5 The Chain Rule

The Chain Rule

We know that the Chain Rule for functions of a single variable gives the rule for differentiating a composite function: If $y = f(x)$ and $x = g(t)$, where f and g are differentiable functions then y is indirectly a differentiable function of t and

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

The Chain Rule

The Chain Rule : Case 1

For functions of more than one variable, the Chain Rule has several versions, each of them giving a rule for differentiating a composite function.

The first version (Theorem 1) deals with the case where $z = f(x, y)$ and each of the variables x and y is, in turn, a function of a variable t.

This means that z is indirectly a function of t, $z = f(g(t), h(t))$, and the Chain Rule gives a formula for differentiating z as a function of t. We assume that f is differentiable.

The Chain Rule: Case 1

We know that this is the case when f_x and f_y are continuous.

1 The Chain Rule (Case 1) Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = g(t)$ and $y = h(t)$ are both differentiable functions of t. Then z is a differentiable function of t and

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

Since we often write $\frac{\partial z}{\partial x}$ in place of $\frac{\partial f}{\partial x}$, we can rewrite the Chain Rule in the form

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Example

If $z = x^2 y + 3xy^4$, where $x = \sin 2t$ and $y = \cos t$, find $\frac{dz}{dt}$ when $t = 0$.

Solution:

The Chain Rule gives

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

$$= (2xy + 3y^4)(2\cos 2t) + (x^2 + 12xy^3)(-\sin t)$$

It's not necessary to substitute the expressions for x and y in terms of t.

Example - Solution

We simply observe that when $t = 0$, we have $x = \sin 0 = 0$ and $y = \cos 0 = 1$.

Therefore

$$\left.\frac{dz}{dt}\right|_{t=0} = (0+3)(2\cos 0) + (0+0)(-\sin 0) = 6$$

The Chain Rule: Case 2

The Chain Rule: Case 2

We now consider the situation where $z = f(x, y)$ but each of x and y is a function of two variables s and t: $x = g(s, t)$, $y = h(s, t)$.

Then z is indirectly a function of s and t and we wish to find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$.

We know that in computing $\frac{\partial z}{\partial t}$ we hold s fixed and compute the ordinary derivative of z with respect to t.

Therefore we can apply Theorem 1 to obtain

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

The Chain Rule: Case 2

A similar argument holds for $\dfrac{\partial z}{\partial s}$ and so we have proved the following version of the Chain Rule.

2 The Chain Rule (Case 2) Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = g(s, t)$ and $y = h(s, t)$ are differentiable functions of s and t. Then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

Case 2 of the Chain Rule contains three types of variables: s and t are **independent** variables, x and y are called **intermediate** variables, and z is the **dependent** variable.

Example

If $z = e^x \sin y$, where $x = st^2$ and $y = s^2 t$, find $\dfrac{\partial z}{\partial s}$ and $\dfrac{\partial z}{\partial t}$.

Solution:

Applying Case 2 of the Chain Rule, we get

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} = \left(e^x \sin y\right)\left(t^2\right) + \left(e^x \cos y\right)\left(2st\right)$$

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} = \left(e^x \sin y\right)\left(2st\right) + \left(e^x \cos y\right)\left(s^2\right)$$

Example – Solution

If we wish, we can now express $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$ solely in terms of *s* and *t* by substituting $x = st^2, y = s^2t,$ to get

$$\frac{\partial z}{\partial s} = t^2 e^{st^2} \sin\left(s^2 t\right) + 2ste^{st^2} \cos\left(s^2 t\right)$$

$$\frac{\partial z}{\partial t} = 2ste^{st^2} \sin\left(s^2 t\right) + s^2 e^{st^2} \cos\left(s^2 t\right)$$

The Chain Rule: Case 2

Notice that Theorem 2 has one term for each intermediate variable and each of these terms resembles the one-dimensional Chain Rule in Equation 1.

To remember the Chain Rule, it's helpful to draw the **tree diagram** in Figure 2.

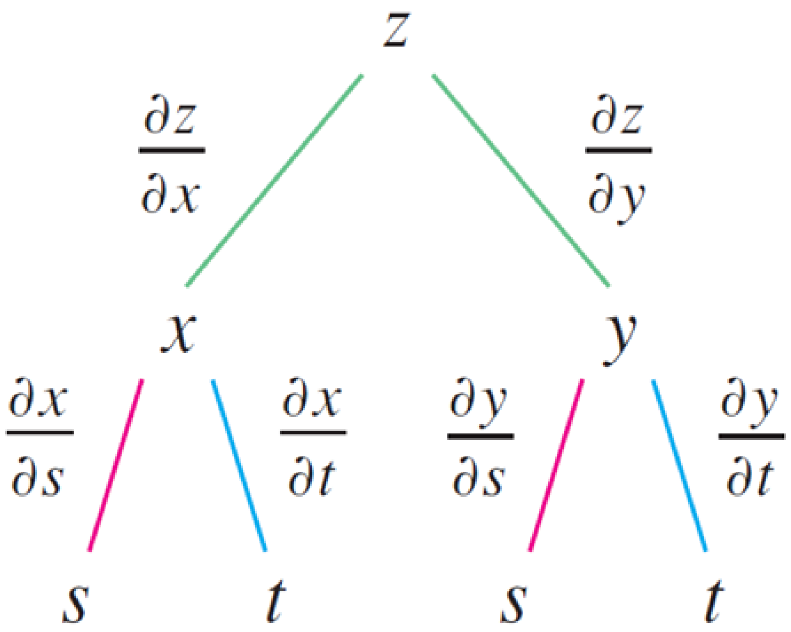


Figure 2

The Chain Rule: Case 2

We draw branches from the dependent variable *z* to the intermediate variables *x* and *y* to indicate that *z* is a function of *x* and *y*. Then we draw branches from *x* and *y* to the independent variables *s* and *t*.

On each branch we write the corresponding partial derivative. To find $\frac{\partial z}{\partial s}$, we find the product of the partial derivatives along each path from *z* to *s* and then add these products:

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

Similarly, we find $\frac{\partial z}{\partial t}$ by using the paths from *z* to *t*.

6 Directional Derivatives and the Gradient Vector

Directional Derivatives and the Gradient Vector

In this section we introduce a type of derivative, called a *directional derivative*, that enables us to find the rate of change of a function of two or more variables in any direction.

Directional Derivatives

Directional Derivatives

We know that if $z = f(x, y)$, then the partial derivatives f_x and f_y are defined as

1

$$f_x(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

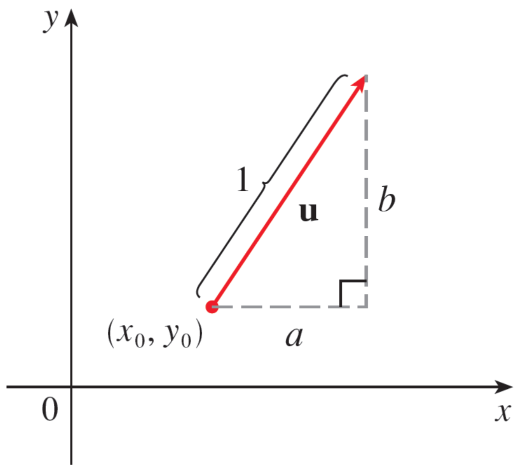
$$f_y(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}$$

and represent the rates of change of z in the x- and y-directions, that is, in the directions of the unit vectors **i** and **j**.

Directional Derivatives

Suppose that we now wish to find the rate of change of z at (x_0, y_0) in the direction of an arbitrary unit vector $\mathbf{u} = \langle a, b \rangle$. (See Figure 2.)

To do this we consider the surface S with the equation $z = f(x, y)$ (the graph of f) and we let $z_0 = f(x_0, y_0)$. Then the point $P(x_0, y_0, z_0)$ lies on S.



A unit vector $\mathbf{u} = \langle a, b \rangle$

Figure 2

Directional Derivatives

The vertical plane that passes through P in the direction of $\mathbf{u}$ intersects S in a curve C. (See Figure 3.)

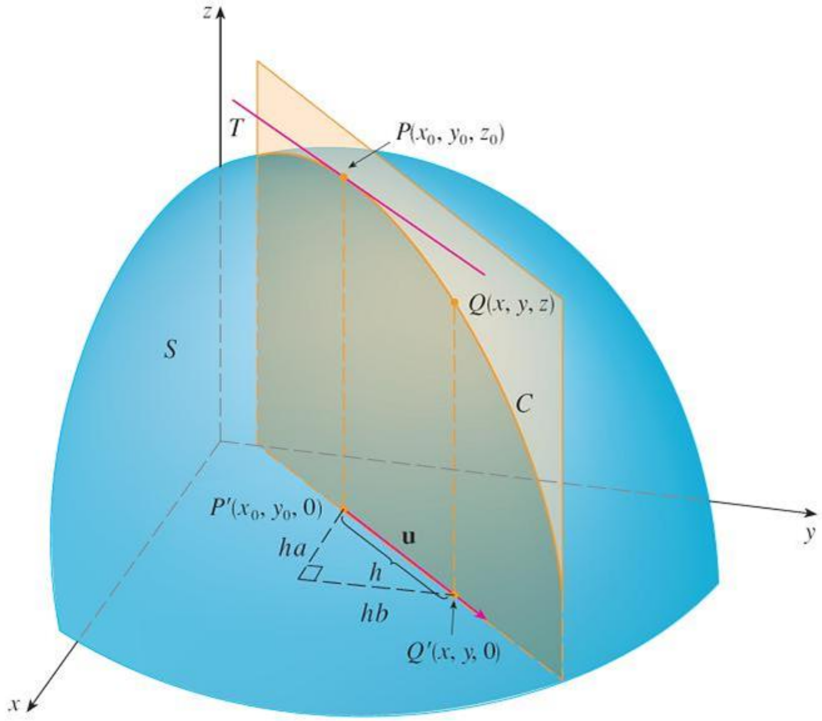


Figure 3

Directional Derivatives

The slope of the tangent line *T* to *C* at the point *P* is the rate of change of *z* in the direction of **u**.

If *Q*(*x*, *y*, *z*) is another point on *C* and P', Q' are the projections of *P*, *Q* onto the *xy*-plane, then the vector $\overrightarrow{P'Q'}$ is parallel to **u** and so

$$\overrightarrow{P'Q'} = h\mathbf{u} = \langle ha, hb \rangle$$

for some scalar *h*. Therefore $x - x_0 = ha$, $y - y_0 = hb$, so $x = x_0 + ha$, $y = y_0 + hb$, and

$$\frac{\Delta z}{h} = \frac{z - z_0}{h} = \frac{f(x_0 + ha,\ y_0 + hb) - f(x_0, y_0)}{h}$$

Directional Derivatives

If we take the limit as $h \to 0$, we obtain the rate of change of z (with respect to distance) in the direction of **u**, which is called the directional derivative of f in the direction of **u**.

2 Definition The **directional derivative** of f at (x_0, y_0) in the direction of a unit vector $\mathbf{u} = \langle a, b \rangle$ is

$$D_{\mathbf{u}} f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha,\ y_0 + hb) - f(x_0, y_0)}{h}$$

if this limit exists.

Directional Derivatives

By comparing Definition 2 with Equations 1, we see that if $\mathbf{u} = \mathbf{i} = \langle 1, 0 \rangle$, then $D_{\mathbf{i}}f = f_x$ and if $\mathbf{u} = \mathbf{j} = \langle 0, 1 \rangle$, then $D_{\mathbf{j}}f = f_y$.

In other words, the partial derivatives of f with respect to x and y are just special cases of the directional derivative.

Directional Derivatives

When we compute the directional derivative of a function defined by a formula, we generally use the following theorem.

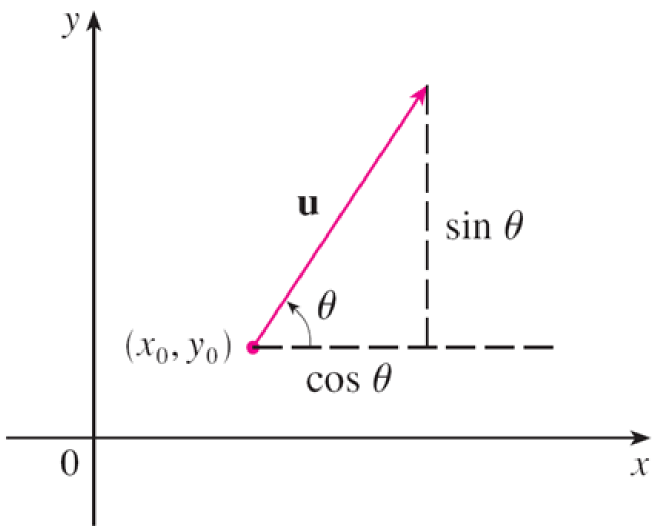
3 Theorem If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $\mathbf{u} = \langle a, b \rangle$ and

$$D_{\mathbf{u}} f(x, y) = f_x(x, y)a + f_y(x, y)b$$

Directional Derivatives

If the unit vector **u** makes an angle θ with the positive *x*-axis (as in Figure 5), then we can write $\mathbf{u} = \langle \cos\theta,\ \sin\theta \rangle$ and the formula in Theorem 3 becomes

$$\textbf{6} \qquad D_{\mathbf{u}} f(x, y) = f_x(x, y)\cos\theta + f_y(x, y)\sin\theta$$



A unit vector $\mathbf{u} = \langle \cos\theta, \sin\theta \rangle$

Figure 5

The Gradient Vector

Directional Derivatives

When we compute the directional derivative of a function defined by a formula, we generally use the following theorem.

3 Theorem If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $\mathbf{u} = \langle a, b \rangle$ and

$$D_{\mathbf{u}} f(x, y) = f_x(x, y)a + f_y(x, y)b$$

The Gradient Vector

Notice from Theorem 3 that the directional derivative of a differentiable function can be written as the dot product of two vectors:

$$\textbf{7} \qquad \begin{aligned} D_{\mathbf{u}} f(x, y) &= f_x(x, y)a + f_y(x, y)b \\ &= \langle f_x(x, y),\, f_y(x, y) \rangle \cdot \langle a, b \rangle \\ &= \langle f_x(x, y),\, f_y(x, y) \rangle \cdot \mathbf{u} \end{aligned}$$

The first vector in this dot product occurs not only in computing directional derivatives but in many other contexts as well.

So we give it a special name (the *gradient* of f) and a special notation (**grad** f or ∇f, which is read "del f").

The Gradient Vector

8 Definition If f is a function of two variables x and y, then the **gradient** of f is the vector function ∇f defined by

$$\nabla f(x,y) = \langle f_x(x,y),\ f_y(x,y) \rangle = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$

Example 3

If $f(x,y) = \sin x + e^{xy}$, then

$$\begin{aligned} \nabla f(x,y) &= \left\langle f_x, f_y \right\rangle \\ &= \left\langle \cos x + ye^{xy},\ xe^{xy} \right\rangle \end{aligned}$$

and $\qquad \nabla f(0,1) = \left\langle 2,0 \right\rangle$

The Gradient Vector

With the notation for the gradient vector, we can rewrite Equation 7 for the directional derivative of a differentiable function as

$$\mathbf{9} \quad D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot \mathbf{u}$$

This expresses the directional derivative in the direction of a unit vector **u** as the scalar projection of the gradient vector onto **u**.