

$$A(\mathbf{x}) = \nabla f(\mathbf{c}) \cdot (\mathbf{x} - \mathbf{c}) + f(\mathbf{c}), \tag{3.3.24}$$

then the remainder function is

$$R(\mathbf{h}) = f(\mathbf{c} + \mathbf{h}) - A(\mathbf{c} + \mathbf{h}) = f(\mathbf{c} + \mathbf{h}) - f(\mathbf{c}) - \nabla f(\mathbf{c}) \cdot \mathbf{h}. \tag{3.3.25}$$

We need to show that $R(\mathbf{h})$ is $o(\mathbf{h})$. Toward that end, for a fixed $\mathbf{h} \neq \mathbf{0}$, define $\varphi : \mathbb{R} \to \mathbb{R}$ by

$$\varphi(t) = f(\mathbf{c} + t\mathbf{h}). \tag{3.3.26}$$

We first note that φ is differentiable with

$$\begin{aligned} \varphi'(t) &= \lim_{s \to 0} \frac{\varphi(t+s) - \varphi(t)}{s} \\ &= \lim_{s \to 0} \frac{f(\mathbf{c} + (t+s)\mathbf{h}) - f(\mathbf{c} + t\mathbf{h})}{s} \end{aligned}$$

$$
\begin{aligned}
&= \|\mathbf{h}\| \lim_{s\to 0} \frac{f\left(\mathbf{c}+t\mathbf{h}+s\|\mathbf{h}\|\frac{\mathbf{h}}{\|\mathbf{h}\|}\right)-f(\mathbf{c}+t\mathbf{h})}{s\|\mathbf{h}\|}\\
&= \|\mathbf{h}\| D_{\frac{\mathbf{h}}{\|\mathbf{h}\|}} f(\mathbf{c}+t\mathbf{h})\\
&= \|\mathbf{h}\| \left(\nabla f(\mathbf{c}+t\mathbf{h})\cdot \frac{\mathbf{h}}{\|\mathbf{h}\|}\right)\\
&= \nabla f(\mathbf{c}+t\mathbf{h})\cdot \mathbf{h}. \qquad (3.3.27)
\end{aligned}
$$

From the Mean Value Theorem of single-variable calculus, it follows that there exists a number s between 0 and 1 such that

$$\varphi'(s) = \varphi(1) - \varphi(0) = f(\mathbf{c}+\mathbf{h}) - f(\mathbf{c}). \qquad (3.3.28)$$

Hence we may write

$$R(\mathbf{h}) = \nabla f(\mathbf{c}+s\mathbf{h})\cdot\mathbf{h} - \nabla f(\mathbf{c})\cdot\mathbf{h} = (\nabla f(\mathbf{c}+s\mathbf{h}) - \nabla f(\mathbf{c}))\cdot\mathbf{h}. \qquad (3.3.29)$$

Applying the Cauchy-Schwarz inequality to (3.3.29),

$$|R(\mathbf{h})| \le \|\nabla f(\mathbf{c}+s\mathbf{h}) - \nabla f(\mathbf{c})\|\|\mathbf{h}\|, \qquad (3.3.30)$$

and so

$$\frac{|R(\mathbf{h})|}{\|\mathbf{h}\|} \le \|\nabla f(\mathbf{c}+s\mathbf{h}) - \nabla f(\mathbf{c})\|. \qquad (3.3.31)$$

Now the partial derivatives of f are continuous, so

$$
\begin{aligned}
\lim_{\mathbf{h}\to\mathbf{0}} \|\nabla f(\mathbf{c}+s\mathbf{h}) - \nabla f(\mathbf{c})\| &= \|\nabla f(\mathbf{c}+s\mathbf{0}) - \nabla f(\mathbf{c})\|\\
&= \|\nabla f(\mathbf{c}) - \nabla f(\mathbf{c})\| \qquad (3.3.32)\\
&= 0.
\end{aligned}
$$

Hence

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{R(\mathbf{h})}{\|\mathbf{h}\|} = 0. \qquad (3.3.33)$$

That is, $R(\mathbf{h})$ is $o(\mathbf{h})$ and A is the best affine approximation to f at $\mathbf{c}$. Thus we have the following fundamental theorem.

Theorem If $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 on an open ball containing the point $\mathbf{c}$, then f is differentiable at $\mathbf{c}$.

Example Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

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

To find the best affine approximation to f at $(1,1)$, we first compute

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

Thus $\nabla f(1,1) = (-4,-2)$ and $f(1,1) = 1$, so the best affine approximation is

$$A(x,y) = (-4,-2)\cdot(x-1, y-1) + 1.$$

Simplifying, we have

$$A(x,y) = -4x - 2y + 7.$$

Example Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by

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

Then

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

Thus, for example, the best affine approximation to f at $(2,1,2)$ is

$$\begin{aligned} A(x,y,z) &= \nabla f(2,1,2)\cdot(x-2, y-1, z-2) + f(2,1,2) \\ &= \frac{1}{3}(2,1,2)\cdot(x-2,y-1,z-2) + 3 \\ &= \frac{2}{3}(x-2) + \frac{1}{3}(y-1) + \frac{2}{3}(z-2) + 3 \\ &= \frac{2}{3}x + \frac{1}{3}y + \frac{2}{3}z. \end{aligned}$$

Now suppose we let (x,y,z) be the lengths of the three sides of a solid block, in which case $f(x,y,z)$ represents the length of the diagonal of the box. Moreover, suppose we measure the sides of the block and find them to have lengths $x = 2+\epsilon_x$, $y = 1+\epsilon_y$, and $z = 2+\epsilon_z$, where $|\epsilon_x| \le h$, $|\epsilon_y| \le h$, and $|\epsilon_z| \le h$ for some positive number h representing the limit of the accuracy of our measuring device. We now estimate the diagonal of the box to be

$$f(2,1,2) = 3$$

with an error of

$$\begin{aligned} |f(2+\epsilon_x, 1+\epsilon_y, 2+\epsilon_z) - f(2,1,2)| &\approx |A(2+\epsilon_x, 1+\epsilon_y, 2+\epsilon_z) - 3| \\ &= \left|\frac{2}{3}\epsilon_x + \frac{1}{3}\epsilon_y + \frac{2}{3}\epsilon_z\right| \\ &\le \frac{2}{3}|\epsilon_x| + \frac{1}{3}|\epsilon_y| + \frac{2}{3}|\epsilon_z| \\ &\le h\left(\frac{2}{3}+\frac{1}{3}+\frac{2}{3}\right) \\ &= \frac{5}{3}h. \end{aligned}$$

That is, we expect our error in estimating the diagonal of the block to be no more that $\frac{5}{3}$ times the maximum error in our measurements of the sides of the block. For example, if the error in our length measurements is off by no more than ± 0.1 centimeters, then our estimate of the diagonal of the box is off by no more than ± 0.17 centimeters.

Note that if $A : \mathbb{R}^n \to \mathbb{R}$ is the best affine approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{c} = (c_1, c_2, \ldots, c_n)$, then the graph of A is the set of all points $(x_1, x_2, \ldots, x_n, z)$ in $\mathbb{R}^{n+1}$ satisfying

$$z = \nabla f(\mathbf{c}) \cdot (x_1 - c_1, x_2 - c_2, \ldots, x_n - c_n) + f(\mathbf{c}). \tag{3.3.34}$$

Letting

$$\mathbf{n} = \left(\frac{\partial}{\partial x_1} f(\mathbf{c}), \frac{\partial}{\partial x_2} f(\mathbf{c}), \ldots, \frac{\partial}{\partial x_n} f(\mathbf{c}), -1\right), \tag{3.3.35}$$

we may describe the graph of A as the set of all points in $\mathbb{R}^{n+1}$ satisfying

$$\mathbf{n} \cdot (x_1 - c_1, x_2 - c_2, \ldots, x_n - c_n, z - f(\mathbf{c})) = 0. \tag{3.3.36}$$

Thus the graph of A is a hyperplane in $\mathbb{R}^{n+1}$ passing through the point $(c_1, c_2, \ldots, c_n, f(\mathbf{c}))$ (a point on the graph of f) with normal vector $\mathbf{n}$.

Definition If $A : \mathbb{R}^n \to \mathbb{R}$ is the best affine approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{c} = (c_1, c_2, \ldots, c_n)$, then we call the graph of A the *tangent hyperplane* to the graph of f at $(c_1, c_2, \ldots, c_n, f(\mathbf{c}))$.

Example We saw above that the best affine approximation to

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

at $(1, 1)$ is

$$A(x, y) = 7 - 4x - 2y.$$

Hence the equation of the tangent plane to the graph of f at is

$$z = 7 - 4x - 2y,$$

or

$$4x + 2y + z = 7.$$

Note that the vector $\mathbf{n} = (4, 2, 1)$ is normal to the tangent plane, and hence normal to the graph of f at $(1, 1, 1)$. The graph of f along with the tangent plane at $(1, 1, 1)$ is shown in Figure 3.3.2.

The chain rule

Suppose $\varphi : \mathbb{R} \to \mathbb{R}^n$ is differentiable at a point c and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at the point $\varphi(c)$. Then the composition of f and φ is a function $f \circ \varphi : \mathbb{R} \to \mathbb{R}$. To compute the derivative of $f \circ \varphi$ at c, we must evaluate

$$(f \circ \varphi)'(c) = \lim_{h \to 0} \frac{f \circ \varphi(c + h) - f \circ \varphi(c)}{h} = \lim_{h \to 0} \frac{f(\varphi(c + h)) - f(\varphi(c))}{h}. \tag{3.3.37}$$

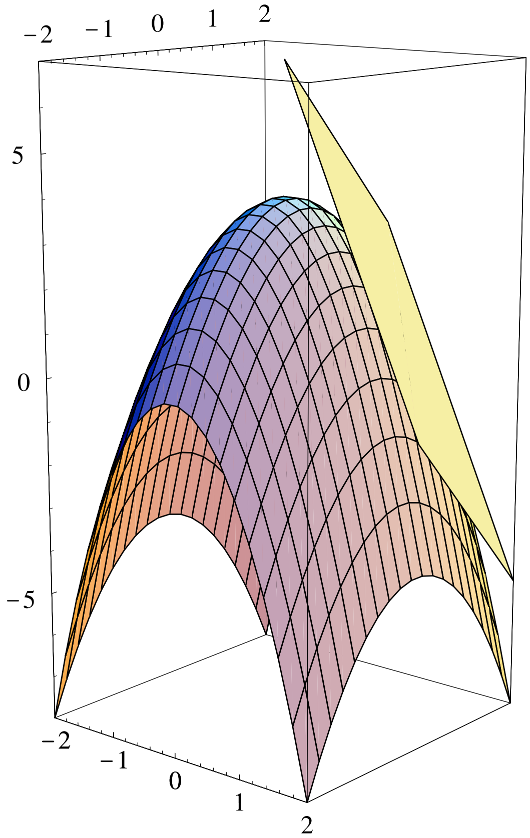

Figure 3.3.2 A plane tangent to the graph of $f(x, y) = 4 - 2x^2 - y^2$

Let A be the best affine approximation to f at $\mathbf{a} = \varphi(c)$ and let $\mathbf{k} = \varphi(c+h) - \varphi(c)$. Then

$$f(\varphi(c+h)) = f(\mathbf{a}+\mathbf{k}) = A(\mathbf{a}+\mathbf{k}) + R(\mathbf{k}), \tag{3.3.38}$$

where $R(\mathbf{k})$ is $o(\mathbf{k})$. Now

$$A(\mathbf{a}+\mathbf{k}) = \nabla f(\mathbf{a}) \cdot \mathbf{k} + f(\mathbf{a}), \tag{3.3.39}$$

so

$$\begin{aligned} f(\varphi(c+h)) - f(\varphi(c)) &= f(\mathbf{a}+\mathbf{k}) - f(\mathbf{a}) \\ &= \nabla f(\mathbf{a}) \cdot \mathbf{k} + R(\mathbf{k}) \\ &= \nabla f(\mathbf{a}) \cdot (\varphi(c+h) - \varphi(c)) + R(\mathbf{k}). \end{aligned} \tag{3.3.40}$$

Substituting (3.3.40) into (3.3.37),we have

$$\begin{aligned} (f \circ \varphi)'(c) &= \lim_{h \to 0} \frac{\nabla f(\mathbf{a}) \cdot (\varphi(c+h) - \varphi(c)) + R(\mathbf{k})}{h} \\ &= \lim_{h \to 0} \nabla f(\mathbf{a}) \cdot \frac{\varphi(c+h) - \varphi(c)}{h} + \lim_{h \to 0} \frac{R(\mathbf{k})}{h} \\ &= \nabla f(\mathbf{a}) \cdot D\varphi(\mathbf{c}) + \lim_{h \to 0} \frac{R(\mathbf{k})}{h}. \end{aligned} \tag{3.3.41}$$

Now $R(\mathbf{k})$ is $o(\mathbf{k})$, so

$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{R(\mathbf{k})}{\|\mathbf{k}\|} = 0,$$

from which it follows that, for any given $\epsilon > 0$, we have

$$\frac{|R(\mathbf{k})|}{\|\mathbf{k}\|} < \epsilon \tag{3.3.42}$$

for sufficiently small $\mathbf{k} \neq 0$. Since $R(\mathbf{0}) = 0$, it follows that

$$|R(\mathbf{k})| < \epsilon\|\mathbf{k}\| \tag{3.3.43}$$

for all $\mathbf{k}$ sufficiently small. Moreover, φ is continuous at c, so we may choose h small enough to guarantee that

$$\mathbf{k} = \varphi(c+h) - \varphi(h)$$

is small enough for (3.3.43) to hold. Hence for sufficiently small $h \neq 0$,

$$\frac{|R(\mathbf{k})|}{h} < \frac{\epsilon\|\mathbf{k}\|}{h}. \tag{3.3.44}$$

Now

$$\lim_{h\to 0} \frac{\|\mathbf{k}\|}{h} = \lim_{h\to 0} \frac{\|\varphi(c+h) - \varphi(c)\|}{h} = \|D\varphi(c)\| \tag{3.3.45}$$

and the choice of ϵ was arbitrary, so it follows that

$$\lim_{h\to\mathbf{0}} \frac{R(\mathbf{k})}{h} = 0. \tag{3.3.46}$$

Hence

$$(f \circ \varphi)'(c) = \nabla f(\mathbf{a}) \cdot D\varphi(c). \tag{3.3.47}$$

This is a version of the chain rule.

Theorem Suppose $\varphi : \mathbb{R} \to \mathbb{R}^n$ is differentiable at c and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\varphi(c)$. Then

$$(f \circ \varphi)'(c) = \nabla f(\varphi(c)) \cdot D\varphi(c). \tag{3.3.48}$$

If we imagine a particle moving along the curve C parametrized by φ, with velocity $\mathbf{v}(t)$ and unit tangent vector $T(t)$ at time t, then (3.3.48) says that the rate of change of f along C at $\varphi(c)$ is

$$\nabla f(\varphi(c)) \cdot \mathbf{v}(c) = \|\mathbf{v}(c)\|\nabla f(\varphi(c)) \cdot T(c) = \|\mathbf{v}(c)\| D_{T(c)} f(\varphi(c)). \tag{3.3.49}$$

In other words, the rate of change of f along C is the rate of change of f in the direction of $T(t)$ multiplied by the speed of the particle moving along the curve.

Example Suppose that the temperature at a point (x, y, z) inside a cubical region of space is given by

$$T(x, y, z) = 80 - 20xe^{-\frac{1}{20}(x^2+y^2+z^2)}.$$

Moreover, suppose a bug flies through this region along the elliptical helix parametrized by

$$\varphi(t) = (\cos(\pi t), 2\sin(\pi t), t).$$

Then

$$\nabla T(x, y, z) = e^{-\frac{1}{20}(x^2+y^2+z^2)}(2x^2 - 20, 2xy, 2xz)$$

and

$$D\varphi(t) = (-\pi\sin(\pi t), 2\pi\cos(\pi t), 1).$$

Hence, for example, if we want to know the rate of change of temperature for the bug at $t = \frac{1}{3}$, we would evaluate

$$D\varphi\left(\frac{1}{3}\right) = \left(-\frac{\sqrt{3}\pi}{2}, \pi, 1\right)$$

and

$$\nabla T\left(\varphi\left(\frac{1}{3}\right)\right) = \nabla T\left(\frac{1}{2}, \sqrt{3}, \frac{1}{3}\right) = e^{-\frac{121}{720}}\left(-\frac{39}{2}, \sqrt{3}, \frac{1}{3}\right),$$

so

$$\begin{aligned}(T\circ\varphi)'\left(\frac{1}{3}\right) &= e^{-\frac{121}{720}}\left(-\frac{39}{2}, \sqrt{3}, \frac{1}{3}\right)\cdot\left(-\frac{\sqrt{3}\pi}{2}, \pi, 1\right)\\ &= e^{-\frac{121}{720}}\left(\frac{39\pi\sqrt{3}}{4} + \sqrt{3}\pi + \frac{1}{3}\right)\\ &= 49.73,\end{aligned}$$

where the final value has been rounded to two decimal places. Hence at that moment the temperature for the bug is increasing at rate of 49.73° per second. We could also express this as

$$\left.\frac{dT}{dt}\right|_{t=\frac{1}{3}} = 49.73°.$$

For an alternative formulation of the chain rule, suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $x_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, n$, are all differentiable and let $w = f(x_1, x_2, \ldots, x_n)$. If x_1, x_2, $\ldots$, x_n are all functions of t, then, by the chain rule,

$$\begin{aligned}\frac{dw}{dt} &= \left(\frac{\partial w}{\partial x_1}, \frac{\partial w}{\partial x_2}, \ldots, \frac{\partial w}{\partial x_n}\right)\cdot\left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}\right)\\ &= \frac{\partial w}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial w}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial w}{\partial x_n}\frac{dx_n}{dt}.\end{aligned} \tag{3.3.50}$$

Example Suppose the dimensions of a box are increasing so that its length, width, and height at time t are, in centimeters,

$$x = 3t,$$

$$y = t^2,$$

and

$$z = t^3,$$

respectively. Since the volume of the box is

$$V = xyz,$$

the rate of change of the volume is

$$\frac{dV}{dt} = \frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt} = 3yz + 2xzt + 3xyt^2.$$

Hence, for example, at $t = 2$ we have $x = 6$, $y = 4$, and $z = 8$, so

$$\left.\frac{dV}{dt}\right|_{t=2} = 96 + 192 + 288 = 576 \text{ cm}^3/\text{sec}.$$

The gradient and level sets

Now consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\mathbf{a}$ on the level set S specified by $f(\mathbf{x}) = c$ for some scalar c. Suppose $\varphi : \mathbb{R} \to \mathbb{R}^n$ is a smooth parametrization of a curve C which lies entirely on S and passes through $\mathbf{a}$. Let $\varphi(b) = \mathbf{a}$. Then the composition of f and φ is a constant function; that is,

$$g(t) = f \circ \varphi(t) = f(\varphi(t)) = c \tag{3.3.51}$$

for all values of t. Thus, using the chain rule,

$$0 = g'(b) = \nabla f(\varphi(b)) \cdot D\varphi(b) = \nabla f(\mathbf{a}) \cdot D\varphi(b). \tag{3.3.52}$$

Hence

$$\nabla f(\mathbf{a}) \perp D\varphi(b). \tag{3.3.53}$$

Now $D\varphi(b)$ is tangent to C at $\mathbf{a}$; moreover, since (3.3.53) holds for any curve in S passing through $\mathbf{a}$, $\nabla f(\mathbf{a})$ is orthogonal to every vector tangent to S. In other words, $\nabla f(\mathbf{a})$ is normal to the hyperplane tangent to S at $\mathbf{a}$. Thus we have the following theorem.

Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable on an open ball containing the point $\mathbf{a}$, and let S be the set of all points in $\mathbb{R}^n$ such that $f(\mathbf{x}) = f(\mathbf{a})$. If $\nabla f(\mathbf{a}) \neq \mathbf{0}$, then the hyperplane with equation

$$\nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0 \tag{3.3.54}$$

is tangent to S at $\mathbf{a}$.

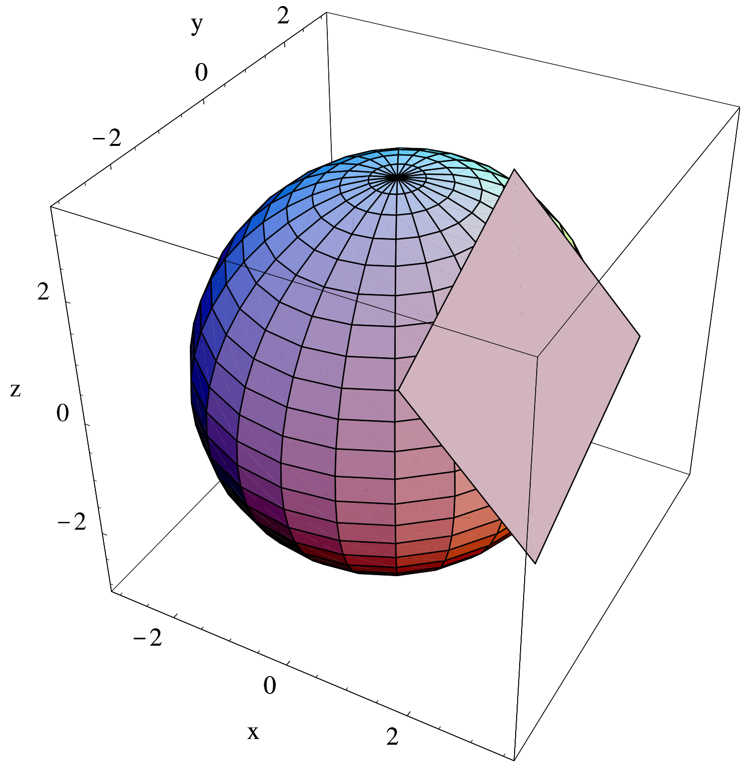

Figure 3.3.3 Sphere with tangent plane

For $n = 2$, the hyperplane described by (3.3.54) will be a tangent line to a curve; for $n = 3$, it will be a tangent plane to a surface.

Example The set of all points S in $\mathbb{R}^3$ satisfying

$$x^2 + y^2 + z^2 = 9$$

is a sphere with radius 3 centered at the origin. We will find an equation for the plane tangent to S at $(2, -1, 2)$. First note that S is a level surface for the function

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

Now

$$\nabla f(x, y, z) = (2x, 2y, 2z),$$

so

$$\nabla f(2, -1, 2) = (4, -2, 4).$$

Thus an equation for the tangent plane is

$$(4, -2, 4) \cdot (x - 2, y + 1, z - 2) = 0,$$

or

$$4x - 2y + 4z = 18.$$

See Figure 3.3.3.

Problems

1. For each of the following, find the best affine approximation to the given function at the specified point $\mathbf{c}$.
   (a) $f(x,y) = 3x^2 + 4y^2 - 2$, $\mathbf{c} = (2,1)$
   (b) $g(x,y) = y^2 - x^2$, $\mathbf{c} = (1,-2)$
   (c) $g(x,y) = y^2 - x^2$, $\mathbf{c} = (0,0)$
   (d) $f(x,y,z) = -\log(x^2 + y^2 + z^2)$, $\mathbf{c} = (1,0,0)$
   (e) $h(w,x,y,z) = w^2 + x^2 + 3y^2 = 2z^2$, $\mathbf{c} = (1,2,-2,1)$

2. For each of the following, find the equation of the plane tangent to the graph of f for the given point $\mathbf{c}$. Plot the graph and the tangent plane together.
   (a) $f(x,y) = 4x^2 + y^2$, $\mathbf{c} = (1,-1)$
   (b) $f(x,y) = \sqrt{9 - x^2 - y^2}$, $\mathbf{c} = (2,1)$
   (c) $f(x,y) = 9 - x^2 - y^2$, $\mathbf{c} = (2,-2)$
   (d) $f(x,y) = 3y^2 - x^2$, $\mathbf{c} = (1,-1)$

3. Suppose $A : \mathbb{R}^n \to \mathbb{R}$ is the best affine approximation to $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{c}$. Explain why $|\nabla f(\mathbf{c}) \cdot \mathbf{h}|$ is a good approximation for $|f(\mathbf{c} + \mathbf{h}) - f(\mathbf{c})|$ when $\|\mathbf{h}\|$ is small. That is, explain why $|\nabla f(\mathbf{c}) \cdot \mathbf{h}|$ is a good approximation for the error in approximating $f(\mathbf{c} + \mathbf{h})$ by $f(\mathbf{c})$.

4. Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x,y,z) = xyz$.
   (a) Find the best affine approximation to f at $(3,2,4)$.
   (b) Suppose x, y, and z represent the length, width, and height of a box. Suppose you measure the length to be $3 \pm h$ centimeters, the width to be $2 \pm h$ centimeters, and the height to be $4 \pm h$ centimeters. Use the best affine approximation from (a) to approximate the maximum error you would make in computing the volume of the box from these measurements.

5. A metal plate is heated so that its temperature at a point (x,y) is

$$T(x,y) = 50y^2 e^{-\frac{1}{5}(x^2+y^2)}.$$

   A bug moves along the ellipse parametrized by

$$\alpha(t) = (\cos(t), 2\sin(t)).$$

   Find the rate of change of temperature for the bug at times $t = 0$, $t = \frac{\pi}{4}$, and $t = \frac{\pi}{2}$.

6. Let x, y, and z be the length, width, and height, respectively, of a box. Suppose the box is increasing in size so that when $x = 3$ centimeters, $y = 2$ centimeters, and $z = 5$ centimeters, the length is increasing at rate of 2 centimeters per second, the width at a rate of 4 centimeters per second, and the height at a rate of 3 centimeters per second.
   (a) Find the rate of change of the volume of the box at this time.
   (b) Find the rate of change of the length of the diagonal of the box at this time.

7. Suppose $w = -\log(x^2 + y^2 + z^2)$ and $(x, y, z) = (4t, \sin(t), \cos(t)$. Find
$$\left.\frac{dw}{dt}\right|_{t=\frac{\pi}{3}}.$$

8. The kinetic energy K of an object of mass m moving in a straight line with velocity v is
$$K = \frac{1}{2}mv^2.$$
If, at time $t = t_0$, $m = 2000$ kilograms, $v = 50$ meters per second, m is decreasing at a rate of 2 kilograms per second, and v is increasing at a rate of 1.5 meters per second per second, find
$$\left.\frac{dK}{dt}\right|_{t=t_0}.$$

9. Each of the following equations specifies some curve in $\mathbb{R}^2$. In each case, find an equation for the line tangent to the curve at the given point $\mathbf{a}$.
   (a) $x^2 + y^2 = 5$, $\mathbf{a} = (2, 1)$
   (b) $2x^2 + 4y^2 = 18$, $\mathbf{a} = (1, -2)$
   (c) $y^2 - x = 0$, $\mathbf{a} = (4, -2)$
   (d) $y^2 - x^2 = 5$, $\mathbf{a} = (-2, 3)$

10. Each of the following equations specifies some surface in $\mathbb{R}^3$. In each case, find an equation for the plane tangent to the surface at the given point $\mathbf{a}$.
    (a) $x^2 + y^2 + z^2 = 14$, $\mathbf{a} = (2, 1, -3)$
    (b) $x^2 + 3y^2 + 2z^2 = 9$, $\mathbf{a} = (2, -1, 1)$
    (c) $x^2 + y^2 - z^2 = 1$, $\mathbf{a} = (1, 2, 2)$
    (d) $xyz = 6$, $\mathbf{a} = (1, 2, 3)$

11. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a, b), $f(a, b) = c$, and $\frac{\partial}{\partial y} f(a, b) \neq 0$. Let C be the level curve of f with equation $f(x, y) = c$. Show that
$$y = -\frac{\dfrac{\partial}{\partial x} f(a, b)}{\dfrac{\partial}{\partial y} f(a, b)}(x - a) + b$$
is an equation for the line tangent to C at (a, b).