Chapter 4
The Derivative

Abstract The derivative of a map is the linear term in its Taylor approximation; it is a map itself. Because linear approximations are simpler than those of higher order, and because linear maps are easier to visualize than nonlinear ones, the derivative is an especially important part of the study of maps. It gives us valuable local information. We study the derivative in this chapter, beginning with the familiar connection to tangents.

4.1 Differentiability

Differentiability of $y = f(x_1, \ldots, x_n)$

Analytically, a function $y = f(x)$ is differentiable at a point if a certain limit exists; geometrically, the graph of the function must have a tangent at that point. When there are several input variables, $y = f(x_1, \ldots, x_p)$, the geometric characterization is the same—the graph must have a tangent—but the analytic one becomes uncertain: Is it enough for the partial derivatives to exist, or must the directional derivatives exist in all directions, or is even more necessary? In this section, we introduce the derivative map to settle the question and to make a clear connection between the analytic and geometric aspects of differentiability.

Differentiability in terms of "little oh"

According to the usual definition, $y = f(x)$ is differentiable at $x = a$ if

$$\lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a)}{\Delta x} = f'(a),$$

for some finite number $f'(a)$ that we then call the derivative of f at a. In that case, we can rewrite the limit expression in the form

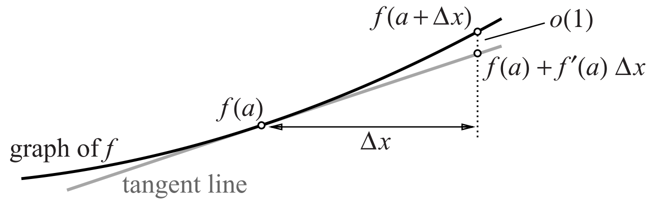
$$\lim_{\Delta x \to 0} \frac{f(a+\Delta x) - f(a) - f'(a)\,\Delta x}{\Delta x} = 0.$$

This says that the numerator, as a function of Δx, vanishes to order greater than 1 (cf. Definition 3.3, p. 85). In other words, the usual definition of differentiability is

J.J. Callahan, *Advanced Calculus: A Geometric View*, Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4419-7332-0_4,

equivalent to the following equality involving “little oh”:

$$\underbrace{f(a+\Delta x)}_{\substack{\text{values of } f\\ \text{near } a}} = \underbrace{f(a)+f'(a)\,\Delta x}_{\substack{\text{values along}\\ \text{tangent line at } a}} + o(1).$$



Differentiability and local linearity

We recognize $y = f(a) + f'(a)\Delta x$ as the formula for the tangent line at $x = a$, so the equation tells us what it means for the graph of f to have a tangent: the gap between the graph of f and its tangent line vanishes more rapidly than the horizontal displacement from the point of tangency. We can take this as the *geometric* definition of differentiability. Another name for differentiability, understood geometrically, is *local linearity*: under sufficiently high magnification (i.e., for Δx sufficiently small), the graph of f at $x = a$ is indistinguishable from the linear graph of the tangent there.

Comparison with Taylor’s theorem

Notice that our new geometric formula for differentiability is similar to Taylor’s formula,

$$f(a+\Delta x) = f(a) + f'(a)\Delta x + O(2).$$

The difference lies solely in the order of vanishing of the remainder; Taylor’s formula has the stronger condition $R_{1,a}(\Delta x) = O(2)$. (For example, $t^{4/3} = o(1)$ but $t^{4/3} \neq O(2)$; see also Exercise 3.15, p. 102.) But the hypothesis that Taylor’s formula rests upon is stronger, too: Taylor’s theorem requires that f have a continuous second derivative on an open interval that contains a and $a + \Delta x$. However, as we have seen, the limit defining the derivative leads us to the formula

$$f(a+\Delta x) = f(a) + f'(a)\Delta x + o(1)$$

that involves “little oh” rather than “big oh.”

Differentiability for $z = f(x,y)$

Let us move on to the differentiability of $z = f(x,y)$ at $(x,y) = (a,b)$, and approach it from the geometric point of view. In terms of coordinates $\Delta x = x - a$, $\Delta y = y - b$ centered at (a,b), an arbitrary plane has a formula that we can write as

$$z = c + p\,\Delta x + q\,\Delta y.$$

We require the gap $f(a+\Delta x, b+\Delta y) - (c + p\,\Delta x + q\,\Delta y)$ to vanish more rapidly than the horizontal displacement $\sqrt{(\Delta x)^2 + (\Delta y)^2}$ to the point of tangency (a,b).

Definition 4.1 *The function $z = f(x,y)$ is* ***differentiable****, or* ***locally linear****, at $(x,y) = (a,b)$ if there are constants c, p, and q for which*

$$f(a+\Delta x, b+\Delta y) - (c + p\,\Delta x + q\,\Delta y) = o(1).$$

In that case, the graph of $z = c + p\,\Delta x + q\,\Delta y$ is the ***tangent plane*** *to the graph of f at the point (a,b).*

Theorem 4.1. *If $z = f(x,y)$ is differentiable at (a,b), then both partial derivatives exist at (a,b) and the equation of the tangent plane there is*

$$z = f(a,b) + f_x(a,b)\,\Delta x + f_y(a,b)\,\Delta y.$$

Proof. In terms of the definition, we must show

$$c = f(a,b), \quad p = \frac{\partial f}{\partial x}(a,b), \quad q = \frac{\partial f}{\partial y}(a,b);$$

in particular, we must show that the two partial derivatives exist. For a start, the expression

$$f(a+\Delta x, b+\Delta y) - (c + p\,\Delta x + q\,\Delta y)$$

must vanish when $\Delta x = \Delta y = 0$. This implies $c = f(a,b)$. Now keep $\Delta y = 0$ but let Δx vary. The hypothesis then becomes

$$f(a+\Delta x, b) - (f(a,b) + p\,\Delta x) = o(1),$$

and it means (cf. Definition 3.3, p. 85)

$$0 = \lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a,b) - p\,\Delta x}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a,b)}{\Delta x} - p.$$

Therefore

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

that is, the partial derivative exists and has the value p. The value of q is determined in a similar way, by fixing $\Delta x = 0$ and letting Δy vary. □

Derivative of f

The partial derivatives that define the tangent plane are, at the same time, the components of the 1×2 matrix that defines the derivative of f at (a,b) (Definition 3.16, p. 99). The following corollary makes explicit this (natural!) connection between differentiability and the derivative.

Corollary 4.2 *If $z = f(x,y)$ is differentiable at (a,b), then the derivative $\mathrm{d}f_{(a,b)} : \mathbb{R}^2 \to \mathbb{R}$ exists and*

$$f(a+\Delta x, b+\Delta y) = f(a,b) + \mathrm{d}f_{(a,b)}(\Delta x, \Delta y) + o(1). \qquad \square$$

Do partial derivatives imply differentiability?

A reasonable question to ask at this point is: if both partial derivatives of $f(x,y)$ exist at (a,b), is f then differentiable at (a,b)? In other words, if the plane

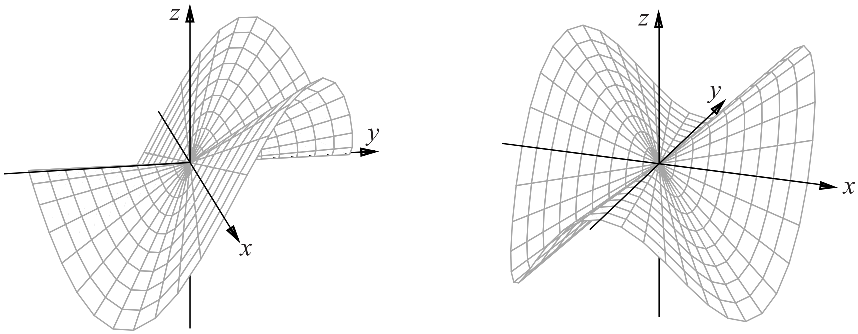
$$z = f(a,b) + f_x(a,b)\,\Delta x + f_y(a,b)\,\Delta y$$

is defined, is it not automatically the tangent plane to the graph of f at (a,b)? Is it not guaranteed that the gap between this plane and the graph of f must vanish to higher order than the horizontal displacement $(\Delta x, \Delta y)$ from the point (a,b)?

Counterexample: the "manta ray"

It fact, the answer is *no*. Here is an example that illustrates the contrary (a *counterexample*):

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



A bundle of lines through the origin

The two figures are different views of the graph of $z = f(x,y)$; it looks vaguely like a manta ray swimming along the y-axis. The essential thing to note is that the graph is made up of a bundle of straight lines through the origin. An easy way to confirm this is to put a polar coordinate overlay on the graph. That is, let $x = r\cos\theta$, $y = r\sin\theta$; then, away from the origin,

$$z = f(x,y) = \frac{r^3 \cos^2\theta \sin\theta}{r^2} = r\cos^2\theta\sin\theta.$$

When θ is fixed (as it is along a radial line), this is the straight line $z = mr$ of slope $m = \cos^2\theta\sin\theta$. The *Mathematica* 5 code that produces the figures makes use of this overlay:

```
ParametricPlot3D[{r Cos[t], r Sin[t], r Cos[t]^2 Sin[t]},
   {r, 0, 1}, {t, 0, 2 Pi}, PlotPoints -> {10, 49},
   BoxRatios -> {1, 1, .8}, Axes -> False, Boxed -> False,
   ViewPoint -> {3.500, -0.680, 1.221}]
```

No plane is tangent to the graph at the origin

Let us now see why no plane is tangent to the graph at the origin. Along the radial line $\theta = \theta_0$, the graph is the straight line of slope $m = \cos^2\theta_0 \sin\theta_0$. If θ_0 is an integer multiple of $\pi/2$, then $m = 0$ and the radial line lies along an axis. So this slope is a partial derivative; we find $f_x(0,0) = f_y(0,0) = 0$. Therefore, if the graph were to have a tangent at the origin, Theorem 4.1 would force it to be the (x,y)-plane itself: $z = 0$. In that case, the slope of the graph in any direction at the origin would be 0. But the figures make it clear (and the formula $m = \cos^2\theta_0\sin\theta_0$ confirms) that the slope of the radial line in the direction $\theta_0 \neq k\pi/2$ is nonzero. The manta ray graph has no tangent plane at the origin; the function f is not differentiable there.

Directional derivatives

So the mere existence of the partial derivatives of $z = f(x,y)$ at a point is not enough to guarantee that f is differentiable at that point, i.e., that its graph has a tangent plane there. But suppose we impose a stronger condition, one requiring that all directional derivatives exist. We recall the definition of a directional derivative for a function of p variables (where p need not equal 2).

Definition 4.2 *Let* $\mathbf{u}$ *be a unit vector; then the* ***directional derivative*** *of* $z = f(\mathbf{x})$ *at the point* $\mathbf{x} = \mathbf{a}$ *in the direction* $\mathbf{u}$ *is*

$$D_{\mathbf{u}} f(\mathbf{a}) = \left.\frac{d}{dt} f(\mathbf{a} + t\mathbf{u})\right|_{t=0},$$

when the expression on the right exists.

Let $\mathbf{u} = \mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (i.e., 1 in the ith place, 0 elsewhere); the derivative in the direction $\mathbf{e}_i$ is just the usual **partial derivative**:

$$D_{\mathbf{e}_i} f(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

Do directional derivatives imply differentiability?

Suppose we now require that the directional derivatives of $z = f(x,y)$ in all directions (and not just the axis directions) exist at a point. Will this stronger condition guarantee that the graph of f has a tangent at that point? In fact, the manta ray is still a counterexample. In the direction $\mathbf{u} = (\cos\theta, \sin\theta)$, the directional derivative of f at the origin is

$$D_{(\cos\theta, \sin\theta)} f(0,0) = \cos^2\theta \sin\theta.$$

(In the given direction, the graph of f is a straight line of slope $\cos^2\theta \sin\theta$.) Thus, even though all the directional derivatives of f exist at the origin, there is (still) no tangent plane. The existence of all directional derivatives of f at a point is not enough to guarantee that f is differentiable there.

Directional derivatives from the derivative

Although the existence of directional derivatives does not guarantee differentiability, the converse is true, according to the following theorem.

Theorem 4.3. *If* $z = f(x,y)$ *is differentiable at* (a,b)*, then all directional derivatives exist at* (a,b)*. In fact,* $D_{(\alpha,\beta)} f(a,b) = \mathrm{d}f_{(a,b)}(\alpha,\beta)$.

Proof. The proof is probably easier to follow in vector notation. We set $(\alpha,\beta) = \mathbf{u}$; then, by definition, the directional derivative is

$$D_{\mathbf{u}} f(\mathbf{a}) = \lim_{t\to 0} \frac{f(\mathbf{a} + t\mathbf{u}) - f(\mathbf{a})}{t} = \lim_{t\to 0} \frac{\mathrm{d}f_{\mathbf{a}}(t\mathbf{u}) + o(1)}{t}.$$

We have $f(\mathbf{a} + t\mathbf{u}) - f(\mathbf{a}) = \mathrm{d}f_{\mathbf{a}}(t\mathbf{u}) + o(1)$ because f is differentiable at $\mathbf{a}$. But then

$$\lim_{t\to 0} \frac{\mathrm{d}f_{\mathbf{a}}(t\mathbf{u}) + o(1)}{t} = \lim_{t\to 0} \frac{t\,\mathrm{d}f_{\mathbf{a}}(\mathbf{u})}{t} + \lim_{t\to 0} \frac{o(1)}{t} = \mathrm{d}f_{\mathbf{a}}(\mathbf{u}) + 0 = \mathrm{d}f_{\mathbf{a}}(\mathbf{u}).$$

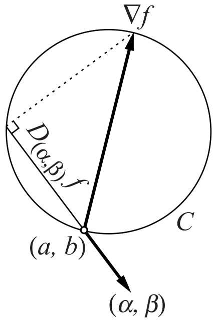
We have used the linearity of $\mathrm{d}f_{\mathbf{a}}$ to write $\mathrm{d}f_{\mathbf{a}}(t\mathbf{u}) = t\,\mathrm{d}f_{\mathbf{a}}(\mathbf{u})$. Also, $o(1)/t \to 0$ as $t \to 0$, by definition of $o(1)$. $\square$

The gradient vector

The gradient of $z = f(x,y)$ at (a,b) is the *vector* $\operatorname{grad} f(a,b) = \nabla f(a,b)$ in $\mathbb{R}^2$ whose components are $f_x(a,b)$ and $f_y(a,b)$. Of course these are, at the same time, the components of the matrix that represents the derivative $\mathrm{d}f_{(a,b)} : \mathbb{R}^2 \to \mathbb{R}^1$. Conceptually, $\nabla f(a,b)$ and $\mathrm{d}f_{(a,b)}$ are different; the first is a vector, the second is a linear map. However, because matrix multiplication involves the scalar (dot) product, the two are connected:

$$\mathrm{d}f_{(a,b)}(\alpha,\beta) = \big(f_x(a,b)\ f_y(a,b)\big)\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \nabla f(a,b)\cdot(\alpha,\beta).$$

This connection allows us to express the previous theorem in a way that is probably more familiar.



Corollary 4.4 *If $z = f(x,y)$ is differentiable at (a,b), then*

$$D_{(\alpha,\beta)}f(a,b) = \nabla f(a,b)\cdot(\alpha,\beta). \qquad \square$$

Let C be the circle that has the vector $\nabla f(a,b)$ as diameter. Then, because the scalar $\nabla f(a,b)\cdot(\alpha,\beta)$ is the perpendicular projection of $\nabla f(a,b)$ on the line in the direction of (α,β), we can realize $D_{(\alpha,\beta)}f$ as the length of the chord of C that lies in the direction (α,β). If (α,β) makes an obtuse angle with ∇f, extend $-(\alpha,\beta)$ and note that then $D_{(\alpha,\beta)}f \le 0$.

The hypothesis that f is differentiable is crucial in the corollary. If we merely know that all the directional derivatives exist (including the partial derivatives), we cannot conclude that $D_{(\alpha,\beta)}f(a,b) = \nabla f(a,b)\cdot(\alpha,\beta)$. The manta ray at the origin is once again a counterexample: we have

$$D_{(\alpha,\beta)}f(0,0) = \alpha^2\beta \ \text{ but } \ \nabla f(0,0)\cdot(\alpha,\beta) = (0,0)\cdot(\alpha,\beta) = 0.$$

Local linearity with level curves

Because we used tangents to define differentiability, it is natural to use tangents to illustrate local linearity. For example, consider the function $f(x,y) = x^2 - y^2$ at the point $(a,b) = (2,-1)$. The graph of $z = f(x,y)$ is a curved surface in $\mathbb{R}^3$ and the graph of the derivative

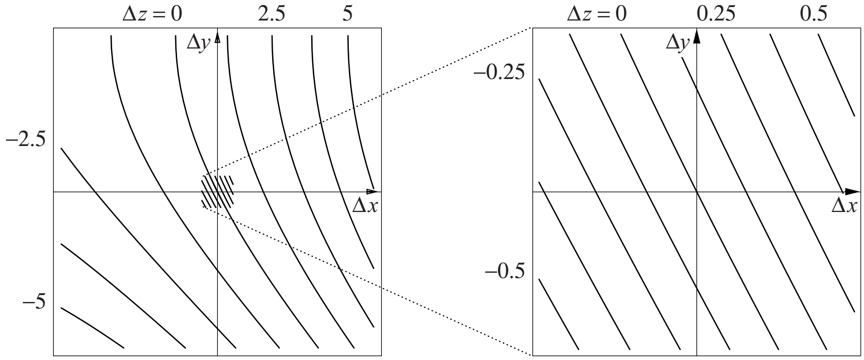
$$\mathrm{d}f_{(2,-1)}(\Delta x,\Delta y) = f_x(2,-1)\Delta x + f_y(2,-1)\Delta y = 4\Delta x + 2\Delta y$$

is a plane. We expect that, under sufficient magnification at the point $(x,y,z) = (2,-1,f(2,-1))$ in $\mathbb{R}^3$, the graph of f will become indistinguishable from the tangent plane; cf. Exercise 4.2.

We, however, take a different approach, comparing instead the level sets of f and $\mathrm{d}f_{(2,-1)}$ in windows centered at $(2,-1)$ in the (x,y)-plane. The window on the left below shows level curves

$$f(2+\Delta x, -1+\Delta y) - f(2,-1) = \Delta z, \quad \Delta z = -5, -3.75, \dots, 5,$$

in the square $1 \leq x \leq 3$, $-2 \leq y \leq 0$. By design, the level curve $\Delta z = 0$ passes through the origin of the $(\Delta x, \Delta y)$-window. Obviously f is nonlinear: its level sets are curved, and they are unequally spaced.



Local linearity emerges

Under tenfold magnification (see the right window; spacing between levels has been cut by the same factor of 10), the local linearity of f begins to emerge. The level curves of f are now essentially straight, parallel, and evenly spaced: the hallmarks of a linear function. Thus, at this magnification, f looks linear. We must now just check that the apparently linear function we see in this window agrees with the derivative,

$$df_{(2,-1)}(\Delta x, \Delta y) = 4\Delta x + 2\Delta y.$$

Comparing level curves of f and $df_{(2,-1)}$

The level curves $4\Delta x + 2\Delta y = C$ of the derivative are parallel straight lines with common slope $\Delta y/\Delta x = -2$, just as in the microscope window on the right. But this is not yet enough; we need to show that a given line represents the same level for f and for $df_{(2,-1)}$. This is made easier by the fact that f itself has a simple formula:

$$\begin{aligned} z &= f(2+\Delta x, -1+\Delta y) \\ &= (2+\Delta x)^2 - (-1+\Delta y)^2 = 4 + 4\Delta x + (\Delta x)^2 - 1 + 2\Delta y - (\Delta y)^2 \\ &= 3 + df_{(2,-1)}(\Delta x, \Delta y) + (\Delta x)^2 - (\Delta y)^2. \end{aligned}$$

Along the diagonals of the window, $(\Delta x)^2 = (\Delta y)^2$, so

$$\Delta z = f(2+\Delta x, -1+\Delta y) - 3 = df_{(2,-1)}(\Delta x, \Delta y);$$

in other words, f and its derivative agree exactly on the diagonals. It follows that, everywhere in the right window, a given level curve for f is indistinguishable from the level curve for $df_{(2,-1)}$ at the same level.

4.2 Maps of the plane

Visualizing maps $\mathbf{f}: \mathbb{R}^2 \to \mathbb{R}^2$

Our goal here is to understand what differentiability means geometrically for a map of the plane. Suppose $\mathbf{f}: U^2 \to \mathbb{R}^2$ has the coordinate form

$$\mathbf{f}: \begin{cases} x = f(u,v), \\ y = g(u,v). \end{cases}$$

Here U^2 is a *window* of the form $|u-a| < p$, $|v-b| < q$ (or, more generally, an *open set* in $\mathbb{R}^2$; cf. Definition 8.4, p. 277). The derivative of $\mathbf{f}$ at the point $\mathbf{a} = (a,b)$ is the linear map $d\mathbf{f}_{\mathbf{a}} : \mathbb{R}^2 \to \mathbb{R}^2$ whose coordinate matrix is

$$d\mathbf{f}_{(a,b)} = \begin{pmatrix} f_u(a,b) & f_v(a,b) \\ g_u(a,b) & g_v(a,b) \end{pmatrix},$$

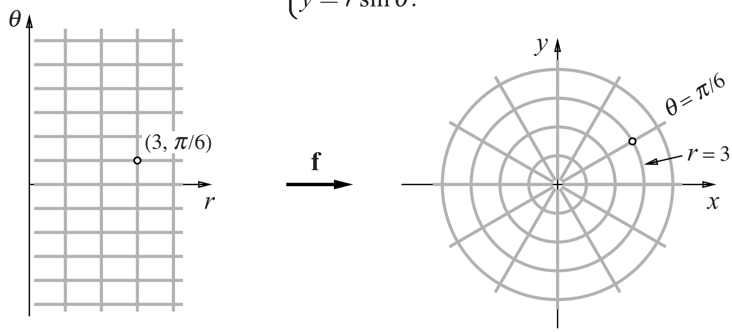
(Definition 3.16, p. 99). The graph of $\mathbf{f}$ and the graph of $d\mathbf{f}_{(a,b)}$ are both 2-dimensional surfaces, but they lie in the 4-dimensional (u,v,x,y)-space, so we cannot visualize them directly. In particular, we cannot see how—or whether—the graph of the derivative is tangent to the graph of the map.

We faced this dimension problem with the graph of a linear map of the plane in Chapter 2. There, we solved the problem by looking at images instead of graphs; we do the same here. Differentiability is then manifested as local linearity: we compare the image of $\mathbf{f}$ in a microscope window centered at $\mathbf{a}$ to the image of the linear map $d\mathbf{f}_{\mathbf{a}}$ in that window.

First example: polar coordinates

The polar coordinate change is the map that pulls back Cartesian coordinates to polar coordinates:

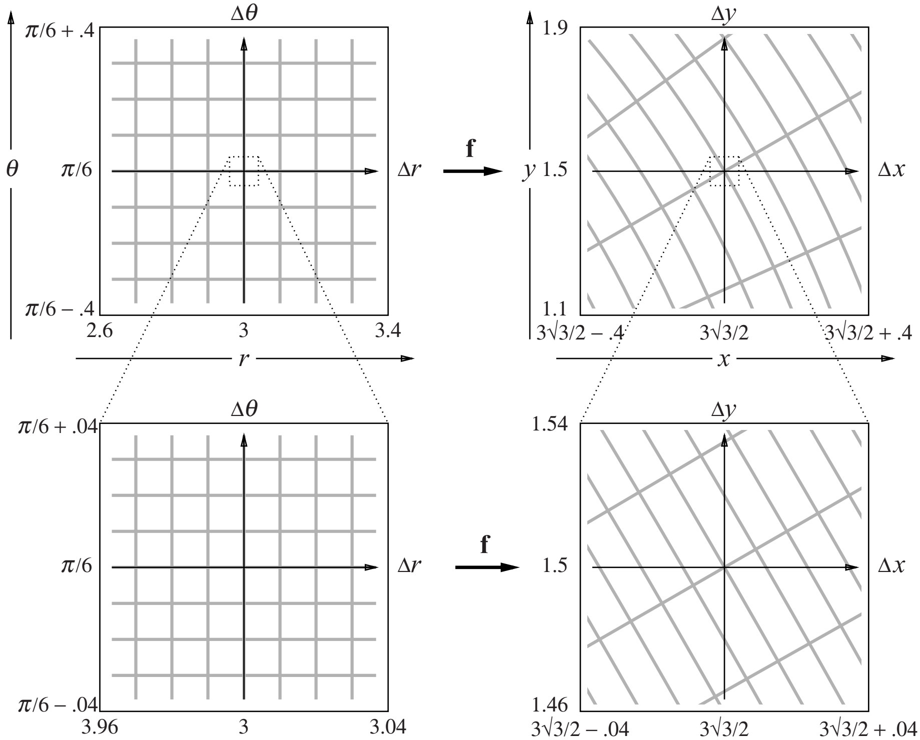
$$\mathbf{f}: \begin{cases} x = r\cos\theta, \\ y = r\sin\theta. \end{cases}$$



The map $\mathbf{f}$ puts a grid of rays and concentric circles on the (x,y)-plane, corresponding to the rectangular grid $\theta =$ constant and $r =$ constant in the (r,θ)-plane itself. By convention, only the positive half of each ray is used; that is, we assume $r > 0$. (In other words, the domain U^2 for $\mathbf{f}$ is the open right half-plane.) Sometimes it is useful to allow $r = 0$, as well. This is the θ-axis; the map $\mathbf{f}$ collapses it to a single point, the origin in the target.

Local behavior of **f** near $(r,\theta) = (3,\pi/6)$

According to Taylor's theorem, any smooth map is approximately a polynomial when its domain is restricted to a small enough region. The closeness of the approximation is directly related to the degree of the polynomial, but even a linear polynomial can provide an impressive approximation. Let us focus on the point $(r,\theta) = (3,\pi/6)$ and see how **f** becomes approximately linear as it acts on smaller and smaller regions centered at this point. In the process we also see that the approximation is precisely the linear term in the Taylor polynomial of **f**, the map that we call the *derivative* of **f** at $(r,\theta) = (3,\pi/6)$ and denote with the symbol $d\mathbf{f}_{(3,\pi/6)}$.



The figure above shows what **f** does to a a grid of squares in a small window centered at the point $(r,\theta) = (3,\pi/6)$. The image is a grid of radial lines and circular arcs in another small window centered at the image point $(x,y) = (3\sqrt{3}/2, 3/2)$. To describe the action of **f** in these windows it is natural to use coordinates Δr, $\Delta\theta$, Δx, and Δy that measure displacements from the center of each window:

$$\Delta r = r - 3, \qquad \Delta x = x - 3\sqrt{3}/2,$$
$$\Delta\theta = \theta - \pi/6, \qquad \Delta y = y - 3/2.$$

Views of **f** in two "microscope" windows

In the figure, there are two pairs of windows at different levels of magnification. In the upper pair the grid spacing is 0.1 units and the windows themselves measure 0.8 units on a side. The radial lines we see in the image window therefore have a separation of $\Delta\theta = 0.1$ radians and the concentric circular arcs have a separation of $\Delta r = 0.1$ units. At this level of magnification the arcs are still noticeably curved.

Each lower window is a tenfold magnification of the center of the window above it. The radial lines in the image are now only $\Delta\theta = 0.01$ radians apart; they look nearly parallel. The concentric circular arcs are spaced $\Delta r = 0.01$ units apart and likewise appear to be straight and parallel. In this microscopic view, **f** looks like a linear map, because it maps a grid of congruent squares to a grid of (nearly) congruent parallelograms, rectangles, in fact.

Near $(3, \pi/6)$, **f** is a stretch and rotation

Can we describe **f** in the lower windows in the fashion of a linear map? The line $\Delta\theta = 0$ in the source (the Δr-axis) is just the horizontal line $\theta = \pi/6$, so its image is the radial line that makes an angle of $\pi/6$ radians, or 30°, as it passes through the origin $(\Delta x, \Delta y) = (0,0)$ of the microscope window in the target. In other words, **f** rotates the Δr-axis by 30°; the figure makes it clear that the whole $(\Delta r, \Delta\theta)$-plane undergoes essentially the same rotation. In addition, before **f** rotates the plane it stretches it vertically; by eye, the stretch factor appears to be about 3.

The derivative $\mathbf{df}_{(r,\theta)}$

Now compare this action with the action of the derivative $\mathbf{df}_{(3,\pi/6)}$. At an arbitrary point $(r,\theta) = (r_0, \theta_0)$, the derivative $d\mathbf{f}_{(r_0,\theta_0)}$ is defined (see p. 112) to be the linear map

$$\mathbf{df}_{(r_0,\theta_0)} : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{pmatrix} \Delta r \\ \Delta\theta \end{pmatrix} \mapsto \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

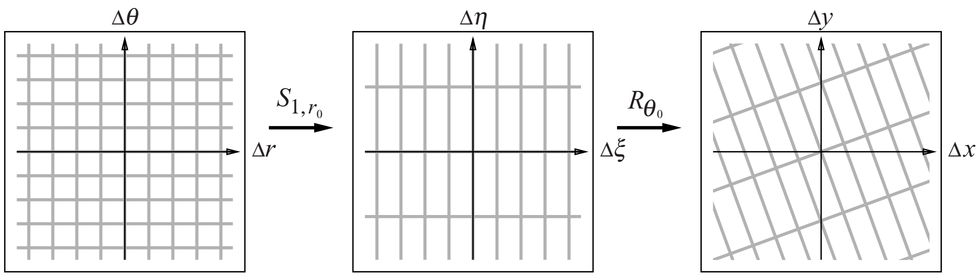
whose matrix is

$$\mathbf{df}_{(r_0,\theta_0)} = \begin{pmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{pmatrix}\Bigg|_{(r,\theta)=(r_0,\theta_0)} = \begin{pmatrix} \cos\theta_0 & -r_0\sin\theta_0 \\ \sin\theta_0 & r_0\cos\theta_0 \end{pmatrix}.$$

The derivative splits into two factors

Notice that $\mathbf{df}_{(r_0,\theta_0)}$ factors neatly into a pair of matrices, a stretch (or *strain*) S_{1,r_0}, followed by a rotation R_{θ_0}:

$$\begin{pmatrix} \cos\theta_0 & -r_0\sin\theta_0 \\ \sin\theta_0 & r_0\cos\theta_0 \end{pmatrix} = \underbrace{\begin{pmatrix} \cos\theta_0 & -\sin\theta_0 \\ \sin\theta_0 & \cos\theta_0 \end{pmatrix}}_{R_{\theta_0}} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & r_0 \end{pmatrix}}_{S_{1,r_0}}.$$



Factoring $\mathbf{df}_{(r_0,\theta_0)}$ means that we can describe its effect in two stages. First, S_{1,r_0} stretches the plane vertically (i.e., in the direction of the $\Delta\theta$-axis) by the factor r_0; then R_{θ_0} rotates the result by θ_0 radians. (The coordinate names $\Delta\xi$, $\Delta\eta$ in the intermediate window are just arbitrary choices.)

In particular, we can now conclude that $d\mathbf{f}_{(3,\pi/6)}$ is a threefold vertical stretch followed by a 30° rotation. But this is exactly how $\mathbf{f}$ itself behaves in a microscope window centered at $(r,\theta) = (3,\pi/6)$. This suggests we say that $\mathbf{f}$ is *locally linear* in a small neighborhood of $(3,\pi/6)$.

$\mathbf{f} \approx d\mathbf{f}$ near $(3,\pi/6)$

The example is leading us to say that a map $\mathbf{f} : U^2 \to \mathbb{R}^2 : \mathbf{u} \mapsto \mathbf{x}$ will be locally linear—or differentiable—at a point $\mathbf{u} = \mathbf{a}$ if

$$\Delta\mathbf{x} = \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) - \mathbf{f}(\mathbf{a})$$

differs from a linear function of $\Delta\mathbf{u}$ by an amount that vanishes faster than $\Delta\mathbf{u}$. Here is a precise definition.

Differentiability and local linearity

Definition 4.3 *The map* $\mathbf{x} = \mathbf{f}(\mathbf{u})$ *is **differentiable**, or **locally linear**, at* $\mathbf{u} = \mathbf{a}$ *if there is a linear map* $\mathbf{L} : \mathbb{R}^2 \to \mathbb{R}^2$, *called the **derivative** of* $\mathbf{f}$ *at* $\mathbf{a}$, *for which*

$$\mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) = \mathbf{f}(\mathbf{a}) + \mathbf{L}(\Delta\mathbf{u}) + \boldsymbol{o}(1).$$

Theorem 4.5. *If* $\mathbf{f} : U^2 \to \mathbb{R}^2$ *is differentiable at* $\mathbf{u} = \mathbf{a}$, *then* $\mathbf{L} = d\mathbf{f}_{\mathbf{a}}$. *In particular, all the partial derivatives appearing in the matrix* $d\mathbf{f}_{\mathbf{a}}$ *exist.*

Proof. See Chapter 4.4, where the theorem is restated (as Theorem 4.6) for the general case $\mathbf{f} : U^p \to \mathbb{R}^q$. □

The microscope equation

The theorem makes it clear that if $\mathbf{f}$ is locally linear at $\mathbf{a}$, then its linear approximation is its derivative $d\mathbf{f}_{\mathbf{a}}$. Note that if we rewrite the **window equation**

$$\Delta\mathbf{x} = \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) - \mathbf{f}(\mathbf{a}) = d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u}) + \boldsymbol{o}(1)$$

without the remainder term $\boldsymbol{o}(1)$, we get an approximation that is, in effect, a new form of the **microscope equation**:

$$\Delta\mathbf{x} \approx d\mathbf{f}_{\mathbf{u}_0}(\Delta\mathbf{u}).$$

In other words, the microscope equation emerges as a (rather condensed) way of expressing the differentiability or local linearity of a map. We have already noted the connection between the microscope equation and Taylor's theorem in Chapter 3 (p. 83; p. 95).

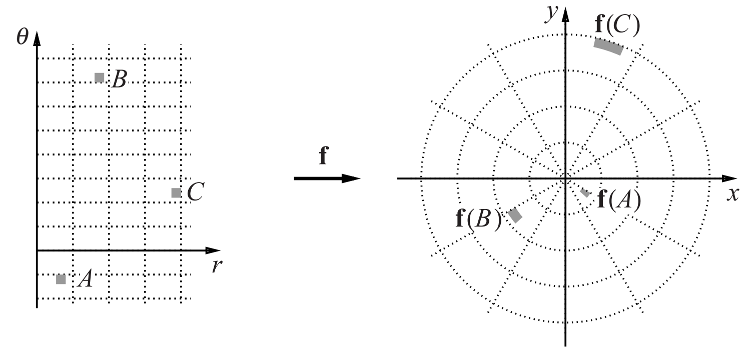
Local area multiplier

Definition 4.4 *If* $\mathbf{f} : U^2 \to \mathbb{R}^2$ *is differentiable at* $\mathbf{a}$, *its **local area multiplier** at* $\mathbf{a}$ *is the area multiplier of its derivative* $d\mathbf{f}_{\mathbf{a}}$.

For the polar coordinate map $\mathbf{x} = \mathbf{f}(\mathbf{r})$, we find that the area multiplier of $d\mathbf{f}_{\mathbf{r}_0}$ is r_0:

$$\det d\mathbf{f}_{\mathbf{r}_0} = \det\begin{pmatrix} \cos\theta_0 & -r_0\sin\theta_0 \\ \sin\theta_0 & r_0\cos\theta_0 \end{pmatrix} = r_0\cos^2\theta_0 + r_0\sin^2\theta_0 = r_0.$$

Thus we say that r_0 is the **local area multiplier** for $\mathbf{f}$ itself at the point $\mathbf{r}_0 = (r_0, \theta_0)$. It is evident in the figure below that the local area multiplier of $\mathbf{f}$ varies from point to point and increases with the radius r; our calculations show that the local multiplier is exactly r.



Local linearity *versus* "looking linear locally"

For plane maps $\mathbf{f}$ like the polar coordinate map, we have used the notions

$$\mathbf{f} \textit{ is arbitrarily close to } d\mathbf{f}_{\mathbf{a}} \textit{ near } \mathbf{a} \quad \text{and} \quad \mathbf{f} \textit{ "looks like" } d\mathbf{f}_{\mathbf{a}} \textit{ near } \mathbf{a}$$

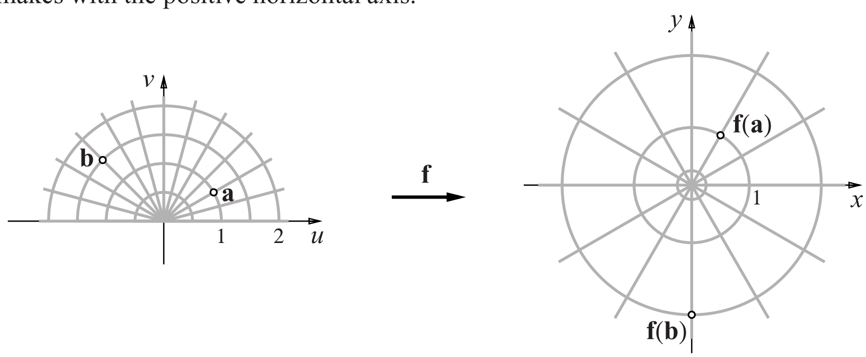
more or less interchangeably. However, the two are subtly different. The first, of course, is what we now call *local linearity*. The second, however, may not be true if $d\mathbf{f}_{\mathbf{a}}$ fails to be invertible; it is a stronger condition. To help bring into sharper focus the distinction between these two notions—and to see how the second condition can fail—we analyze a second map.

Second example: a quadratic map

Consider the quadratic map $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, defined by the equations

$$\mathbf{f} : \begin{cases} x = u^2 - v^2, \\ y = 2uv. \end{cases}$$

Although the action of the polar coordinate change map was immediately evident on a global level (i.e., on the entire right half-plane $r > 0$), the same is not true for the quadratic map $\mathbf{f}$. However, the action is not hard to describe; we now show $\mathbf{f}$ squares the distance of any point from the origin and doubles the angle that point makes with the positive horizontal axis.



Polar coordinate overlays

We can show that $\mathbf{f}$ acts this way by translating our formulas for $\mathbf{f}$ into polar coordinates, because they provide the angles and distances we wish to measure.

That is, think of polar coordinates (r,θ) as an "overlay" on the (x,y)-plane, and introduce the same overlay on the (u,v)-plane using new polar coordinates (ρ,φ): $u=\rho\cos\varphi$, $v=\rho\sin\varphi$. Then the formulas for $\mathbf{f}$ that define x and y in terms of u and v translate into expressions for r and θ in terms of ρ and φ:

$$\begin{aligned} r\cos\theta &= x = u^2-v^2 = \rho^2\cos^2\varphi-\rho^2\sin^2\varphi = \rho^2\cos 2\varphi,\\ r\sin\theta &= y = 2uv = 2\rho\cos\varphi\cdot\rho\sin\varphi = \rho^2\sin 2\varphi. \end{aligned}$$

In other words, $r\cos\theta=\rho^2\cos 2\varphi$ and $r\sin\theta=\rho^2\sin 2\varphi$, so

$$r=\rho^2 \text{ and } \theta=2\varphi.$$

f doubles angles and squares distances from the origin

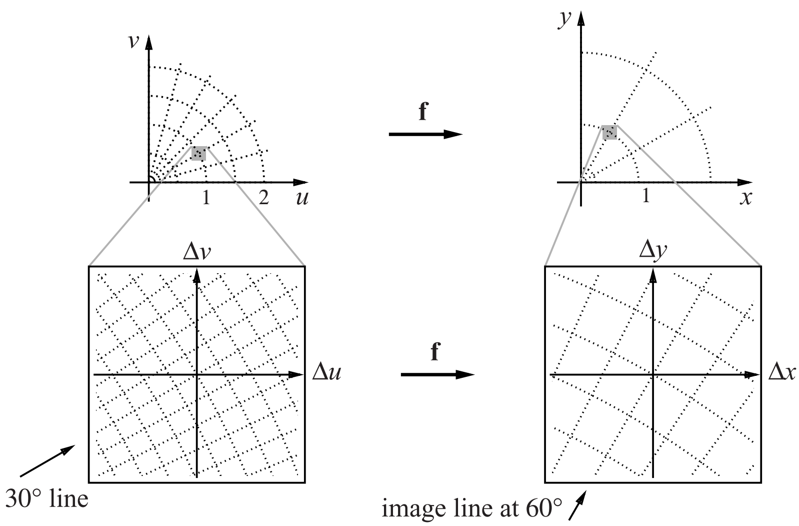
Thus, in terms of the polar coordinate overlays on the source and target, $\mathbf{f}$ squares the distance of a point from the origin ($r=\rho^2$) and it doubles the angle that that point makes with the horizontal ($\theta=2\varphi$).

f is a "double cover"

The angle-doubling means $\mathbf{f}$ "fans out" the upper half-plane $v\geq 0$ in the source to cover the entire target (x,y)-plane. The lower half-plane $v\leq 0$ also covers the entire (x,y)-plane, so the source covers the target twice, except for the origin. More precisely, let V^2 be the plane minus the origin: $V^2=\mathbb{R}^2\setminus(0,0)$. Then $\mathbf{f}:V^2\to V^2$ is a 2–1 map. Every point in the target V^2 is the image of exactly two points in the source V^2 (that lie 180° apart at the same distance from the origin). The unit circle maps to itself. A concentric circle inside the unit circle maps to another one even closer to the origin; one outside the unit circle is mapped to another farther from the origin.

Behavior of **f** near $(u,v)=(\sqrt{3}/2,1/2)$

With this clear picture of the global behavior of $\mathbf{f}$, it is easy to analyze its local behavior near a given point. For example, take the point $(u,v)=(\sqrt{3}/2,1/2)$. Its image is $(x,y)=(1/2,\sqrt{3}/2)$. These points are 1 unit from the origin ($\rho=r=1$) and make angles of $\varphi=30°=\pi/6$ radians and $\theta=60°=\pi/3$ radians, respectively.



The figure above uses the polar coordinate overlays to show how **f** acts in microscope windows centered at these two points. In the polar grid in the source window, the spacing between adjacent circular arcs is $\Delta\rho = 1/36 \approx 0.028$ units; the spacing between radial line segments is $\Delta\varphi = 1.5^\circ \approx 0.026$ radians. Because these numbers are nearly equal and relatively small, the grid looks approximately square.

Rotation and linear expansion

The target window shows the image of the source grid under the map **f**. The 30°-line from the source gives us our bearings. At the macroscopic level, **f** maps it to the 60°-line, so in the microscope window **f** rotates it by 30°. The entire grid is carried along by this action, so **f** rotates all points in the source by 30°, more or less. Obviously, there is also linear expansion we must take into account. The image grid is still approximately square; thus **f** must be close to a uniform dilation. Moreover, a single square in the image grid is about the size of a 2×2 square in the source grid, so the linear expansion factor is about 2 (and the area expansion factor is about 4). Therefore, in the microscope windows it appears that **f** approximately doubles all lengths and rotates points by about 30°, or $\pi/6$ radians. In other words, **f** approximates the linear map $2R_{\pi/6}$ in a small neighborhood of $(u,v) = (\sqrt{3}/2, 1/2)$.

The derivative of **f** at $(\sqrt{3}/2, 1/2)$

We expect, therefore, that the derivative of **f** at $(\sqrt{3}/2, 1/2)$ must equal $2R_{\pi/6}$. Can we confirm this? First of all, at an arbitrary point $(u,v) = (a,b)$, the derivative of **f** is given by the matrix

$$\mathbf{df}_{(a,b)} = \begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix}\Bigg|_{(u,v)=(a,b)} = \begin{pmatrix} 2a & -2b \\ 2b & 2a \end{pmatrix} = 2\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Therefore,

$$\mathbf{df}_{(\sqrt{3}/2,1/2)} = 2\begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix} = 2\begin{pmatrix} \cos\pi/6 & -\sin\pi/6 \\ \sin\pi/6 & \cos\pi/6 \end{pmatrix} = 2R_{\pi/6},$$

so $\mathbf{df}_{(\sqrt{3}/21/2)}$ is indeed the local linear approximation to **f** at $(\sqrt{3}/2, 1/2)$.

$\mathbf{df}_{(a,b)}$ is a rotation–dilation matrix

Before we consider the local behavior of **f** at a second point, we pause to note that our formula, above, for the derivative $\mathbf{df}_{(a,b)}$ shows that it is a rotation–dilation matrix (cf. p. 39 ff.). We exclude the special case $(a,b) = (0,0)$, where $\mathbf{df}_{(0,0)}$ is just the zero matrix. Thus, assuming $(a,b) \neq (0,0)$,

$$\mathbf{df}_{(a,b)} = 2\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = 2\sqrt{a^2+b^2}\, R_{\arctan(b/a)}.$$

This says that $\mathbf{df}_{(a,b)}$ is rotation by $\theta = \arctan(b/a)$ followed by a uniform linear dilation by the factor $2\sqrt{a^2+b^2}$. The local area multiplier for **f** at (a,b) is therefore $4(a^2+b^2)$.

Conformal maps

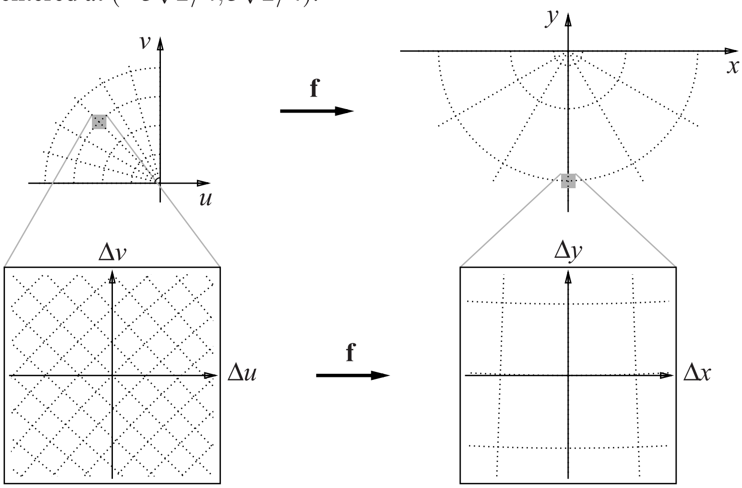
In Euclidean geometry, a rotation–dilation matrix such as $\mathbf{df}_{(a,b)}$, with $(a,b) \neq (0,0)$, is also known as a *similarity transformation*. Even when $\mathbf{df}_{(a,b)}$ alters lengths (i.e., when $2\sqrt{a^2+b^2} \neq 1$), angles remain unchanged.; therefore, a plane figure and its image under $\mathbf{df}_{(a,b)}$ are *similar*. A map such as **f** whose derivative is a similarity at each point in an open region is said to be **conformal** on that region.

Behavior of **f** near $(-3\sqrt{2}/4, 3\sqrt{2}/4)$

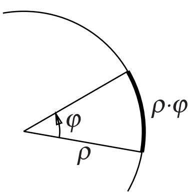
For our second illustration, let us study the local behavior of **f** at the point $(u, v) = (-3\sqrt{2}/4, 3\sqrt{2}/4)$ (so $(\rho, \varphi) = (3/2, 3\pi/4)$). The derivative is

$$\mathbf{df}_{(-3\sqrt{2}/4, 3\sqrt{2}/4)} = \begin{pmatrix} -3\sqrt{2}/2 & -3\sqrt{2}/2 \\ 3\sqrt{2}/2 & -3\sqrt{2}/2 \end{pmatrix} = 3R_{3\pi/4}.$$

Therefore, if **f** is locally linear at $(-3\sqrt{2}/4, 3\sqrt{2}/4)$, we expect **f** will approximately triple all lengths and rotate all points by $3\pi/4$ radians (i.e., 135°) in a microscope window centered at $(-3\sqrt{2}/4, 3\sqrt{2}/4)$.



The figure shows the action of **f**. To describe it, we again use a polar grid overlay in the source. At the macroscopic level, **f** "fans out" the second quadrant to cover the third and fourth quadrants. At the microscopic level, the spacing between concentric circular arcs in the source grid is $\Delta\rho = 1/36 \approx 0.028$ units, just as it was in our first illustration. However, we have reduced the spacing between radial lines in the grid to $\Delta\varphi = 1° = \pi/180$ radians, but because $\rho \approx 1.5$ in the window, the width between adjacent rays is about $1.5 \times \pi/180 \approx 0.026$ units. (By the definition of radian measure, an angle of φ radians at the center of a circle of radius ρ cuts off an arc of length $\rho \cdot \varphi$ on the circle.) This adjustment keeps the source grid roughly square.



It is evident from the image in the target window that **f** is, once again, approximately a rotation coupled with a uniform dilation. This time the 135°-line is our landmark; **f** maps it to the 270°-line in the target, rotating all points in the window therefore by about 135°. A square in the image grid is about the size of a 3×3 square in the source grid, so the linear dilation factor is about 3. Hence d**f** is indeed the local linear approximation to **f** at $(-3\sqrt{2}/4, 3\sqrt{2}/4)$.

Locally linear versus *looking linear locally*

Our third illustration analyzes the action of **f** at the origin. Here we finally get to see the difference between *looking linear* and *local linearity*. At the origin, the local action of **f** is the same as its global action: in any microscope window, no matter how small, **f** doubles angles and squares lengths. But no linear map does this, so **f** near the origin does not "look like" any linear map. In particular, $\mathbf{f}$ does not "look like" its derivative $d\mathbf{f}_{(0,0)}$ there. Nevertheless, $\mathbf{f}$ *is* locally linear at the origin; we show this in order to confirm that $\mathbf{f}$ is well-approximated by $d\mathbf{f}_{(0,0)}$ there.

Local linearity for the quadratic map

How can this be? How can $\mathbf{f}$ be *locally linear* at a point and not *look linear* at that point? How can a map be "well-approximated" by a linear map and not "look like" that linear map? The explanation lies in the definition of local linearity. A map $\mathbf{f}$ is **locally linear** at $\mathbf{u} = \mathbf{a}$ (p. 115) if

$$\Delta\mathbf{x} = \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) - \mathbf{f}(\mathbf{a}) = d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u}) + \boldsymbol{o}(1);$$

that is, if the difference $\Delta\mathbf{x} - d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})$ vanishes more rapidly than $\Delta\mathbf{u}$. To see that this is indeed true for our quadratic map, note first that we can compute $\Delta\mathbf{x} = (\Delta x, \Delta y)$ exactly:

$$\begin{aligned}
\Delta x &= (a + \Delta u)^2 - (b + \Delta v)^2 - (a^2 - b^2) \\
&= 2a\,\Delta u - 2b\,\Delta v + (\Delta u)^2 - (\Delta v)^2, \\
\Delta y &= 2(a + \Delta x)(b + \Delta v) - 2ab \\
&= 2b\,\Delta u + 2a\,\Delta v + 2\,\Delta x\,\Delta y.
\end{aligned}$$

These window equations take the vector form

$$\Delta\mathbf{x} = d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u}) + \mathbf{f}(\Delta\mathbf{u}),$$

showing us that the remainder term is just $\mathbf{f}(\Delta\mathbf{u})$. But $\mathbf{f}(\Delta\mathbf{u}) = \boldsymbol{O}(2)$ because $\mathbf{f}(\Delta\mathbf{u})$ is quadratic, and this, in turn, implies the weaker condition $\mathbf{f}(\Delta\mathbf{u}) = \boldsymbol{o}(1)$ for local linearity. Thus, $\mathbf{f}$ is "well-approximated" by the linear map $d\mathbf{f}_{\mathbf{a}}$ near $\mathbf{a}$, for every point $\mathbf{a}$, *including the origin*. But whether $\mathbf{f}$ "looks like" its linear approximation $d\mathbf{f}_{\mathbf{a}}$ in a microscope window centered at $\mathbf{a}$ will depend on the relative sizes of the two terms in the formula for $\Delta\mathbf{x}$.

At $(\sqrt{3}/2, 1/2)$, $2R_{\pi/6}(\Delta\mathbf{u})$ dominates

For example, in the window centered at $\mathbf{a} = (\sqrt{3}/2, 1/2)$ (i.e., the first window), the window equation for $\mathbf{f}$ is

$$\Delta\mathbf{x} = 2R_{\pi/6}(\Delta\mathbf{u}) + \mathbf{f}(\Delta\mathbf{u}).$$

The linear term is the rotation and uniform dilation $2R_{\pi/6}(\Delta\mathbf{u})$. This linear map is invertible, so it vanishes exactly to order 1 (Exercise 3.28, p.104). By contrast, the second term is the remainder $\mathbf{f}(\Delta\mathbf{u})$ and, as such, vanishes at least to order 2. Thus, when $\Delta\mathbf{u} \approx \mathbf{0}$ (in other words, in the microscope window), the linear term *dominates*, precisely because it vanishes to a lower order in $\Delta\mathbf{u}$. This is why the map $\mathbf{f}$ looks like its linear approximation $2R_{\pi/6}$ near $(\sqrt{3}/2, 1/2)$.

The behavior of $\mathbf{f}$ in the second window (where $\mathbf{a} = (-3\sqrt{2}/4, 3\sqrt{2}/4)$) is entirely similar: the linear term $3R_{3\pi/4}(\Delta\mathbf{u})$ is invertible so it again dominates the quadratic one $\mathbf{f}(\Delta\mathbf{u})$. Thus in the second window $\mathbf{f}$ looks like its linear approximation $3R_{3\pi/4}(\Delta\mathbf{u})$.

At the origin, the window equation still expresses the local linearity of $\mathbf{f}$, but it has the fundamentally different form

At $(0,0)$, $\mathbf{f}(\Delta\mathbf{u})$ dominates

$$\Delta\mathbf{x} = d\mathbf{f}_{\mathbf{0}}(\Delta\mathbf{u}) + \boldsymbol{O}(2) = \mathbf{0} + \mathbf{f}(\Delta\mathbf{u}).$$

The linear term, which had been dominant in the other windows, here contributes nothing. It vanishes to *infinite* order, meaning it vanishes at least to order p for every $p > 0$. The value of $\Delta\mathbf{x}$ is determined solely by the quadratic term $\mathbf{f}(\Delta\mathbf{u})$. This is what accounts for the angle-doubling and distance-squaring. By default, $\mathbf{f}(\Delta\mathbf{u})$ is the dominant term; in fact, it vanishes exactly to order 2 (see Exercise 4.12), so we are justified in saying it dominates any map (such as $d\mathbf{f}_{\mathbf{0}}$) that vanishes to higher order. Thus, in a microscope window centered at the origin, $\mathbf{f}$ does not look like its linear approximation, because the linear approximation is not the dominant term in $\Delta\mathbf{x}$. Instead, $\mathbf{f}$ looks like (indeed, is equal to) the quadratic term $\mathbf{f}(\Delta\mathbf{u})$.

$\mathbf{f}$ "looks like" $d\mathbf{f}$ if $d\mathbf{f}$ is invertible

In summary, $\mathbf{f}$ will look like its linear approximation when that linear approximation is invertible, but need not otherwise. At the moment, we are relying only on an intuitive undertanding of what it means for one map to "look like" another. We make the idea precise in the chapter on inverse maps (Chapter 5), where we say that two maps look alike if we can transform one into the other by a coordinate change. This is the same approach we took in Chapter 2 for linear maps. In that case, the coordinate change also involved finding a certain inverse map.

4.3 Parametrized surfaces

Parametrizing a surface

For another useful set of examples to illustrate the role of the derivative, we turn to surfaces in $\mathbb{R}^3$ given parametrically. Such a surface is the image of a map $\mathbf{f} : U^2 \to \mathbb{R}^3$ of a 2-dimensional region U^2 (in the same way that a parametrized curve is the image of a 1-dimensional interval). Our aim is to see how the derivative $d\mathbf{f}_{\mathbf{u}}$ is related to the map $\mathbf{f}$ near $\mathbf{u}$.

Example: unit sphere

Our first example is the unit sphere in $\mathbb{R}^3$, given parametrically as

$$\mathbf{f} : \begin{cases} x = \cos\theta\,\cos\varphi, \\ y = \sin\theta\,\cos\varphi, \\ z = \sin\varphi. \end{cases}$$

The image is indeed the unit sphere centered at the origin because every image point is exactly 1 unit from the origin:

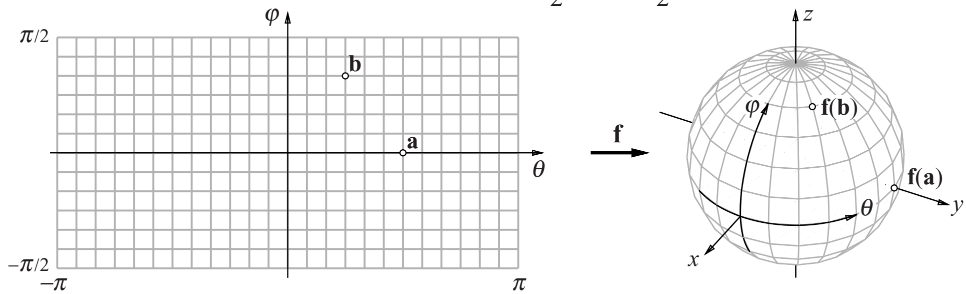
$$x^2 + y^2 + z^2 = \cos^2\theta\,\cos^2\varphi + \sin^2\theta\cos^2\varphi + \sin^2\varphi = \cos^2\varphi + \sin^2\varphi = 1.$$

Because $\cos\theta$ and $\sin\theta$ have period 2π, it is sufficient to take $-\pi \le \theta \le \pi$. When $\varphi = 0$, we have

$$x = \cos\theta, \quad y = \sin\theta, \quad z = 0;$$

thus $\mathbf{f}(\theta,0)$ traces out the unit circle in the (x,y)-plane, the *equator* of the sphere. When $\varphi = \pm\pi/2$, we have $x = y = 0$ and $z = \pm 1$, the north and south poles of the sphere. It follows that $\mathbf{f}$ already covers the entire image even if we restrict θ and φ to the rectangular domain U^2:

$$-\pi \le \theta \le \pi, \quad -\frac{\pi}{2} \le \varphi \le \frac{\pi}{2}.$$



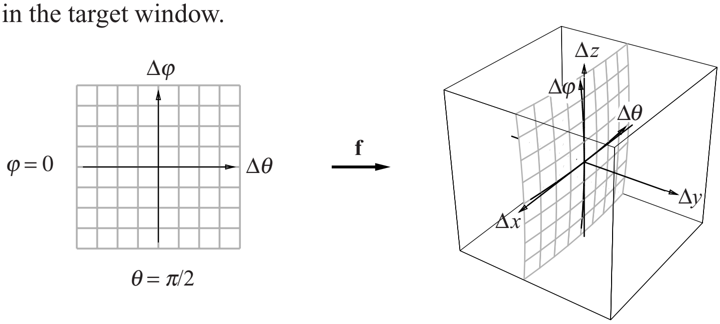
θ = longitude; φ = latitude

Note how the images of the θ- and φ-axes appear on the sphere, as the equator and the prime meridian, respectively. The parameters θ and φ are evidently just the familiar *longitude* and *latitude*. The points $\mathbf{a}$ and $\mathbf{b}$ that are marked on the (θ,φ)-plane, and their images $\mathbf{f}(\mathbf{a})$ and $\mathbf{f}(\mathbf{b})$ on the sphere, are the two sites where we compare $\mathbf{f}$ with its derivative.

We first view the action of $\mathbf{f}$ itself in a microscope window centered at each point, and then compare that with the action of the derivative at that point. The source is 2-dimensional, so a window will still be a small square. The target, however, is 3-dimensional, so each target window will be a small cube.

Action of $\mathbf{f}$ at $(\pi/2,0)$

The first point $\mathbf{a} = (\theta,\varphi) = (\pi/2,0)$ has its image $\mathbf{f}(\mathbf{a})$ on the equator at 90° east longitude, a point that has target coordinates $(x,y,z) = (0,1,0)$. In the figure below, the microscope window centered at $(\theta,\varphi) = (\pi/2,0)$ is a square 0.2 units on a side; the target window is a cube of the same dimensions centered at $(x,y,z) = (0,1,0)$. Following the figure is *Mathematica* 5code that produces the image of the $(\Delta\theta,\Delta\varphi)$-plane in the target window.




```
ParametricPlot3D[{Cos[u] Sin[v], Sin[u] Sin[v], Cos[v]},
    {u, Pi/2 - 0.1, Pi/2 + 0.1}, {v, -0.1, 0.1},
    PlotRange->{{-0.1,0.1}, {0.9,1.1}, {-0.1,0.1}},
    ViewPoint->{3.103, 2.109, 2.299}, PlotPoints->9]
```

The image is the portion of the sphere that lies in this window; it is nearly flat because the window is small. It approximates the plane $\Delta y = 0$, that is, the $(\Delta x, \Delta z)$-plane. It appears that $\mathbf{f}$ preserves lengths and angles: the image grid has the same size and shape as the source grid. The image of the $\Delta\varphi$-axis does not quite coincide with the Δz-axis, but is tangent to it. Likewise, the image of the $\Delta\theta$-axis is tangent to the Δx-axis, but has the opposite orientation.

The target window shows a small piece of the sphere

Let us now determine the action of the derivative $\mathbf{df}_{(\pi,2,0)}$ in the same microscope window. At an arbitrary point (θ, φ), the derivative map $\mathbf{df}_{(\theta,\varphi)} : \mathbb{R}^2 \to \mathbb{R}^3 : \Delta\boldsymbol{\theta} \mapsto \Delta\mathbf{x}$ given by the 3×2 matrix

Action of $\mathbf{df}_{(\pi/2,0)}$

$$\mathbf{df}_{(\theta,\varphi)} = \begin{pmatrix} -\sin\theta\cos\varphi & -\cos\theta\sin\varphi \\ \cos\theta\cos\varphi & -\sin\theta\sin\varphi \\ 0 & \cos\varphi \end{pmatrix}.$$

For each (θ, φ) (except when $\cos\varphi = 0$), the image will therefore be a plane in $\mathbb{R}^3$. When $(\theta, \varphi) = (\pi/2, 0)$, the map $\Delta\mathbf{x} = \mathbf{df}_{(\pi/2,0)}(\Delta\boldsymbol{\theta})$ is

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta\theta \\ \Delta\varphi \end{pmatrix}, \quad \text{or just} \quad \begin{aligned} \Delta x &= -\Delta\theta, \\ \Delta y &= 0, \\ \Delta z &= \Delta\varphi. \end{aligned}$$

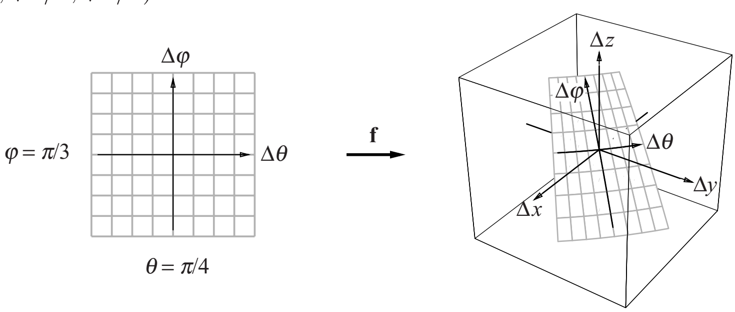
This is relatively easy to interpret. The equation $\Delta y = 0$ tells us the image is the $(\Delta x, \Delta z)$-plane. The $\Delta\varphi$-axis is mapped to the Δz-axis without stretching, and the $\Delta\theta$-axis is mapped to the Δx-axis, reversing direction but without stretching.

Our visual evidence indicates that $\mathbf{df}_{(\pi/2,0)}$ is just the "flattening-out" of $\mathbf{f}$ in a microscope window centered at $(\pi/2, 0)$. It seems reasonable to say that $\mathbf{f}$ is locally linear at $(\pi/2, 0)$ and "looks like" its derivative there.

$\mathbf{f}$ looks like $\mathbf{df}_{(\pi/2,0)}$ near $(\pi/2, 0)$

At the second point $\mathbf{b} = (\theta, \varphi) = (\pi/4, \pi/3)$, both $\mathbf{f}$ and its derivative are a bit more complicated to describe. The image $\mathbf{f}(\mathbf{b})$ lies in the northern hemisphere, at 60° north latitude and 45° east longitude; its target coordinates are $(x,y,z) = (\sqrt{2}/4, \sqrt{2}/4, \sqrt{3}/2)$.

Action of $\mathbf{f}$ at $(\pi/4, \pi/3)$

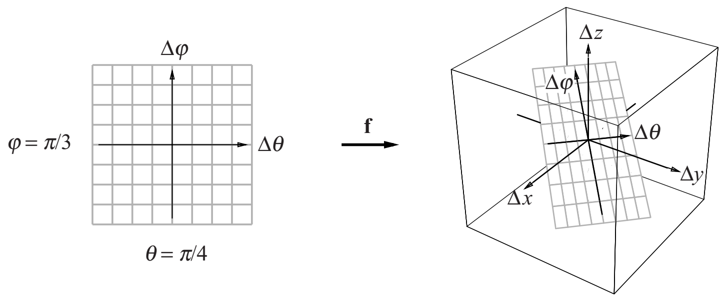


In the figure above, the microscope windows (both the square and the cube) are again 0.2 units on a side. The image is nearly flat, and is only about half as wide as it is tall. The image of the $\Delta\theta$-axis is horizontal; that is, it lies in the $(\Delta x, \Delta y)$-plane. The image of the $\Delta\varphi$-axis lies in the vertical plane where $\Delta x = \Delta y$. It would seem that we cannot tell by eye the angle between this image and the Δz-axis. But remember that we are just viewing a small portion of the sphere at 60° north latitude. That makes it obvious that the image of the $\Delta\varphi$-axis makes an angle of 60° with the vertical at the center of the window.

f does not *yet* look linear

The sides of the image seem pinched together, more so at the top than at the bottom; the grid of latitude and longitude lines is relatively far from rectangular. This is only to be expected, though. Away from the equator, longitude lines do pinch together toward the poles. In a linear map, parallel lines always have parallel images, so we must conclude that **f** does not look linear, at least at this scale.

But according to Taylor's theorem, **f** is indeed locally linear (everywhere away from the poles). We see that better when we magnify the view; the figure below is a tenfold magnification over the previous one.



f under further magnification

At this magnification, the quadrilaterals in the image grid now look like congruent rectangles, so **f** now looks like a linear map. In the $\Delta\varphi$-direction, lengths are unaltered (the image rectangles are as tall as the original squares in the source), but in the $\Delta\theta$-direction, lengths are halved.

Action of $d\mathbf{f}_{(\pi/4,\pi/3)}$

The map $\Delta\mathbf{x} = d\mathbf{f}_{(\pi/4,\pi/3)}(\Delta\boldsymbol{\theta})$ defined by the derivative is

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -\sqrt{2} & -\sqrt{6} \\ \sqrt{2} & -\sqrt{6} \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \Delta\theta \\ \Delta\varphi \end{pmatrix}.$$

The two vectors

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} -\sqrt{2}/4 \\ \sqrt{2}/4 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} -\sqrt{6}/4 \\ -\sqrt{6}/4 \\ 1/2 \end{pmatrix},$$

are the images of the unit vectors on the $\Delta\theta$- and $\Delta\varphi$-axes, respectively. We can see immediately that these image vectors are orthogonal, and you can check that their lengths are 1/2 and 1, respectively. Thus, a square grid in the $(\Delta\theta, \Delta\varphi)$-plane has for its image a rectangular grid with rectangles exactly half as wide as the squares. The image of the $\Delta\theta$-axis lies in the $(\Delta x, \Delta y)$-plane because the image vector has $\Delta z = 0$. For a similar reason, the image of the $\Delta\varphi$-axis lies in the vertical plane $\Delta x = \Delta y$. Moreover, because the dot product of the image vector with the unit vector in the Δz-direction is $2/4 = 1/2$, the image makes an angle of 60° with the Δz-axis.

Once again we have compelling visual evidence. This time it indicates that $\mathbf{df}_{(\pi/4,\pi/3)}$ matches $\mathbf{f}$ in a sufficiently small microscope window centered at the point $(\pi/4,\pi/3)$. It seems reasonable to say that $\mathbf{f}$ is locally linear at $(\pi/4,\pi/3)$ and "looks like" its derivative there.

f looks like $\mathbf{df}_{(\pi/4,\pi/3)}$ near $(\pi/4,\pi/3)$

We noted that unit vectors on the $\Delta\theta$- and $\Delta\varphi$-axes are mapped to orthogonal vectors in the target that have lengths 1/2 and 1, respectively. Thus, a unit square maps to a rectangle with area 1/2; the local area multiplier at $(\theta,\varphi) = (\pi/4,\pi/3)$ appears to be 1/2. In fact, the area multiplier for the linear map $\mathbf{df}_{(\pi/4,\pi/3)} : \mathbb{R}^2 \to \mathbb{R}^3$ is (Theorem 2.25, p. 55)

Area magnification at $(\pi/4,\pi/3)$

$$\sqrt{\begin{vmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{6}}{4} \\ 0 & \frac{1}{2} \end{vmatrix}^2 + \begin{vmatrix} 0 & \frac{1}{2} \\ -\frac{\sqrt{2}}{4} & -\frac{\sqrt{6}}{4} \end{vmatrix}^2 + \begin{vmatrix} -\frac{\sqrt{2}}{4} & -\frac{\sqrt{6}}{4} \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{6}}{4} \end{vmatrix}^2} = \sqrt{\tfrac{1}{32} + \tfrac{1}{32} + \tfrac{6}{32}} = \tfrac{1}{2}.$$

A similar calculation at $(\theta,\varphi) = (\pi/2,0)$ (using the matrix for $\mathbf{df}_{(\pi/2,0)}$) gives a local area magnification factor of

$$\sqrt{\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}^2 + \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}^2 + \begin{vmatrix} -1 & 0 \\ 0 & 0 \end{vmatrix}^2} = \sqrt{0+1+0} = 1;$$

this agrees with our discussion of the local action of $\mathbf{f}$ in the microscope window at $(\pi/2,0)$.

These examples lead us to the following definition.

Definition 4.5 *If the surface parametrization* $\mathbf{f} : U^2 \to \mathbb{R}^3$ *is differentiable at* $\mathbf{a}$*, its* ***local area multiplier*** *is the area multiplier of its derivative* $\mathbf{df_a} : \mathbb{R}^2 \to \mathbb{R}^3$.

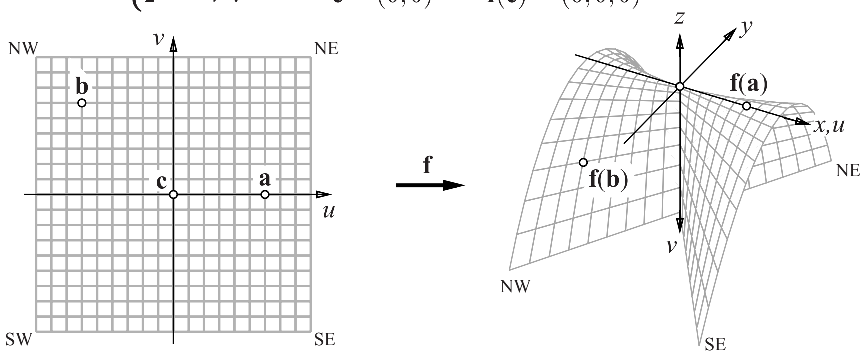
Local area multiplier

At an arbitrary point (θ,φ) in the domain of the sphere parametrization, the local area multiplier is $\cos\varphi$; see the exercises.

The next example is called a *crosscap*. It has a simple parametrization $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ in terms of polynomials defined on the entire plane.

The crosscap

$$\mathbf{f} : \begin{cases} x = u, \\ y = uv, \\ z = -v^2. \end{cases} \qquad \begin{aligned} \mathbf{a} &= (1,0) & \mathbf{f}(\mathbf{a}) &= (1,0,0) \\ \mathbf{b} &= (-1,1) & \mathbf{f}(\mathbf{b}) &= (-1,-1,-1) \\ \mathbf{c} &= (0,0) & \mathbf{f}(\mathbf{c}) &= (0,0,0) \end{aligned}$$



The image is a kind of parabolic arch that "crosses through" itself in the way shown in the figure. The u-axis is mapped to the x-axis along the ridge of the arch. The image of the v-axis folds back on itself along the line of self-intersection; both halves map to the negative z-axis. We do a local analysis at three different points. This time, though, we first compute the derivative and then compare it to the map itself.

Action of $d\mathbf{f}$ at $\mathbf{a} = (1,0)$

At an arbitrary point $\mathbf{p} = (p,q)$, the derivative $d\mathbf{f}_\mathbf{p} : \mathbb{R}^2 \to \mathbb{R}^3$ is given by the 3×2 matrix

$$\begin{pmatrix} 1 & 0 \\ q & p \\ 0 & -2q \end{pmatrix}.$$

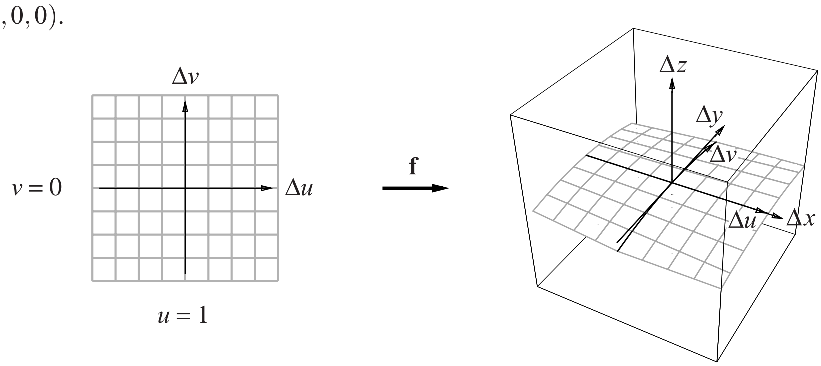
At $(p,q) = \mathbf{a} = (1,0)$, the derivative $d\mathbf{f}_\mathbf{a}$ is the map

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad \text{or} \quad \begin{aligned} \Delta x &= \Delta u, \\ \Delta y &= \Delta v, \\ \Delta z &= 0. \end{aligned}$$

This is just the identity map of the $(\Delta u, \Delta v)$-plane to the $(\Delta x, \Delta y)$-plane in the target. All lengths and angles are preserved; the local area magnification factor is therefore equal to 1.

Action of $\mathbf{f}$ near $\mathbf{a} = (1,0)$

Compare this with the action of $\mathbf{f}$ itself in a square microscope window, 0.2 units on a side, centered at $(p,q) = (1,0)$. Its image is the portion of the crosscap that appears in the small cubical window of the same dimensions, centered at $\mathbf{f}(\mathbf{a}) = (1,0,0)$.

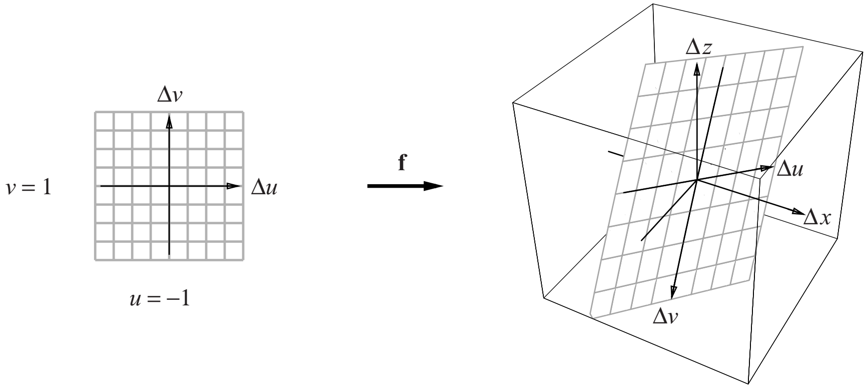


Apart from the slight curving (which would become even less noticeable if we increased the magnification), the image is the same size and shape as the source: $\mathbf{f}$ essentially preserves all lengths and angles in mapping the microscope window to the target, so $\mathbf{f}$ "looks like" $d\mathbf{f}_{(1,0)}$ near $(1,0)$.

Action of $d\mathbf{f}$ at $\mathbf{b} = (-1,1)$

At $(p,q) = \mathbf{b} = (-1,1)$, the derivative $d\mathbf{f}_\mathbf{b}$ is the map

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad \text{or} \quad \begin{aligned} \Delta x &= \Delta u, \\ \Delta y &= \Delta u - \Delta v, \\ \Delta z &= -2\Delta v. \end{aligned}$$



As always, the image is spanned by the images of the unit vectors in the Δu- and Δv-directions. In this case, the spanning vectors are $(1,1,0)$ and $(0,-1,-2)$ in $(\Delta x, \Delta y, \Delta z)$-space. The first lies in the $(\Delta x, \Delta y)$-plane, and the second lies in the $(\Delta y, \Delta z)$-plane, pointing downward (i.e., in the negative Δz-direction). The image of a coordinate grid of unit squares in the source consists of congruent parallelograms whose sides have lengths $\sqrt{2}$ and $\sqrt{5}$. Locally, areas are tripled (see the exercises).

Action of $\mathbf{f}$ near $\mathbf{b} = (-1,1)$

The figure above shows the action of $\mathbf{f}$ itself in a microscope window centered at $(p,q) = \mathbf{b} = (-1,1)$. The source window is a square 0.2 units on a side; the image cube is larger, about 0.4 units on a side, so that it can contain the entire image. As we see, the image of the Δu-axis lies in the $(\Delta x, \Delta y)$-plane, whereas the image of the Δv-axis lies in the $(\Delta y, \Delta z)$-plane, oriented so that the positive Δv-axis points down. The image coordinate grid appears to consist of congruent parallelograms that are taller than they are wide. The same figure serves to represent both $d\mathbf{f}_{\mathbf{b}}$ and $\mathbf{f}$ near $\mathbf{b}$; it shows that $\mathbf{f}$ "looks like" its linear approximation near $\mathbf{b} = (-1,1)$.

Action of $d\mathbf{f}$ at $\mathbf{c} = (0,0)$

At the origin the situation is not so simple. Perhaps this is to be expected, because it is the place where the crosscap "crosses" itself. The derivative $d\mathbf{f}_{\mathbf{c}}$ is the linear map

$$\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad \text{or} \quad \begin{aligned} \Delta x &= \Delta u, \\ \Delta y &= 0, \\ \Delta z &= 0. \end{aligned}$$

The rank of the matrix has dropped to 1. This means the image has dimension 1 instead of 2; instead of a plane, it has collapsed to a line. Indeed, the equations indicate the image is just the Δx-axis. Furthermore, the local area magnification factor is equal to 0.

Action of $\mathbf{f}$ near $\mathbf{c} = (0,0)$

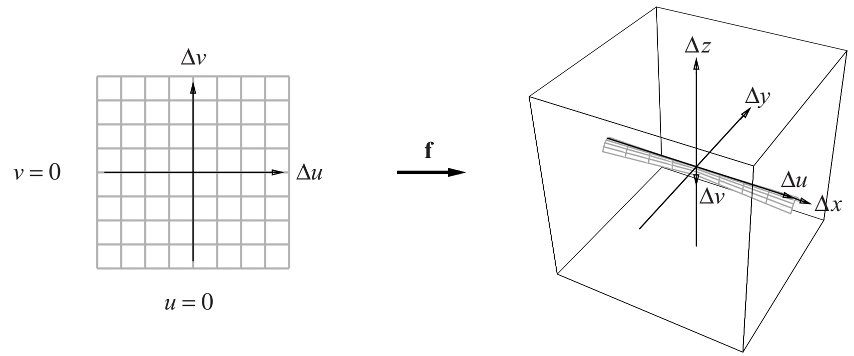
In the exercises you compute the window equation

$$\Delta \mathbf{x} = \mathbf{f}(\mathbf{p} + \Delta \mathbf{u}) - \mathbf{f}(\mathbf{p}) = d\mathbf{f}_{\mathbf{p}}(\Delta \mathbf{u}) + \mathbf{R}_{1,\mathbf{p}}(\Delta \mathbf{u})$$

for the crosscap map $\mathbf{f}$ at an arbitrary point $\mathbf{p}$. Because $\mathbf{f}$ is a simple quadratic map, you get an explicit (quadratic) formula for the remainder $\mathbf{R}_{1,\mathbf{p}}$. At the origin, $\mathbf{p} = \mathbf{c} = \mathbf{0}$, the formula for $\Delta \mathbf{x}$ reduces to

$$\begin{aligned}
\Delta\mathbf{x} &= d\mathbf{f_0}(\Delta\mathbf{u}) + \mathbf{R}_{1,\mathbf{0}}(\Delta\mathbf{u}),\\
\Delta x &= \quad \Delta u \quad + \quad 0 \quad = O(1),\\
\Delta y &= \quad 0 \quad + \quad \Delta u\,\Delta v \quad = O(2),\\
\Delta z &= \quad 0 \quad + \quad -(\Delta v)^2 \quad = O(2).
\end{aligned}$$

Thus, in a microscope window centered at $\mathbf{x} = \mathbf{0}$, the values of Δy and Δz are an order of magnitude smaller than Δx. And that is what we see in the figure below.



The image has been squeezed in the Δv-direction so that it fits into a narrow tube along the Δx-axis. The source and target windows are both 0.2 units on a side, but the tube's dimensions are only 0.01×0.01 in the Δy- and Δz-directions; they are an order of magnitude smaller than the long dimension.

How $\mathbf{f}$ differs from $d\mathbf{f_0}$

Does $\mathbf{f}$ "look like" $d\mathbf{f_0}$ in this window? Not quite. In the Δx-direction, the derivative $d\mathbf{f_0}$ vanishes only to order 1 but the remainder $\mathbf{R}_{1,\mathbf{0}}$ vanishes to infinite order. The derivative dominates, and $\mathbf{f}$ does indeed look like $d\mathbf{f_0}$ in that direction. But in the Δy- and Δz-directions, $d\mathbf{f_0}$ now vanishes to infinite order, but $\mathbf{R}_{1,\mathbf{0}}$—and thus $\mathbf{f}$ itself—vanishes only to order 2. The remainder dominates; $\mathbf{f}$ therefore looks like the remainder in those directions, and not like its derivative. So, even though $\mathbf{f}$ is "well-approximated" by its derivative at $\mathbf{u} = \mathbf{0}$ (i.e., even though $\mathbf{f}$ is differentiable), it does not look like its derivative in a microscope window centered there.

$\mathbf{f}$ "looks like" $d\mathbf{f}$ if $d\mathbf{f}$ has maximal rank

In our study of the quadratic map in the previous section we noted that the quadratic map (pp. 116–121) failed to look like its derivative locally at a point where the derivative itself failed to be invertible. Invertibility is out of the question here, because the source and target have different dimensions (the derivative is not a square matrix). The proper analogue is *maximal rank*:

If $d\mathbf{f_a}$ has maximal rank, $\mathbf{f}$ will *look like* $d\mathbf{f_a}$ near $\mathbf{a}$.

We investigate this point further in the chapter on implicit functions (Chapter 6).

4.4 The chain rule

In elementary calculus, every formula is assembled from a few simple types of functions—think of them as "atoms"—by using arithmetic and composition. We calculate the derivative of any such formula by knowing the derivatives of the atoms and the rules for differentiating arbitrary sums, products, quotients, and chains (or compositions). In this section we develop rules for derivatives of maps that generalize the sum, product, and chain rules. As in the one-variable case, the most important is the chain rule; with it we show how the derivatives of a map and its inverse are related.

Differentiability of a map

Before considering the differentiation rules, we must say what it means for a map between spaces of arbitrary dimension to be differentiable. Our definition is just the generalization of the ones we have used in special cases (Definitions 4.1, p. 106 and 4.3, p. 115); U^p is a *window* of the form $|u_i - a_i| < q_i$, $i = 1, \cdots, p$.

Definition 4.6 *The map* $\mathbf{f} : U^p \to \mathbb{R}^q$ *is* ***differentiable****, or* ***locally linear****, at* $\mathbf{u} = \mathbf{a}$ *if there is a linear map* $\mathbf{L} : \mathbb{R}^p \to \mathbb{R}^q$*, called the* ***derivative*** *of* $\mathbf{f}$ *at* $\mathbf{a}$*, for which*

$$\mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) = \mathbf{f}(\mathbf{a}) + \mathbf{L}(\Delta\mathbf{u}) + \boldsymbol{o}(1).$$

Theorem 4.6. *Suppose* $\mathbf{f} : U^p \to \mathbb{R}^q$ *is differentiable at* $\mathbf{u} = \mathbf{a}$*; then* $\mathbf{L} = d\mathbf{f}_{\mathbf{a}}$*. In particular, if the component functions of* $\mathbf{f}$ *are* $f_i(\mathbf{u})$*, then all the partial derivatives* $\partial f_i / \partial u_j(\mathbf{a})$ *exist.*

Proof. Suppose the element in the ith row and jth column of the matrix representing $\mathbf{L}$ is ℓ_{ij}. We show that the partial derivative $\partial f_i / \partial u_j(\mathbf{a})$ exists and is equal to ℓ_{ij}. Let $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$, the vector in $\mathbb{R}^p$ with 1 in the jth coordinate and 0 elsewhere. Then, by definition,

$$\frac{\partial f_i}{\partial u_j}(\mathbf{a}) = \lim_{h \to 0} \frac{f_i(\mathbf{a} + h\mathbf{e}_j) - f_i(\mathbf{a})}{h}.$$

By hypothesis, $f_i(\mathbf{a} + h\mathbf{e}_j) - f_i(\mathbf{a}) = L_i(h\mathbf{e}_j) + R_i(h\mathbf{u}) = hL_i(\mathbf{e}_j) + R_i(h\mathbf{u})$, where L_i is the ith component of $\mathbf{L}$ and R_i is the ith component of the remainder map that is represented by the symbol $\boldsymbol{o}(1)$. Therefore,

$$\begin{aligned}\frac{\partial f_i}{\partial u_j}(\mathbf{a}) &= \lim_{h \to 0} \frac{f_i(\mathbf{a} + h\mathbf{e}_j) - f_i(\mathbf{a})}{h} = \lim_{h \to 0} \frac{hL_i(\mathbf{e}_j) + R_i(h\mathbf{e}_j)}{h} \\ &= L_i(\mathbf{e}_j) + \lim_{h \to 0} \frac{R_i(h\mathbf{e}_j)}{h} = \ell_{ij}.\end{aligned}$$

The final equation holds because $L_i(\mathbf{e}_j) = \ell_{ij}$, and $R_i(h\mathbf{e}_j)/h \to 0$ is simply a consequence of $R_i(\Delta\mathbf{u}) = o(1)$. □

Differentiability is local linearity

The theorem shows $\mathbf{L}$ is unique. We continue to equate differentiability and local linearity: $\mathbf{f}$ is differentiable at $\mathbf{a}$ if and only if $\Delta\mathbf{x} = \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) - \mathbf{f}(\mathbf{a})$ agrees with the linear map $d\mathbf{f_a}(\Delta\mathbf{u})$ to order greater than 1 as $\Delta\mathbf{u} \to \mathbf{0}$. We use this characterization repeatedly in the rest of the section.

Derivative of a sum

Given two maps $\mathbf{f}, \mathbf{g} : U^p \to \mathbb{R}^q$, their sum and difference,

$$(\mathbf{f} \pm \mathbf{g})(\mathbf{u}) = \mathbf{f}(\mathbf{u}) \pm \mathbf{g}(\mathbf{u}),$$

are themselves maps of the same type: $\mathbf{f} \pm \mathbf{g} : U^p \to \mathbb{R}^q$. The following theorem extends to maps the familiar rule: *the derivative of a sum is the sum of the derivatives*.

Theorem 4.7. *If* $\mathbf{f}$ *and* $\mathbf{g}$ *are differentiable on* U^p*, then so are* $\mathbf{f} \pm \mathbf{g}$*, and*

$$d(\mathbf{f} \pm \mathbf{g})_{\mathbf{a}}(\Delta\mathbf{u}) = d\mathbf{f_a}(\Delta\mathbf{u}) \pm d\mathbf{g_a}(\Delta\mathbf{u})$$

at any point $\mathbf{a}$ *in* U^p.

Proof. The statements are probably intuitively clear, but we prove them to illustrate our characterization of a differentiable map. Because $\mathbf{f}$ and $\mathbf{g}$ are differentiable at $\mathbf{a}$, we can write

$$\begin{aligned}(\mathbf{f} \pm \mathbf{g})(\mathbf{a} + \Delta\mathbf{u}) &= \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) \pm \mathbf{g}(\mathbf{a} + \Delta\mathbf{u}) \\ &= \mathbf{f}(\mathbf{a}) + d\mathbf{f_a}(\Delta\mathbf{u}) + \boldsymbol{o}(1) \pm (\mathbf{g}(\mathbf{a}) + d\mathbf{g_a}(\Delta\mathbf{u}) + \boldsymbol{o}(1)) \\ &= (\mathbf{f} \pm \mathbf{g})(\mathbf{a}) + (d\mathbf{f_a} \pm d\mathbf{g_a})(\Delta\mathbf{u}) + \boldsymbol{o}(1).\end{aligned}$$

Thus $(\mathbf{f} \pm \mathbf{g})(\mathbf{a} + \Delta\mathbf{u}) - (\mathbf{f} \pm \mathbf{g})(\mathbf{a})$ agrees with the linear map $(d\mathbf{f_a} \pm d\mathbf{g_a})(\Delta\mathbf{u})$ to order greater than 1; by Theorem 4.6, the maps $\mathbf{f} \pm \mathbf{g}$ are differentiable at $\mathbf{a}$, and $d(\mathbf{f} \pm \mathbf{g})_{\mathbf{a}} = d\mathbf{f_a} \pm d\mathbf{g_a}$. □

Derivative of a scalar product

There are product rules, too, at least when the products themselves are defined in meaningful ways. For example, we can compute the ordinary cross-product (or vector product) of maps whose target is $\mathbb{R}^3$, and we can compute the dot product (or scalar product) of any maps whose targets have the same dimension.

Theorem 4.8. *If* $\mathbf{f}, \mathbf{g} : U^p \to \mathbb{R}^q$ *are differentiable on* U^p*, then so is the scalar product* $\mathbf{f} \cdot \mathbf{g} : U^p \to \mathbb{R}$*, and*

$$d(\mathbf{f} \cdot \mathbf{g})_{\mathbf{a}}(\Delta\mathbf{u}) = \mathbf{f}(\mathbf{a}) \cdot d\mathbf{g_a}(\Delta\mathbf{u}) + d\mathbf{f_a}(\Delta\mathbf{u}) \cdot \mathbf{g}(\mathbf{a}).$$

Proof. By definition of the scalar product function, and then by the differentiability of $\mathbf{f}$ and $\mathbf{g}$, we have

$$\begin{aligned}(\mathbf{f} \cdot \mathbf{g})(\mathbf{a} + \Delta\mathbf{u} &= \mathbf{f}(\mathbf{a} + \Delta\mathbf{u}) \cdot \mathbf{g}(\mathbf{a} + \Delta\mathbf{u}) \\ &= \big(\mathbf{f}(\mathbf{a}) + d\mathbf{f_a}(\Delta\mathbf{u}) + \boldsymbol{o}(1)\big) \cdot \big(\mathbf{g}(\mathbf{a}) + d\mathbf{g_a}(\Delta\mathbf{u}) + \boldsymbol{o}(1)\big).\end{aligned}$$

When we expand the right-hand side, we get nine individual scalar terms, five of which have $\boldsymbol{o}(1)$ as a factor. Those five therefore all vanish to order greater than 1, so we combine them into a single (scalar) symbol $o(1)$:

$$(\mathbf{f}\cdot\mathbf{g})(\mathbf{a}+\Delta\mathbf{u}) = \mathbf{f}(\mathbf{a})\cdot\mathbf{g}(\mathbf{a}) + \mathbf{f}(\mathbf{a})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot\mathbf{g}(\mathbf{a}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + o(1).$$

To decide whether $\mathbf{f}\cdot\mathbf{g}$ is differentiable, Theorem 4.6 suggests we rewrite the last equation (after setting $\mathbf{f}(\mathbf{a})\cdot\mathbf{g}(\mathbf{a}) = (\mathbf{f}\cdot\mathbf{g})(\mathbf{a})$) in the form

$$(\mathbf{f}\cdot\mathbf{g})(\mathbf{a}+\Delta\mathbf{u}) - (\mathbf{f}\cdot\mathbf{g})(\mathbf{a}) = \mathbf{f}(\mathbf{a})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot\mathbf{g}(\mathbf{a}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + o(1).$$

Let us now consider, in turn, the first three terms on the right. We know $d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u})$ is a linear function of $\Delta\mathbf{u}$, and so is its dot product with the scalar $\mathbf{f}(\mathbf{a})$. The second term is likewise a linear function of $\Delta\mathbf{u}$. In the third term, each of the two factors is linear; by Exercise 3.28 (p. 104), each factor vanishes at least to order 1. That is, there are constants $C_{\mathbf{f}}$ and $C_{\mathbf{g}}$ such that

$$\|d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\| \le C_{\mathbf{f}}\|\Delta\mathbf{u}\|, \quad \|d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u})\| \le C_{\mathbf{g}}\|\Delta\mathbf{u}\|.$$

Therefore, because $|A\cdot B| \le \|A\|\|B\|$ for any two vectors in $\mathbb{R}^q$,

$$|d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u})| \le \|d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\|\|d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u})\| \le C_{\mathbf{f}}C_{\mathbf{g}}\|\Delta\mathbf{u}\|^2,$$

implying that $d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u})$ is $O(2)$ and hence can be absorbed into the term denoted $o(1)$. Thus we see $(\mathbf{f}\cdot\mathbf{g})(\mathbf{a}+\Delta\mathbf{u}) - (\mathbf{f}\cdot\mathbf{g})(\mathbf{a})$ agrees with the linear function $\mathbf{f}(\mathbf{a})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot\mathbf{g}(\mathbf{a})$ to order greater than 1, so it follows that $d(\mathbf{f}\cdot\mathbf{g})_{\mathbf{a}}(\Delta\mathbf{u}) = \mathbf{f}(\mathbf{a})\cdot d\mathbf{g}_{\mathbf{a}}(\Delta\mathbf{u}) + d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\cdot\mathbf{g}(\mathbf{a})$. □

The chain rule for one-variable functions

We turn now to the chain rule. For functions of a single variable, the chain rule is commonly written two different ways, corresponding to the two ways we write derivatives. Suppose $s = f(u)$ and $x = \varphi(s)$; then $x = \varphi(f(u))$ and x therefore depends on u through the action of a new function composed of f and φ. We write the composed function as $x = (\varphi\circ f)(u)$, and write its derivative in terms of the derivatives of the components f and φ as either

$$\frac{dx}{du} = \frac{dx}{ds}\frac{ds}{du} \quad \text{or} \quad (\varphi\circ f)'(a) = \varphi'(f(a))\,f'(a).$$

These are two formulations of the chain rule. The first uses the *Leibniz notation* for derivatives; its appeal is that it looks like an ordinary rule for multiplying fractions. The second calls attention to the fact that the derivatives of the individual functions φ and f must be evaluated at different points. It also reminds us that x depends on u one way (namely, through $\varphi\circ f$) but on s a different way (namely, through φ alone). The Leibniz notations dx/du and dx/ds suggest the same thing, though somewhat more obliquely.

Diagrams of maps

Now suppose $\mathbf{f}: U^p \to \mathbb{R}^q$ and $\boldsymbol{\varphi}: S^q \to \mathbb{R}^r$ are differentiable maps with the image of $\mathbf{f}$ contained in the domain of $\boldsymbol{\varphi}$: $\mathbf{f}(U^p) \subseteq S^q$. Then the composite $\boldsymbol{\varphi}\circ\mathbf{f}: U^p \to \mathbb{R}^r$ is defined for all $\mathbf{u}$ in U^p: $(\boldsymbol{\varphi}\circ\mathbf{f})(\mathbf{u}) = \boldsymbol{\varphi}(\mathbf{f}(\mathbf{u}))$. Visually, we can think of $\mathbf{f}$ and $\boldsymbol{\varphi}$ as maps coming one after another in a linear "chain" as on the left, below.

However, to show how they are related to the composite map $\boldsymbol{\varphi} \circ \mathbf{f}$, it is natural to put the three maps in a triangle:

$$U^p \xrightarrow{\ \mathbf{f}\ } S^q \xrightarrow{\ \boldsymbol{\varphi}\ } \mathbb{R}^r$$

$$\begin{tikzcd} & S^q \arrow[dr, "\boldsymbol{\varphi}"] & \\ U^p \arrow[ur, "\mathbf{f}"] \arrow[rr, "\boldsymbol{\varphi} \circ \mathbf{f}"'] & & \mathbb{R}^r \end{tikzcd}$$

The chain rule for maps

The chain rule for maps says that the derivative of a composite is the composite of its derivatives. We state and prove the theorem below. You should check that the proof is just a rigorous version of the following "plausibility argument" based on the microscope equations for $\mathbf{f}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\varphi} \circ \mathbf{f}$. Starting with

$$\Delta\mathbf{s} \approx d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u}) \text{ and } \Delta\mathbf{x} \approx d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}(\Delta\mathbf{s}),$$

it follows that

$$\Delta\mathbf{x} \approx d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\big) = (d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}})(\Delta\mathbf{u}).$$

But, by definition, $\Delta\mathbf{x} \approx d(\boldsymbol{\varphi} \circ \mathbf{f})_{\mathbf{a}}(\Delta\mathbf{u})$; because the linear map in the microscope equation is unique, $d(\boldsymbol{\varphi} \circ \mathbf{f})_{\mathbf{a}} = d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}}$.

Theorem 4.9 (Chain rule). *If* $\mathbf{f} : U^p \to S^q$ *is differentiable at* $\mathbf{a}$*, and* $\boldsymbol{\varphi} : S^q \to \mathbb{R}^r$ *is differentiable at* $\mathbf{f}(\mathbf{a})$*, then the composite map* $\boldsymbol{\varphi} \circ \mathbf{f} : U^p \to \mathbb{R}^r$ *is differentiable at* $\mathbf{a}$ *and*

$$d(\boldsymbol{\varphi} \circ \mathbf{f})_{\mathbf{a}} = d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}}.$$

$$\begin{tikzcd} & S^q \arrow[dr, "\boldsymbol{\varphi}"] & \\ U^p \arrow[ur, "\mathbf{f}"] \arrow[rr, "\boldsymbol{\varphi} \circ \mathbf{f}"'] & & \mathbb{R}^r \end{tikzcd} \qquad \begin{tikzcd} & \mathbb{R}^q \arrow[dr, "d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}"] & \\ \mathbb{R}^p \arrow[ur, "d\mathbf{f}_{\mathbf{a}}"] \arrow[rr, "d(\boldsymbol{\varphi} \circ \mathbf{f})_{\mathbf{a}}"'] & & \mathbb{R}^r \end{tikzcd}$$

Proof. It is possible to prove this result in terms of the component functions of the maps. However, we work directly with the maps themselves. According to our characterization of differentiability (and taking into account that $d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\big) = d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})$ by definition), we must therefore show

$$(\boldsymbol{\varphi} \circ \mathbf{f})(\mathbf{a} + \Delta\mathbf{u}) - (\boldsymbol{\varphi} \circ \mathbf{f})(\mathbf{a}) = d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\big) + \boldsymbol{o}(1).$$

This will prove that the derivative of $\boldsymbol{\varphi} \circ \mathbf{f}$ at $\mathbf{a}$ is $d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}}$.

To begin, the differentiability of $\mathbf{f}$ at $\mathbf{a}$ allows us to write

$$(\boldsymbol{\varphi} \circ \mathbf{f})(\mathbf{a} + \Delta\mathbf{u}) = \boldsymbol{\varphi}\big(\mathbf{f}(\mathbf{a} + \Delta\mathbf{u})\big) = \boldsymbol{\varphi}\Big(\underbrace{\mathbf{f}(\mathbf{a})}_{\mathbf{b}} + \underbrace{d\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u}) + \boldsymbol{o}(\Delta\mathbf{u})}_{\Delta\mathbf{s}}\Big).$$

We write $\boldsymbol{o}(\Delta\mathbf{u})$ here, instead of just $\boldsymbol{o}(1)$, to stress that this particular remainder vanishes (to order greater than 1) with $\Delta\mathbf{u}$. Now use the differentiability of $\boldsymbol{\varphi}$ at $\mathbf{f}(\mathbf{a}) = \mathbf{b}$ to expand the right-hand side. This yields a second remainder $\boldsymbol{o}(\Delta\mathbf{s})$ that vanishes with $\Delta\mathbf{s}$ and is thus distinct from $\boldsymbol{o}(\Delta\mathbf{u})$:

$$\begin{aligned}\boldsymbol{\varphi}(\mathbf{b})+\mathrm{d}\boldsymbol{\varphi}_{\mathbf{b}}(\Delta\mathbf{s})+\boldsymbol{o}(\Delta\mathbf{s}) &= \boldsymbol{\varphi}(\mathbf{f}(\mathbf{a}))+\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})+\boldsymbol{o}(\Delta\mathbf{u})\big)+\boldsymbol{o}(\Delta\mathbf{s})\\ &= (\boldsymbol{\varphi}\circ\mathbf{f})(\mathbf{a})+\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\big)+\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(\boldsymbol{o}(\Delta\mathbf{u})\big)+\boldsymbol{o}(\Delta\mathbf{s}).\end{aligned}$$

We used the linearity of $\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}$ to split the second term into two. We write the new remainder as $\boldsymbol{o}(\Delta\mathbf{s})$ to indicate that it is a function of $\Delta\mathbf{s}$ rather than $\Delta\mathbf{u}$ and that, as such, it vanishes to order greater than 1 with $\Delta\mathbf{s}$. At this stage we have

$$(\boldsymbol{\varphi}\circ\mathbf{f})(\mathbf{a}+\Delta\mathbf{u})-(\boldsymbol{\varphi}\circ\mathbf{f})(\mathbf{a})=\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})\big)+\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}\big(\boldsymbol{o}(\Delta\mathbf{u})\big)+\boldsymbol{o}(\Delta\mathbf{s}).$$

It remains only to show that the last two terms on the right vanish to order greater than 1 in $\Delta\mathbf{u}$.

Lemma 4.1. $\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}(\boldsymbol{o}(\Delta\mathbf{u}))=\boldsymbol{o}(\Delta\mathbf{u})$.

Proof. Because $\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}$ is a linear map, we know (Exercise 3.28, p. 104) there is a positive constant C for which $\|\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}(\boldsymbol{o}(\Delta\mathbf{u}))\|\le C\|\boldsymbol{o}(\Delta\mathbf{u}))\|$. Therefore,

$$\lim_{\Delta\mathbf{u}\to\mathbf{0}}\frac{\|\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}(\boldsymbol{o}(\Delta\mathbf{u}))\|}{\|\Delta\mathbf{u}\|}\le\lim_{\Delta\mathbf{u}\to\mathbf{0}}\frac{C\|\boldsymbol{o}(\Delta\mathbf{u})\|}{\|\Delta\mathbf{u}\|}=0,$$

by the definition of $\boldsymbol{o}(\Delta\mathbf{u})$. Thus $\mathrm{d}\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})}(\boldsymbol{o}(\Delta\mathbf{u}))=\boldsymbol{o}(\Delta\mathbf{u})$. □

Lemma 4.2. *Let* $\Delta\mathbf{s}=\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})+\boldsymbol{o}(\Delta\mathbf{u})$*; then* $\Delta\mathbf{s}=\boldsymbol{O}(\Delta\mathbf{u})$.

Proof. The first term, $\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{u})$, is linear, so by Exercise 3.28.a, it vanishes at least to order 1 in $\Delta\mathbf{u}$. The second term certainly vanishes at least to order 1 in $\Delta\mathbf{u}$, so $\Delta\mathbf{s}=\boldsymbol{O}(\Delta\mathbf{u})$. □

Lemma 4.3. *If* $\Delta\mathbf{s}=\boldsymbol{O}(\Delta\mathbf{u})$*, then* $\boldsymbol{o}(\Delta\mathbf{s})=\boldsymbol{o}(\Delta\mathbf{u})$.

Proof. We must show $\|\boldsymbol{o}(\Delta\mathbf{s})\|/\|\Delta\mathbf{u}\|\to 0$ as $\Delta\mathbf{u}\to\mathbf{0}$; note that there are different variables in the numerator and the denominator. The two variables are linked, however: $\Delta\mathbf{s}=\boldsymbol{O}(\Delta\mathbf{u})$. In fact, this hypothesis means $\Delta\mathbf{s}\to\mathbf{0}$ as $\Delta\mathbf{u}\to\mathbf{0}$, suggesting that we write

$$\frac{\|\boldsymbol{o}(\Delta\mathbf{s})\|}{\|\Delta\mathbf{u}\|}=\frac{\|\boldsymbol{o}(\Delta\mathbf{s})\|}{\|\Delta\mathbf{s}\|}\cdot\frac{\|\Delta\mathbf{s}\|}{\|\Delta\mathbf{u}\|}.$$

Now the second factor on the right is bounded as $\Delta\mathbf{u}\to\mathbf{0}$, because $\Delta\mathbf{s}=\boldsymbol{O}(\Delta\mathbf{u})$. The first factor tends to zero as $\Delta\mathbf{s}\to\mathbf{0}$, by definition of $\boldsymbol{o}(\Delta\mathbf{s})$. Because $\Delta\mathbf{s}\to\mathbf{0}$ as $\Delta\mathbf{u}\to\mathbf{0}$, it appears we have shown that $\|\boldsymbol{o}(\Delta\mathbf{s})\|/\|\Delta\mathbf{u}\|$ does indeed tend to 0 as $\Delta\mathbf{u}\to\mathbf{0}$.

But $\Delta\mathbf{s}$ may be zero for some $\Delta\mathbf{u}\ne\mathbf{0}$, so the first factor $\|\boldsymbol{o}(\Delta\mathbf{s})\|/\|\Delta\mathbf{s}\|$ is undefined and the argument fails. We need to avoid quotients here. Fortunately, Exercise 3.17 (p. 102) provides an alternate formulation of "little oh" without quotients. The alternate formulation of the condition $\boldsymbol{o}(\Delta\mathbf{s})=\boldsymbol{o}(\Delta\mathbf{u})$ that we seek to prove is as follows. For any given $\varepsilon>0$, we must be able to find a $\delta>0$ so that

$$\|\boldsymbol{o}(\Delta\mathbf{s})\|\le\varepsilon\|\Delta\mathbf{u}\|\quad\text{when}\quad\|\Delta\mathbf{u}\|<\delta.$$

To find δ, first note that $\Delta\mathbf{s} = \boldsymbol{O}(\Delta\mathbf{u})$ means, by definition, that there are positive constants δ_1 and C for which $\|\Delta\mathbf{s}\| \le C\|\Delta\mathbf{u}\|$ when $\|\Delta\mathbf{u}\| < \delta_1$. The alternate formulation of "little oh" implies that, for the ε already given, we can choose δ_2 so that

$$\|\boldsymbol{o}(\Delta\mathbf{s})\| \le \frac{\varepsilon}{C}\|\Delta\mathbf{s}\| \quad \text{when } \|\Delta\mathbf{s}\| < \delta_2.$$

Finally, if we let δ be the smaller of δ_1 and δ_2/C, then $\|\Delta\mathbf{u}\| < \delta$ implies first that $\|\Delta\mathbf{s}\| \le C\|\Delta\mathbf{u}\| < C\delta \le \delta_2$ and consequently that

$$\|\boldsymbol{o}(\Delta\mathbf{s})\| \le \frac{\varepsilon}{C}\|\Delta\mathbf{s}\| \le \varepsilon\|\Delta\mathbf{u}\|. \qquad \square$$

To complete the proof of the theorem, we just need to combine the results of Lemmas 4.2 and 4.3 to conclude that if $\Delta\mathbf{s} = d\mathbf{f}_\mathbf{a}(\Delta\mathbf{u}) + \boldsymbol{o}(\Delta\mathbf{u})$, then $\boldsymbol{o}(\Delta\mathbf{s}) = \boldsymbol{o}(\Delta\mathbf{u})$. $\square$

Example: a chain of maps of the plane

Here is an example that shows how the chain rule works for a pair of maps of the plane. The first, $\mathbf{f}$, is the polar coordinate map and the second, $\boldsymbol{\varphi}$, is the conformal quadratic map; these are Examples 1 and 2 in Chapter 4.2.

$$\boldsymbol{\varphi} : \begin{cases} x = u^2 - v^2, \\ y = 2uv. \end{cases} \qquad \mathbf{f} : \begin{cases} u = \rho\cos\varphi, \\ v = \rho\sin\varphi, \end{cases}$$

Their composite is

$$\boldsymbol{\varphi}\circ\mathbf{f} : \begin{cases} x = \rho^2\cos^2\varphi - \rho^2\sin^2\varphi = \rho^2\cos 2\varphi, \\ y = 2\rho\cos\varphi\cdot\rho\sin\varphi = \rho^2\sin 2\varphi. \end{cases}$$

With these formulas for the component functions of $\boldsymbol{\varphi}\circ\mathbf{f}$, we can compute the derivative directly, without using the chain rule. At an arbitrary point (ρ,φ), the derivative is

$$d(\boldsymbol{\varphi}\circ\mathbf{f})_{(\rho,\varphi)} = \begin{pmatrix} 2\rho\cos 2\varphi & -2\rho^2\sin 2\varphi \\ 2\rho\sin 2\varphi & 2\rho^2\cos 2\varphi \end{pmatrix}.$$

Let us compare this with the derivative obtained with the chain rule. We start with the derivatives of the individual maps $\boldsymbol{\varphi}$ and $\mathbf{f}$:

$$d\boldsymbol{\varphi}_{(u,v)} = \begin{pmatrix} 2u & -2v \\ 2v & 2u \end{pmatrix}, \quad d\mathbf{f}_{(\rho,\varphi)} = \begin{pmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{pmatrix}.$$

These have been evaluated at arbitrary points in the domains of the maps. But, in the chain rule, $d\boldsymbol{\varphi}$ must be evaluated at $\mathbf{f}(\rho,\varphi) = (\rho\cos\varphi, \rho\sin\varphi)$. Thus,

$$d\boldsymbol{\varphi}_{\mathbf{f}(\rho,\varphi)} = \begin{pmatrix} u & -v \\ v & u \end{pmatrix}\Bigg|_{\substack{u=\rho\cos\varphi \\ v=\rho\sin\varphi}} = \begin{pmatrix} 2\rho\cos\varphi & -2\rho\sin\varphi \\ 2\rho\sin\varphi & 2\rho\cos\varphi \end{pmatrix},$$

and the matrix product we seek is

$$
\begin{aligned}
d\boldsymbol{\varphi}_{\mathbf{f}(\rho,\varphi)} \circ d\mathbf{f}_{(\rho,\varphi)} &= \begin{pmatrix} 2\rho\cos\varphi & -2\rho\sin\varphi \\ 2\rho\sin\varphi & 2\rho\cos\varphi \end{pmatrix} \begin{pmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{pmatrix} \\
&= \begin{pmatrix} 2\rho\cos^2\varphi - 2\rho\sin^2\varphi & -2\rho^2\cos\varphi\sin\varphi - 2\rho^2\sin\varphi\cos\varphi \\ 2\rho\sin\varphi\cos\varphi + 2\rho\cos\varphi\sin\varphi & -2\rho^2\sin^2\varphi + 2\rho^2\cos^2\varphi \end{pmatrix} \\
&= \begin{pmatrix} 2\rho\cos 2\varphi & -2\rho^2\sin 2\varphi \\ 2\rho\sin 2\varphi & 2\rho^2\cos 2\varphi \end{pmatrix} = d(\boldsymbol{\varphi} \circ \mathbf{f})_{(\rho,\varphi)}.
\end{aligned}
$$

The chain rule evidently holds in this case.

For a second example, we consider two maps of the plane defined by arbitrary component functions:

Example: arbitrary maps of the plane

$$
\boldsymbol{\varphi} : \begin{cases} x = \varphi(s,t), \\ y = \psi(s,t), \end{cases} \qquad \mathbf{f} : \begin{cases} s = f(u,v), \\ t = g(u,v). \end{cases}
$$

The derivatives of these maps are the matrices

$$
d\boldsymbol{\varphi}_{(s,t)} = \begin{pmatrix} \dfrac{\partial\varphi}{\partial s} & \dfrac{\partial\varphi}{\partial t} \\[2ex] \dfrac{\partial\psi}{\partial s} & \dfrac{\partial\psi}{\partial t} \end{pmatrix}, \quad d\mathbf{f}_{(u,v)} = \begin{pmatrix} \dfrac{\partial f}{\partial u} & \dfrac{\partial f}{\partial v} \\[2ex] \dfrac{\partial g}{\partial u} & \dfrac{\partial g}{\partial v} \end{pmatrix},
$$

whose product is the derivative of the composite map $\boldsymbol{\varphi} \circ \mathbf{f}$:

$$
d(\boldsymbol{\varphi} \circ \mathbf{f})_{(u,v)} = \begin{pmatrix} \dfrac{\partial\varphi}{\partial s}\dfrac{\partial f}{\partial u} + \dfrac{\partial\varphi}{\partial t}\dfrac{\partial g}{\partial u} & \dfrac{\partial\varphi}{\partial s}\dfrac{\partial f}{\partial v} + \dfrac{\partial\varphi}{\partial t}\dfrac{\partial g}{\partial v} \\[2ex] \dfrac{\partial\psi}{\partial s}\dfrac{\partial f}{\partial u} + \dfrac{\partial\psi}{\partial t}\dfrac{\partial g}{\partial u} & \dfrac{\partial\psi}{\partial s}\dfrac{\partial f}{\partial v} + \dfrac{\partial\psi}{\partial t}\dfrac{\partial g}{\partial v} \end{pmatrix}.
$$

When we write out the components of the composite map,

$$
\boldsymbol{\varphi} \circ \mathbf{f} : \begin{cases} x = \varphi(f(u,v), g(u,v)), \\ y = \psi(f(u,v), g(u,v)), \end{cases}
$$

we express x and y directly as functions of u and v; if we use $\partial x/\partial u$, and so forth, to denote the partial derivatives of these functions, then

$$
d(\boldsymbol{\varphi} \circ \mathbf{f})_{(u,v)} = \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\[2ex] \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}.
$$

We now have two formulas for the derivative $d(\boldsymbol{\varphi} \circ \mathbf{f})_{(u,v)}$. Together they give us the chain rule for the individual component functions:

$$\frac{\partial x}{\partial u}=\frac{\partial \varphi}{\partial s}\frac{\partial f}{\partial u}+\frac{\partial \varphi}{\partial t}\frac{\partial g}{\partial u}, \qquad \frac{\partial x}{\partial v}=\frac{\partial \varphi}{\partial s}\frac{\partial f}{\partial v}+\frac{\partial \varphi}{\partial t}\frac{\partial g}{\partial v},$$
$$\frac{\partial y}{\partial u}=\frac{\partial \psi}{\partial s}\frac{\partial f}{\partial u}+\frac{\partial \psi}{\partial t}\frac{\partial g}{\partial u}, \qquad \frac{\partial y}{\partial v}=\frac{\partial \psi}{\partial s}\frac{\partial f}{\partial v}+\frac{\partial \psi}{\partial t}\frac{\partial g}{\partial v}.$$

Chain rule for component functions

There are clear patterns in these four equations that allow us to see what form the component derivatives will take in the general case. We start with two variables (u and v) that first determine two others (s and t) directly and then two more (x and y) indirectly. The partial derivative of x with respect to u, for example, must have terms that take into account how x varies with u via s (viz. $\partial\varphi/\partial s\cdot\partial f/\partial u$) and via t ($\partial\varphi/\partial t\cdot\partial g/\partial u$)).

In the general case, p variables ($u_1, \ldots, u_p$) first determine the values of q new variables ($s_1, \ldots, s_q$) directly, and then r additional variables ($x_1, \ldots, x_r$) indirectly:

$$x_k=\varphi_k(s_1,\ldots,s_q) \text{ and } s_j=f_j(u_1,\ldots,u_p),$$

for $k=1,\ldots,r$ and $j=1,\ldots,q$. Thus, a partial derivative of x_k, for example, must take into account how x_k varies with each u_i via each of the q intermediate variables s_j:

$$\frac{\partial x_k}{\partial u_i}=\frac{\partial \varphi_k}{\partial s_1}\frac{\partial f_1}{\partial u_i}+\cdots+\frac{\partial \varphi_k}{\partial s_q}\frac{\partial f_q}{\partial u_i}, \quad i=1,\ldots,p.$$

The following theorem summarizes this discussion, with the single variable y replacing the various $x_1, \ldots, x_r$.

Theorem 4.10. *Suppose the functions $y=\varphi(s_1,\ldots,s_q)$ and $s_j=f_j(u_1,\ldots,u_p)$, with $j=1,\ldots,q$, are all differentiable; then*

$$\frac{\partial y}{\partial u_i}=\sum_{j=1}^{q}\frac{\partial \varphi}{\partial s_j}\frac{\partial f_j}{\partial u_i}, \quad i=1,\ldots,p. \qquad \square$$

Derivative of the inverse

The following corollary of the chain rule says that the derivative of the inverse (of a given map) is the inverse of the derivative (of that map).

Corollary 4.11 *Suppose $\mathbf{f}: U^n \to S^n$ is invertible, and $\mathbf{f}^{-1}: S^n \to U^n$ is its inverse. Suppose that both $\mathbf{f}$ and $\mathbf{f}^{-1}$ are differentiable; then*

$$(\mathrm{d}\mathbf{f}_{\mathbf{u}})^{-1}=\mathrm{d}(\mathbf{f}^{-1})_{\mathbf{f}(\mathbf{u})}.$$

Proof. Let $\mathbf{I}: U^n \to U^n$ be the identity map, $\mathbf{I}(\mathbf{u})=\mathbf{u}$. Then $\mathbf{f}^{-1}\circ\mathbf{f}=\mathbf{I}$; by the chain rule

$$I=\mathrm{d}\mathbf{I}_{\mathbf{u}}=\mathrm{d}(\mathbf{f}^{-1})_{\mathbf{f}(\mathbf{u})}\circ\mathrm{d}\mathbf{f}_{\mathbf{u}},$$

where I is the linear map represented by the $n\times n$ identity matrix. The equation implies that $\mathrm{d}(\mathbf{f}^{-1})_{\mathbf{f}(\mathbf{u})}$ is the inverse of $\mathrm{d}\mathbf{f}_{\mathbf{u}}$. $\square$

Local orientation and volume magnification

The corollary focuses our attention on maps $\mathbf{f}: U^n \to S^n$ whose source and target have the same dimension. We have already studied some examples when $n=2$ in the second section of this chapter; in particular, we saw we could use the area multiplier of the derivative to assign a local area multiplier (p. 115) to the map itself at each point. The following definition carries these ideas over to higher dimensions.

Definition 4.7 *Suppose* $\mathbf{f} : U^n \to \mathbb{R}^n$ *is differentiable; the* ***local volume multiplier of*** $\mathbf{f}$ *at* $\mathbf{a}$*, written* $J_{\mathbf{f}}(\mathbf{a})$*, is* $\det d\mathbf{f}_{\mathbf{a}}$*, the volume multiplier of the derivative of* $\mathbf{f}$ *at* $\mathbf{a}$*. Also,* $\mathbf{f}$ *is* ***orientation-preserving*** *or* ***reversing*** *at* $\mathbf{a}$ *according as its derivative* $d\mathbf{f}_{\mathbf{a}}$ *is orientation-preserving or reversing.*

Volume magnification in a chain

The chain rule implies that the local volume multiplier of a composite is the product of their individual multipliers, as we would expect.

Corollary 4.12 *If* $\mathbf{f} : U^n \to S^n$ *and* $\boldsymbol{\varphi} : S^n \to \mathbb{R}^n$ *are differentiable, then*

$$J_{\boldsymbol{\varphi}\circ\mathbf{f}}(\mathbf{a}) = J_{\boldsymbol{\varphi}}(\mathbf{f}(\mathbf{a}))\, J_{\mathbf{f}}(\mathbf{a}).$$

Proof. The proof is just a consequence of the fact that the determinant of a product is the product of the determinants:

$$\begin{aligned} J_{\boldsymbol{\varphi}\circ\mathbf{f}}(\mathbf{a}) = \det d(\boldsymbol{\varphi}\circ\mathbf{f})_{\mathbf{a}} &= \det(d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \circ d\mathbf{f}_{\mathbf{a}}) \\ &= \det d\boldsymbol{\varphi}_{\mathbf{f}(\mathbf{a})} \det d\mathbf{f}_{\mathbf{a}} = J_{\boldsymbol{\varphi}}(\mathbf{f}(\mathbf{a}))\, J_{\mathbf{f}}(\mathbf{a}). \end{aligned}$$ □

The Jacobian

The traditional name for $J_{\mathbf{f}}(\mathbf{a})$ is the **Jacobian**; hence the letter "J". In this context, the matrix $d\mathbf{f}_{\mathbf{a}}$ is the **Jacobian matrix** (the Jacobian itself is always the *determinant*). The Jacobian plays a central role in multiple integrals. We show it is the analogue of the factor $\varphi'(s)$ that appears in the transformation $dx = \varphi'(s)\,ds$ of differentials (pp. 3–5). For that reason, we write it another way (also traditional) that suggests the connection with derivatives more directly. To illustrate, let

$$\mathbf{f} : \begin{cases} x = f(u,v), \\ y = g(u,v); \end{cases}$$

then our alternate notation for the Jacobian of $\mathbf{f}$ is

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial(f,g)}{\partial(u,v)}.$$

Jacobian notation

Here we write $J(u,v)$ without the subscript for the map $\mathbf{f}$. This is frequently done, and it directs attention to the Jacobian as a function of the input variables. The second and third expressions are the more common ones; they remind us that the Jacobian involves partial derivatives. The Jacobian of the polar coordinate map is

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \dfrac{\partial}{\partial r} r\cos\theta & \dfrac{\partial}{\partial\theta} r\cos\theta \\ \dfrac{\partial}{\partial r} r\sin\theta & \dfrac{\partial}{\partial\theta} r\sin\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

This agrees with our original determination of the local area multiplier for polar coordinates (p. 115).

Definition 4.8 *Suppose* $\mathbf{f} : U^n \to \mathbb{R}^n$ *is differentiable and has component functions*

$$x_i = f_i(u_1, \ldots, u_n), \quad i = 1, \ldots, n.$$

Then the ***Jacobian*** *of* **f** *is the determinant*

$$J(u_1, \ldots, u_n) = \frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} = \frac{\partial(f_1, \ldots, f_n)}{\partial(u_1, \ldots, u_n)} = \det\left(\frac{\partial f_i}{\partial u_j}\right) = \det d\mathbf{f}_{\mathbf{u}}.$$

The following are restatements of Corollaries 4.11 and 4.12 using Jacobian notation. Because they deal with inverses and with Jacobians, it becomes practical to replace function names by names of output variables (e.g., to replace $s_i = f_i(u_1, \ldots, u_n)$ by $s_i = s_i(u_1, \ldots, u_n)$).

Corollary 4.13 *If a map and its inverse are both differentiable,*

$$\mathbf{f} : \begin{cases} s_1 = s_1(u_1, \ldots, u_n), \\ \quad \vdots \\ s_n = s_n(u_1, \ldots, u_n), \end{cases} \qquad \mathbf{f}^{-1} : \begin{cases} u_1 = u_1(s_1, \ldots, s_n), \\ \quad \vdots \\ u_n = u_n(s_1, \ldots, s_n), \end{cases}$$

then

$$\frac{\partial(u_1, \ldots, u_n)}{\partial(s_1, \ldots, s_n)} = \frac{1}{\dfrac{\partial(s_1, \ldots, s_n)}{\partial(u_1, \ldots, u_n)}}. \qquad \square$$

Chain rule for Jacobians

Corollary 4.14 *If the following are differentiable,*

$$\boldsymbol{\varphi} : \begin{cases} x_1 = x_1(s_1, \ldots, s_n), \\ \quad \vdots \\ x_n = x_n(s_1, \ldots, s_n), \end{cases} \qquad \mathbf{f} : \begin{cases} s_1 = s_1(u_1, \ldots, u_n), \\ \quad \vdots \\ s_n = s_n(u_1, \ldots, u_n), \end{cases}$$

then

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} = \frac{\partial(x_1, \ldots, x_n)}{\partial(s_1, \ldots, s_n)} \, \frac{\partial(s_1, \ldots, s_n)}{\partial(u_1, \ldots, u_n)}. \qquad \square$$

These results obviously remind us of the one-variable cases: if $u = u(s)$ is the inverse of $s = s(u)$, and $x = x(s)$, then

$$\frac{du}{ds} = \frac{1}{ds/du} \quad \text{and} \quad \frac{dx}{du} = \frac{dx}{ds}\,\frac{ds}{du}.$$

Local area multiplier on a surface patch

Although Jacobians are determinants and consequently involve an equal number of input and output variables, they can appear in other circumstances. For example, the parametrization of a surface patch, $\mathbf{f}: U^2 \to \mathbb{R}^3$, involves three functions of two real variables:

$$\mathbf{f}: \begin{cases} x = f(u,v), \\ y = g(u,v), \\ z = h(u,v). \end{cases}$$

The linear map $d\mathbf{f}_{\mathbf{a}}: \mathbb{R}^2 \to \mathbb{R}^3$ has an area magnification factor that is a kind of "Pythagorean formula" (Theorem 2.25, p. 55). The formula involves the 2×2 minors of $d\mathbf{f}_{\mathbf{a}}$, which can be written in a simple and direct way using Jacobians; the result is given in the following definition.

Definition 4.9 *The* ***local area multiplier*** *for the parametrized surface patch* $\mathbf{f}: U^2 \to \mathbb{R}^3 : (u,v) \mapsto (x,y,z)$ *is*

$$M(u,v) = \sqrt{\left[\frac{\partial(y,z)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(z,x)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^2}.$$

Example: the crosscap

For example, the crosscap we analyzed earlier (pp. 126–128) has the parametrization

$$\mathbf{f}: \begin{cases} x = u, \\ y = uv, \\ z = -v^2, \end{cases} \qquad d\mathbf{f}_{\mathbf{u}} = \begin{pmatrix} 1 & 0 \\ v & u \\ 0 & -2v \end{pmatrix},$$

so the local area multiplier is

$$\sqrt{\begin{vmatrix} v & u \\ 0 & -2v \end{vmatrix}^2 + \begin{vmatrix} 0 & -2v \\ 1 & 0 \end{vmatrix}^2 + \begin{vmatrix} 1 & 0 \\ v & u \end{vmatrix}^2} = \sqrt{4v^4 + 4v^2 + u^2}.$$

At $(u,v) = (1,0)$, the multiplier is 1. This agrees with what we saw earlier: near $(1,0)$, $\mathbf{f}$ preserves areas. At $(u,v) = (-1,1)$, the multiplier is 3. This too agrees with our earlier analysis: near $(-1,1)$, $\mathbf{f}$ triples areas.

A mean-value theorem for maps

The chain rule gives us a way to prove the mean-value theorem for maps of the form $\mathbf{f}: U^p \to \mathbb{R}^q$. The mean-value theorem says that, for any two points $\mathbf{a}$ and $\mathbf{b}$ in U^p,

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| \leq M\|\mathbf{b} - \mathbf{a}\|,$$

where M is a bound on the size of the derivative of $\mathbf{f}$ at points along the line from $\mathbf{a}$ to $\mathbf{b}$. We need to establish what the "size" of the derivative is.

The *norm* of a derivative

In Theorem 3.8 and the discussion preceding it (pp. 76–77), we had $q = 1$. Therefore we were able to identify the derivative with a vector and the size of the derivative with the magnitude of that vector. When $q \geq 2$, the derivative is a more general linear map; to measure its size, we use its *norm* (see Exercise 3.28.b, p. 104). The norm of $d\mathbf{f}_{\mathbf{u}}$ is

$$|||d\mathbf{f}_{\mathbf{u}}||| = \max_{\|\Delta\mathbf{v}\|=1} \|d\mathbf{f}_{\mathbf{u}}(\Delta\mathbf{v})\|;$$

we have $\|d\mathbf{f_u}(\Delta\mathbf{v})\| \le |||d\mathbf{f_u}|||\,\|\Delta\mathbf{v}\|$ for all $\Delta\mathbf{v}$ in $\mathbb{R}^p$. The norm of a linear map is the largest amount by which it stretches any vector.

Theorem 4.15 (Mean-value theorem). *Suppose the map $\mathbf{f}: U^p \to \mathbb{R}^q$ is continuously differentiable and U^p contains every point on the line segment from $\mathbf{a}$ to $\mathbf{b}$. Then*

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| \le \max_{\mathbf{u}} |||d\mathbf{f_u}|||\,\|\mathbf{b}-\mathbf{a}\|,$$

where the maximum is taken over all points $\mathbf{u}$ on the line from $\mathbf{a}$ to $\mathbf{b}$.

Write the difference as an integral

Proof. We begin by constructing the analogue of the error formula

$$\int_0^1 f'(a+t\Delta x)\,\Delta x\,dt = f(a+\Delta x) - f(a)$$

with which we began the discussion of Taylor's theorem (cf. pp. 78–79). Set $\Delta\mathbf{u} = \mathbf{b}-\mathbf{a}$; then $\mathbf{u}(t) = \mathbf{a}+t\Delta\mathbf{u}$, $0 \le t \le 1$, is the line segment from $\mathbf{a}$ to $\mathbf{b}$. All the points on this segment are also in U^p; therefore the map

$$\boldsymbol{\varphi}(t) = \mathbf{f}(\mathbf{u}(t)) = \mathbf{f}(\mathbf{a}+t\Delta\mathbf{u})$$

is continuously differentiable on $[0,1]$. By the chain rule,

$$\boldsymbol{\varphi}'(t) = d\boldsymbol{\varphi}_t = d\mathbf{f}_{\mathbf{u}(t)}(d\mathbf{u}_t) = d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u});$$

we have used $d\mathbf{u}_t = \mathbf{u}'(t) = \Delta\mathbf{u}$. Thus

$$\begin{aligned}\int_0^1 d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u})\,dt &= \int_0^1 \boldsymbol{\varphi}'(t)\,dt = \boldsymbol{\varphi}(1)-\boldsymbol{\varphi}(0)\\ &= \mathbf{f}(\mathbf{a}+\Delta\mathbf{u}) - \mathbf{f}(\mathbf{a}) = \mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a}).\end{aligned}$$

This is, in fact, just Taylor's formula with remainder in degree 0 for the map $\mathbf{f}$. It implies

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| = \left\|\int_0^1 d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u})\,dt\right\| \le \int_0^1 \|d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u})\|\,dt.$$

Because $\Delta\mathbf{u} = \mathbf{b}-\mathbf{a}$ is fixed, we have

$$\|d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u})\| \le \max_{0\le t\le 1} \|d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}(\Delta\mathbf{u})\| \le \max_{0\le t\le 1} |||d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}|||\,\|\Delta\mathbf{u}\|.$$

The right-hand side is independent of t, so (with $\mathbf{u} = \mathbf{a}+t\Delta\mathbf{u}$)

$$\|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})\| \le \max_{0\le t\le 1} |||d\mathbf{f}_{\mathbf{a}+t\Delta\mathbf{u}}|||\,\|\Delta\mathbf{u}\| \le \max_{\mathbf{u}} |||d\mathbf{f_u}|||\,\|\mathbf{b}-\mathbf{a}\|. \qquad \square$$

Exercises

4.1. Determine the derivative of the given function $z = f(x,y)$ at the given point $(x,y) = (a,b)$. Write the derivative as a linear function of the variables $\Delta x = x - a$, $\Delta y = y - b$.

a. $f(x,y) = 7x - 3y + 9$, $(a,b) = (4,5)$.
b. $f(x,y) = 1 - \cos x + y^2/2$, $(a,b) = (0,1)$.
c. $f(x,y) = \arctan(y/x)$, $(a,b) = (4,-3)$.
d. $f(x,y) = \alpha x^2 + 2\beta xy + \gamma y^2 + \delta x + \varepsilon y + \kappa$, (a,b) arbitrary (and α, β, γ, δ, ε, κ are all constants).

4.2. a. Let $f(x,y) = x^2 - y^2$. Plot the graph of $z = f(x,y)$ in a window centered at $(x,y) = (2,-1)$, making the window small enough for the graph to appear to be a flat plane.
b. Confirm that the plane appearing in (a) is the graph of the derivative $df_{(2,-1)}$. Because the derivative is expressed in terms of the displacement variables $\Delta x = x - 2$, $\Delta y = y + 1$, it is necessary to include the constant term $f(2,-1) = 3$ in the equation for the second graph.
c. Construct a contour plot of $f(x,y)$ in two windows centered at $(x,y) = (2,-1)$. Make the first window 2.0 units on a side, and make the second 0.2. Confirm that f changes its appearance from nonlinear to linear from the first to the second window, and confirm that the level curves of f are indistinguishable from the level curves of its derivative there.

4.3. a. Determine the equation of the tangent plane to the graph of $z = f(x,y) = \sin x \sin y$ at the point $(x,y) = (\pi/3, -\pi/2)$.
b. Sketch, together, the graph of f and this tangent plane over the square $0 \le x \le \pi$, $-\pi \le y \le 0$.
c. Select a smaller square centered at $(\pi/3, -\pi/2)$ on which the graph of f and this tangent plane become indistinguishable; sketch the two surfaces over that square.
d. Sketch together contour plots of f and the derivative of f at $(\pi/3, -\pi/2)$ on the square $0 \le x \le \pi$, $-\pi \le y \le 0$. Sketch them again in the smaller square you selected in part (c).

4.4. a. Explain what $\cos t - 1 = o(1)$ means, and then show that it is true. (Suggestion: Use l'Hôpital's rule.)
b. Is $\sin t = o(1)$ true? Explain.
c. Explain what $\sin t - t = o(2)$ means, and then show that it is true. Is it true that $\sin t - t = o(3)$? Explain. Is $\sin t - t = O(3)$ true? Explain.

4.5. Show $f(x) = x^2 \sin(1/x)$ is differentiable at $x = 0$, that $f'(0) = 0$, and even that $f(\Delta x) = f(0) + f'(0)\Delta x + O(2)$. Show that, nevertheless, $f''(0)$ does not exist. Moreover, show that $f'(x)$ is not a continuous function: If $x_n = 1/(2\pi n)$, then $f'(x_n) = 1$; however, $f'(x_n) \not\to f'(0)$ even though $x_n \to 0$.

4.6. Let $z = f(x,y)$ be the "manta ray" function (pp. 108–109). Show analytically that $z = 0$ is not the tangent plane to f at the origin by showing the gap $f(\Delta x, \Delta y) - f(0,0)$ does not vanish to order greater than 1; that is, show directly (cf. Definition 3.14, p. 98) that the ratio

$$\frac{f(\Delta x, \Delta y) - f(0,0)}{\|(\Delta x, \Delta y)\|} = \frac{(\Delta x)^2 \Delta y / ((\Delta x)^2 + (\Delta y)^2)}{\sqrt{(\Delta x)^2 + (\Delta y)^2}}$$

does not have the limit 0 as $(\Delta x, \Delta y) \to (0,0)$.

4.7. Let $f(0,0) = 0$, $f(x,y) = \dfrac{2xy}{\sqrt{x^2+y^2}}$ for $(x,y) \neq (0,0)$.

a. Sketch the graph of $z = f(x,y)$ near the origin. Use a polar coordinate overlay to clarify the picture.
b. Show that the partial derivatives $f_x(0.0)$ and $f_y(0,0)$ exist, and determine their values.
c. Show that the directional derivative $D_{\mathbf{u}} f(0,0)$ does not exist if $\mathbf{u}$ is not an axis direction. Explain this result in terms of the graph of f.
d. Conclude that f, like the "manta ray" counterexample, fails to be differentiable at the origin. (In fact, for both this function and the "manta ray," $f(x,y) = O(1)$ is true but $f(x,y) = o(1)$ is false.)

4.8. Let $f(0,0) = 0$, $f(x,y) = \dfrac{x^3 y}{x^4 + y^2}$ for $(x,y) \neq (0,0)$.

a. Sketch the graph of $z = f(x,y)$ near the origin. (A polar coordinate overlay is not as helpful here; it does not simplify our view of the graph.)
b. Show that the directional derivative $D_{\mathbf{u}} f(0,0)$ exists and equals 0 in every direction $\mathbf{u}$. (In particular, $f_x(0,0) = f_y(0,0) = 0$; therefore, if f were differentiable at $(0,0)$, the tangent plane to its graph at the origin would be the (x,y)-plane.)
c. Compute the partial derivatives $f_x(x,y)$ and $f_y(x,y)$ at any arbitrary point, and show they are not continuous at $(x,y) = (0,0)$.
d. Add to your sketch in part (a) the curve $z = f(x,x^2)$ in the graph of f that lies over the parabola $y = x^2$. Show that $z = x/2$ along the parabola, implying that f vanishes exactly to order 1 on the parabola. Conclude that f cannot be differentiable at the origin.

4.9. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$\mathbf{f} : \begin{cases} x = au + bv + k, \\ y = cu + dv + l, \end{cases} \qquad \mathbf{g} : \begin{cases} r = \alpha x + \beta y + \kappa, \\ s = \gamma x + \delta y + \lambda. \end{cases}$$

a. Determine the derivative $d\mathbf{f}_{\mathbf{u}}$ at an arbitrary point $\mathbf{u} = (u, v)$. Does the derivative depend on (u, v)? How is $d\mathbf{f}_{\mathbf{u}}$ related to $\mathbf{f}$?
b. Determine $d\mathbf{g}_{\mathbf{x}}$ and then use the (components of the) substitution $\mathbf{x} = \mathbf{f}(\mathbf{u})$ to express the derivative in terms of $\mathbf{u}$: $d\mathbf{g}_{\mathbf{x}} = d\mathbf{g}_{\mathbf{f}(\mathbf{u})}$.
c. Compute the components of the composite map $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$. That is, determine r and s in terms of u and v; $\mathbf{h}(\mathbf{u}) = \mathbf{g}(\mathbf{f}(\mathbf{u}))$.
d. Determine $d\mathbf{h}_{\mathbf{u}}$ and verify that $d\mathbf{h}_{\mathbf{u}} = d\mathbf{g}_{\mathbf{f}(\mathbf{u})} \cdot d\mathbf{f}_{\mathbf{u}}$. (This is the chain rule; see Theorem 4.9, p. 132.)

4.10. Let $\mathbf{f} : U^2 \to \mathbb{R}^2$ be the polar coordinate map, and let $\mathbf{f}^{-1} : \mathbb{R}^2 \to U^2$ be its inverse:

$$\mathbf{f} : \begin{cases} x = r\cos\theta, \\ y = r\sin\theta, \end{cases} \qquad \mathbf{f}^{-1} : \begin{cases} r = \sqrt{x^2 + y^2}, \\ \theta = \arctan(y/x). \end{cases}$$

a. Compute the matrix of the derivative $d\mathbf{f}^{-1}_{\mathbf{x}}$ at an arbitrary point $\mathbf{x} = (x, y)$.
b. Compute the inverse matrix $(d\mathbf{f}_{\mathbf{r}})^{-1}$ at an arbitrary point $\mathbf{r} = (r, \theta)$.
c. Use the coordinate change $\mathbf{x} = \mathbf{f}(\mathbf{r})$ to express $d\mathbf{f}^{-1}_{\mathbf{x}}$ in terms of $\mathbf{r}$: $d\mathbf{f}^{-1}_{\mathbf{x}} = d\mathbf{f}^{-1}_{\mathbf{f}(\mathbf{r})}$. Then verify that $(d\mathbf{f}_{\mathbf{r}})^{-1} = d\mathbf{f}^{-1}_{\mathbf{f}(\mathbf{r})}$. (The derivatives of inverse maps are themselves inverses of each other. To see this it is first necessary, however, to express them in terms of the same variables. See Theorem 4.11, p. 136.)
d. Compute the determinants

$$\det d\mathbf{f}^{-1}_{\mathbf{x}} \quad \text{and} \quad \det d\mathbf{f}_{\mathbf{r}},$$

and show that they are reciprocals. Use an appropriate change of variables on one of the expressions to compare the two.

4.11. Let $\mathbf{x} = \mathbf{f}(\mathbf{r})$ be the polar coordinate map of the previous exercise, and let $\mathbf{u} = \mathbf{g}(\mathbf{x})$ be the map

$$\mathbf{g} : \begin{cases} u = x^2 + y^2, \\ v = x^2 - y^2. \end{cases}$$

a. Determine $d\mathbf{g}_{\mathbf{x}}$ at an arbitrary point $\mathbf{x} = (x, y)$; then use the coordinate change $\mathbf{x} = \mathbf{f}(\mathbf{r})$ to express the derivative in terms of $\mathbf{r} = (r, \theta)$: $d\mathbf{g}_{\mathbf{f}(\mathbf{r})}$.
b. Compute the components of the composite map $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$. That is, determine u and v in terms of r and θ; $\mathbf{u} = \mathbf{h}(\mathbf{r}) = \mathbf{g}(\mathbf{f}(\mathbf{r}))$.
c. Verify that

$$d\mathbf{h}_{\mathbf{r}} = \begin{pmatrix} 2r & 0 \\ 2r\cos 2\theta & -r^2 \sin 2\theta \end{pmatrix}$$

and also verify that $d\mathbf{h}_{\mathbf{r}} = d\mathbf{g}_{\mathbf{f}(\mathbf{r})} \cdot d\mathbf{f}_{\mathbf{r}}$. (This is another instance of the chain rule.)

In Exercises 4.12–4.20, $\mathbf{f}: \mathbb{R}^2 \to \mathbb{R}^2$ is the quadratic map discussed in the text:

$$\mathbf{f}: \begin{cases} x = u^2 - v^2, \\ y = 2uv. \end{cases}$$

4.12. Compute $\|\mathbf{f}(\Delta\mathbf{u})\|^2$ to show that $\|\mathbf{f}(\Delta\mathbf{u})\| = \|\Delta\mathbf{u}\|^2$. Note: this means that $\mathbf{f}(\Delta\mathbf{u})$ vanishes exactly to order 2 (by the extension of Definition 3.4 used in Exercise 3.28). That is, there are positive constants C_1, C_2 for which

$$C_1 \leq \frac{\|\mathbf{f}(\Delta\mathbf{u})\|}{\|\Delta\mathbf{u}\|^2} \leq C_2$$

for all $\Delta\mathbf{u} \neq \mathbf{0}$. (It is evident we can take $C_1 = C_2 = 1$.)

4.13. Let $\mathcal{U}^2 = \{(x,y) \mid y > 0\}$ be the upper half-plane, and let $\mathbf{g}_+ : \mathcal{U}^2 \to \mathbb{R}^2$ be the map

$$\mathbf{g}_+ : \begin{cases} u = \sqrt{\dfrac{\sqrt{x^2+y^2}+x}{2}}, \\[2ex] v = \sqrt{\dfrac{\sqrt{x^2+y^2}-x}{2}}. \end{cases}$$

a. Show that $\mathbf{f}(\mathbf{g}_+(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathcal{U}^2$. In other words, $\mathbf{g}_+$ is a (partial) inverse for $\mathbf{f}$.

b. Describe the action of $\mathbf{g}_+$ in terms of polar coordinate overlays on the source and target (cf. pp. 116–121). In other words, describe what happens to the angle a point makes with the positive horizontal axis change, and what happens to its distance from the origin change.

c. Describe the image of $\mathbf{g}_+(\mathcal{U}^2)$

4.14. Show that $\mathbf{g}_+(\mathbf{x})$ (Exercise 4.13) can be extended to the two sides of the x-axis ($y = 0$) as follows:

$$\begin{matrix} u = \sqrt{x}, \\ v = 0, \end{matrix} \quad \text{if } x \geq 0; \qquad \begin{matrix} u = 0, \\ v = \sqrt{|x|}, \end{matrix} \quad \text{if } x \leq 0.$$

What, therefore, is the image of the x-axis under this extension of $\mathbf{g}_+$?

4.15. a. Compute $d(\mathbf{g}_+)_\mathbf{x}$, where $\mathbf{g}_+$ is the map of Exercise 4.13.

b. Use the coordinate change $x = u^2 - v^2$, $y = 2uv$ (provided by the inverse map $\mathbf{f}$) to express $d(\mathbf{g}_+)_\mathbf{x}$ in terms of u and v, giving $d(\mathbf{g}_+)_{\mathbf{f}(\mathbf{u})}$.

c. Verify that $d(\mathbf{g}_+)_{\mathbf{f}(\mathbf{u})}$ is the inverse of $d\mathbf{f}_\mathbf{u}$.

4.16. The object of this exercise is to study the action of $\mathbf{g}_+$ (Exercise 4.13) in a microscope window centered at $\mathbf{x}_0 = (1/2, \sqrt{3}/2)$.

a. Determine the center $\mathbf{g}_+(\mathbf{x}_0)$ of the target window.

b. Explain why $\mathbf{g}_+$ maps the 60°-line in the window at $\mathbf{x}_0$ to the 30°-line in the target window. (Suggestion: Use results from Exercise 4.13.)

c. Show that $d(\mathbf{g}_+)_{\mathbf{x}_0} = \lambda R_\theta$, for certain $\lambda > 0$ and $\theta < 0$; R_θ is rotation by θ radians. What are the values of λ and θ? Does $\mathbf{g}_+$ "look like" its derivative $d(\mathbf{g}_+)_{\mathbf{x}_0}$ in this microscope window? Explain, in terms of what you know about the action of $\mathbf{g}_+$.

4.17. Let $\mathcal{L}^2 = \{(x,y) \mid y < 0\}$ be the lower half-plane, and let $\mathbf{g}_- : \mathcal{L}^2 \to \mathbb{R}^2$ be the map

$$\mathbf{g}_- : \begin{cases} u = -\sqrt{\dfrac{\sqrt{x^2+y^2}+x}{2}}, \\ v = \sqrt{\dfrac{\sqrt{x^2+y^2}-x}{2}}. \end{cases}$$

Note that $\mathbf{g}_-$ differs from $\mathbf{g}_+$ only in the sign of u.

a. Show that $\mathbf{f}(\mathbf{g}_-(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathcal{L}^2$. In other words, $\mathbf{g}_-$ is also a partial inverse for $\mathbf{f}$.

b. Describe the action of $\mathbf{g}_-$ in terms of polar coordinate overlays on the source and target (cf. Exercise 4.13).

c. Determine the image $\mathbf{g}_-(\mathcal{L}^2)$.

4.18. a. Show that $\mathbf{g}_-(\mathbf{x})$ can be extended to the two sides of the x-axis ($y = 0$) as follows:

$$\begin{matrix} u = -\sqrt{x}, \\ v = 0, \end{matrix} \quad \text{if } x \geq 0; \qquad \begin{matrix} u = 0, \\ v = \sqrt{|x|}, \end{matrix} \quad \text{if } x \leq 0.$$

What, therefore, is the image of the x-axis under this extension of $\mathbf{g}_-$?

b. Show that $\mathbf{g}_+$ and $\mathbf{g}_-$ agree on the negative x-axis but disagree on the positive x-axis. Excluding, therefore, the positive x-axis from the domain of $\mathbf{g}_-$, explain how $\mathbf{g}_+$ and $\mathbf{g}_-$ together define a single map on the whole plane $\mathbb{R}^2$ that serves as an inverse for $\mathbf{f}$. Show that this combined map is not continuous across the positive x-axis.

4.19. a. Compute $d(\mathbf{g}_-)_{\mathbf{x}}$, where $\mathbf{g}_-$ is given in Exercise 4.17.

b. Use the coordinate change $x = u^2 - v^2$, $y = 2uv$ (provided by the inverse map $\mathbf{x} = \mathbf{f}(\mathbf{u})$) to express $d(\mathbf{g}_-)_{\mathbf{x}}$ in terms of $\mathbf{u} = (u,v)$, giving $d(\mathbf{g}_-)_{\mathbf{f}(\mathbf{u})}$.

c. Verify that $d(\mathbf{g}_-)_{\mathbf{f}(\mathbf{u})}$ is the inverse of $d\mathbf{f}_{\mathbf{u}}$.

4.20. The object of this exercise is to study the action of $\mathbf{g}_-$ (Exercise 4.17) in a microscope window centered at $\mathbf{x}_0 = (0, -9/4)$.

a. Determine the center $\mathbf{g}_-(\mathbf{x}_0)$ of the target window.

b. Explain why $\mathbf{g}_-$ maps the 270°-line in the window at $\mathbf{x}_0$ to the 135°-line in the target window.

c. Show that $d(\mathbf{g}_-)_{\mathbf{x}_0} = \lambda R_\theta$, for certain $\lambda > 0$ and $\theta < 0$. What are the values of λ and θ? Does $\mathbf{g}_-$ "look like" its derivative $d(\mathbf{g}_-)_{\mathbf{x}_0}$ in this microscope window? Explain, in terms of what you know about the action of $\mathbf{g}_-$.

4.21. Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2 : (u,v) \mapsto (x,y)$ be the map defined by

$$\mathbf{f} : \begin{cases} x = u, \\ y = v^2. \end{cases}$$

a. Describe the global behavior of $\mathbf{f}$ by showing what happens to an ordinary Cartesian grid in the source. In particular, indicate the effect of the equation $y = v^2$. This map is sometimes called a *fold*; your picture should explain why.
b. Determine the derivative $d\mathbf{f}_{(a,b)}$ at each point $(u,v) = (a,b)$.
c. Show that the local area multiplier of $\mathbf{f}$ at (a,b) is b. Hence the local area multiplier along the horizontal axis (the u-axis) is 0; why? What feature of the map $\mathbf{f}$ does this reflect?
d. Sketch the effect of $\mathbf{f}$ in a microscope window centered at $(u,v) = (3,2)$, and indicate how this corresponds to the effect of the derivative $d\mathbf{f}_{(3,2)}$.
e. Sketch the effect of $\mathbf{f}$ in a microscope window centered at $(u,v) = (3,0)$, and indicate how this corresponds to the effect of the derivative $d\mathbf{f}_{(3,0)}$. (Note: the local area multiplier here is 0, and the derivative $d\mathbf{f}_{(3,0)}$ is non-invertible. Thus we do not expect $\mathbf{f}$ to look like $d\mathbf{f}_{(3,0)}$ in a microscope window centered at $(3,0)$.)

4.22. a. Obtain the derivative $d\mathbf{q}_{\mathbf{a}}$ of the map

$$\mathbf{q} : \begin{cases} x = u^3 - 3uv^2, \\ y = 3u^2v - v^3. \end{cases}$$

b. Show that $d\mathbf{q}_{(a,b)}$ is a similarity transformation (cf. p. 118), that is, a rotation by an angle θ combined with a uniform dilation by a factor λ. Determine θ and λ in terms of a and b. Conclude that $\mathbf{q}$ is conformal on the whole plane minus the origin. Why must the origin be excluded?
c. Use a polar coordinate overlay to create a description of the action of $\mathbf{q}$ that is analogous to the description of the quadratic map (as one that doubles angles and squares distances from the origin).

4.23. Repeat all the steps of the previous exercise for the map

$$\mathbf{s} : \begin{cases} x = u^4 - 6u^2v^2 + v^4, \\ y = 4u^3v - 4uv^3. \end{cases}$$

In particular, find $\lambda = 4(a^2+b^2)^{3/2}$ and $\theta = \arctan\left(\dfrac{3a^2b - b^3}{a^3 - a3b^2}\right)$.

4.24. a. Show that the local area magnification factor at the parameter point (θ, φ) on the unit sphere is $\cos\varphi$.

b. Show that the arc length of the "parallel of latitude" at latitude φ_0 is $2\pi\cos\varphi_0$. Show that the arc length of a "meridian of longitude" at longitude θ_0 is π, independent of θ_0.

4.25. a. Compute the matrix of the derivative $\mathbf{df}_{(\pi/3,\pi/6)}$ of the unit sphere map $\mathbf{f}$, and describe its image in $\mathbb{R}^3$.

b. Sketch the image of the unit sphere map $\mathbf{f}$ in a microscope window at the point $(\pi/3, \pi/6)$. Show that the window can be made small enough so the image is indistinguishable from the image of the derivative $\mathbf{df}_{(\pi/3,\pi/6)}$.

4.26. Show that the local area magnification factor at the parameter point (p,q) on the crosscap is $\sqrt{p^2+4q^2+4q^4}$. Confirm that the local area magnification factors at the points **a**, **b**, and **c** discussed in the text are, respectively, 1, 3, and 0.

4.27. Show that the window equation

$$\Delta\mathbf{x} = \mathbf{f}(\mathbf{p}+\Delta\mathbf{u}) - \mathbf{f}(\mathbf{p}) = d\mathbf{f}_{\mathbf{p}}(\Delta\mathbf{u}) + \mathbf{R}_{1,\mathbf{p}}(\Delta\mathbf{u})$$

for the crosscap parametrization $\mathbf{f}$ at an arbitrary point $\mathbf{p} = (p,q)$ can be written with the remainder $\mathbf{R}_{1,\mathbf{p}}$ explicitly as

$$\begin{array}{rcccl}
\Delta\mathbf{x} = & \mathbf{df}_{\mathbf{p}}(\Delta\mathbf{u}) & + & \mathbf{R}_{1,\mathbf{p}}(\Delta\mathbf{u}), \\
\Delta x = & \Delta u & + & 0, \\
\Delta y = & q\,\Delta u + p\,\Delta v & + & \Delta u\,\Delta v, \\
\Delta z = & -2q\,\Delta v & + & -(\Delta v)^2.
\end{array}$$

Note that $\mathbf{R}_{1,\mathbf{p}}(\Delta\mathbf{u})$ is purely quadratic in $\Delta\mathbf{u}$ and is independent of $\mathbf{p}$.

In exercises 4.28–4.30, the map $\mathbf{t}_{R,a} : U^2 \to \mathbb{R}^3$ (where $0 < a < R$) parametrizes a **torus**:

$$\mathbf{t}_{R,a} : \begin{cases} x = (R + a\cos\varphi)\cos\theta, \\ y = (R + a\cos\varphi)\sin\theta, \\ z = a\sin\varphi, \end{cases} \qquad \begin{array}{c} 0 \le \theta \le 2\pi, \\ -\pi \le \varphi \le \pi. \end{array}$$

4.28. Make a sketch of the entire image of $\mathbf{t}_{3,1}$. From this, describe what R and a measure on the torus. What happens to the shape of the torus if $R < a$?

4.29. a. Compute the (matrix of the) derivative $d(\mathbf{t}_{R,a})_{(\theta,\varphi)}$ at an arbitrary point (θ, φ) and for arbitrary R and a.

b. Determine the local area magnification factor of $\mathbf{t}_{R,a}$ at an arbitrary point (θ, φ), and confirm that it is independent of θ. For which value of φ is the factor largest, and for which is it smallest? Is this consistent with your sketch of the action of $\mathbf{t}_{3,1}$?

4.30. a. Sketch the image of $\mathbf{t}_{3,1}$ in a microscope window at $(\theta,\varphi) = (\pi/4,\pi/3)$, and compare it to the image of the derivative of $\mathbf{t}_{3,1}$ at that point.

b. Do the same at the point $(\theta,\varphi) = (0,\pi/2)$.

4.31. Let $x = g(v) > 0$ be a smooth function for $a \le v \le b$. The surface parametrized as

$$\mathbf{r} : \begin{cases} x = g(v)\cos\theta, \\ y = g(v)\sin\theta, \\ z = v, \end{cases} \qquad \begin{array}{c} 0 \le \theta \le 2\pi, \\ a \le v \le b, \end{array}$$

is a **surface of revolution**. The curve $x = g(z)$ is called its **generator**.

a. Determine the derivative $d\mathbf{r}_{(\theta,v)}$ at an arbitrary point, and determine the local area magnification factor.

b. Confirm that the local area magnification factor is independent of θ, and is smallest where $g(v)$ has its minimum.

c. Show that when $g(v) = 2 + \sin kv$, and k is properly chosen, the largest local area magnification factor does not occur where $g(v)$ has its maximum.

4.32. Suppose $y = f(x)$ is differentiable at $x = a$; show that $df_a(\Delta x) = f'(a)\cdot\Delta x$.

4.33. Show that a linear map is differentiable everywhere, and is its own derivative: if $\mathbf{L} : \mathbb{R}^p \to \mathbb{R}^q$ is linear then $d\mathbf{L}_\mathbf{a} = \mathbf{L}$, for every $\mathbf{a}$ in $\mathbb{R}^p$.

4.34. Let $\mathbf{f} : U^2 \to \mathbb{R}^2 : (\rho,\varphi) \mapsto (u,v)$ be the polar coordinate map (p. 134), and let

$$\mathbf{f}^{-1} : \begin{cases} \rho = \sqrt{u^2+v^2}, \\ \varphi = \arctan(y/x), \end{cases}$$

be its inverse. Show that their derivatives are inverses; that is, show that

$$d(\mathbf{f}^{-1})_{\mathbf{f}(\rho,\varphi)} = (d\mathbf{f}_{(\rho,\varphi)})^{-1}.$$

Note that, for the equality to hold, the two sides must be expressed in terms of the same variables; thus, the derivative $d\mathbf{f}^{-1}_{(u,v)}$ must be determined at the point $(u,v) = (\rho,\varphi) = (\rho\cos\varphi,\rho\sin\varphi)$.

4.35. Determine the Jacobians $\partial(x,y)/\partial(u,v)$ and $\partial(u,v)/\partial(x,y)$ when

$$x = u^3 - 3uv^2,$$
$$y = 3u^2v - v^3.$$

4.36. Let $\varphi(t) = \Phi(x(t),y(t))$, where Φ is differentiable and $x = x(t)$ and $y = y(t)$ be differentiable functions of t.

a. Verify that $\varphi'(t) = \Phi_x(x(t),y(t))x'(t) + \Phi_y(x(t),y(t))y'(t)$, or, more briefly, $\varphi' = \operatorname{grad}\Phi\cdot(x',y')$.

b. Suppose Φ is a *potential* for $\mathbf{F}$: $\mathbf{F}(\mathbf{x}) = \operatorname{grad}\Phi(\mathbf{x})$ (cf. p. 25), and $\vec{C}$ is an oriented curve. Fill in the details to show that

$$\int_{\vec{C}} \mathbf{F}\cdot d\mathbf{x} = \int_a^b d\varphi = \Phi(\mathbf{x})\Big|_{\text{start of }\vec{C}}^{\text{end of }\vec{C}}.$$

4.37. Suppose $\mathbf{f} : U^2 \to \mathbb{R}^2 : (u,v) \mapsto (x(u,v), y(u,v))$ is a continuously differentiable map, not necesarily invertible. Let $\vec{C}$ be a piecewise-smooth oriented curve in U^2 for which $\mathbf{f}(\vec{C})$ is also piecewise smooth and oriented (cf. p. 9). Assume $\vec{C}$ and $\mathbf{f}(\vec{C})$ have a common decomposition into smooth oriented curves:

$$\vec{C} = \vec{C}_1 + \cdots + \vec{C}_m, \quad \mathbf{f}(\vec{C}) = \mathbf{f}(\vec{C}_1) + \cdots + \mathbf{f}(\vec{C}_m);$$

each $\vec{C}_i$ and $\mathbf{f}(\vec{C}_i)$ is either a simple closed curve or a simple curve (i.e., no self-intersections). If $\mathbf{u} = \mathbf{u}_i(t)$, $a_i \le t \le b_i$ is a continuously differentiable parametrization of $\vec{C}_i$, $i = 1,\ldots,m$, then $\mathbf{x} = \mathbf{f}(\mathbf{u}_i(t))$ is a continuously differentiable parametrization of $\mathbf{f}(\vec{C}_i)$.

a. Let $P(x,y)$ and $Q(x,y)$ be continuously differentiable functions defined on $\mathbf{f}(U^2)$; show that

$$\int_{\mathbf{f}(\vec{C}_i)} P\,dx + Q\,dy = \int_{\vec{C}_i} (P^* x_u + Q^* y_u)\,du + (P^* x_v + Q^* y_v)\,dv.$$

Here, $P^* = P^*(u,v) = P(x(u,v), y(u,x))$, $x_u = \partial x/\partial u$, and so forth. The equation describes how the path integral on the left is transformed into the one on the right by the change of variables $(x,y) = \mathbf{f}(u,v)$.

b. Deduce that

$$\int_{\mathbf{f}(\vec{C})} P\,dx + Q\,dy = \int_{\vec{C}} (P^* x_u + Q^* y_u)\,du + (P^* x_v + Q^* y_v)\,dv.$$

4.38. Let $\mathbf{f} : (r,\theta) \mapsto (x,y)$ be the polar coordinate map, and let $\vec{C}$ be any continuously differentiable oriented curve in the (r,θ)-plane with $r > 0$. Determine how the path integral

$$I = \int_{\mathbf{f}(\vec{C})} \frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$$

is transformed by polar coordinates. Use the transformed integral to show that

$$I = \Delta\theta = \theta\Big|_{\text{start of }\mathbf{f}(\vec{C})}^{\text{end of }\mathbf{f}(\vec{C})}.$$

4.39. Let $\mathbf{f} : (x,y) = (u^2 - v^2, 2uv)$ be the quadratic map, and let $\vec{C}$ be any continuously differentiable oriented curve in the (u,v)-plane that avoids the origin. Show that

$$\int_{\mathbf{f}(\vec{C})} \frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy = 2\int_{\vec{C}} \frac{-v}{u^2+v^2}\,du + \frac{u}{u^2+v^2}\,dv$$

and conclude that

$$\int_{\mathbf{f}(\vec{C})} \frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy = 2\arctan\left(\frac{v}{u}\right)\Bigg|_{\text{start of }\vec{C}}^{\text{end of }\vec{C}}.$$