LESSON 10

The Chain Rule

Contents

10.1 The Chain Rule, Case 1

Let us recall that, given a function $f(x) = g(h(x))$ that is differentiable on some continuous interval (a, b), then $f'(x) = g'(h(x)) \cdot h'(x)$. Or, in Leibniz notation

$$\frac{df}{dx} = \frac{dg}{dh} \cdot \frac{dh}{dx}.$$

This could be used to differentiate a function such as $f(x) = \sin(x^2)$ where we recognize this as a composite function with *sine* being the "outside" function and x^2 being the inside function. The usefulness of the chain rule was also apparent when differentiating an equation in x and y, where we take the derivative with respect to x and keep in mind that y is actually a function of x. For example:

$$\begin{aligned} xy^2 &= x^2 + 2y \\ \frac{d}{dx}(xy^2) &= \frac{d}{dx}(x^2 + 2y) \\ y^2 + x \cdot (2y \cdot y') &= 2x + 2y' \\ y^2 + 2xy \cdot y' &= 2x + 2y' \end{aligned}$$

Here we note that the chain rule was used on the derivative of y^2 which is really the composite function $(y(x))^2$ with the squaring being the outside function and $y(x)$ the inside function so the derivative of y^2 with respect to x is $2y \cdot y'$.

The first case of the chain rule we will look at for our purposes has actually already been explored minimally in applications of rates of change. Let's take a look at the following example:

Example 10.1.1 Two cars are traveling to a common destination. The first car, car A, is traveling due north at 60 mi/hr while the second car, car B, is traveling due west at 50 mi/hr. At what rate are the cars approaching each other when car A is 0.3 mi and car B is 0.4 mi from their destination? Let x be the distance car A is from their destination, y be the distance car B is from the destination, and z be the distance between the cars.

We note that, in the above example, the distance between the two cars can be described by the equation $z = \sqrt{x^2 + y^2}$ which is actually a function in two variables so we can investigate it using multi-variable calculus. Moreover, we note that x and y are both functions of time. Specifically $x(t) = -60t + x_0$ and $y(t) = -50t + y_0$ for some initial distances, x_0 and y_0 from their destination. Here we were looking for $\frac{dz}{dt}$ which we did by taking the derivative of both sides with respect to t by thinking of all variables as functions of time. Specifically, we got

$$\frac{dz}{dt} = \frac{x\frac{dx}{dt} + y\frac{dy}{dt}}{\sqrt{x^2 + y^2}}.$$

This is exactly case 1 of the chain rule for multi-variable functions. We state this rule explicitly here:

Theorem 10.1.1

Suppose that $z = f(x, y)$ is differentiable in x and y while $x = g(t)$ and $y = h(t)$ are differentiable in t. Then z is a differentiable function of t and is found to be

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

Proof : The proof will done after the following example and exercise.

Example 10.1.2 Given $z = \sqrt{x^2 + y^2}$, where $x(t) = -60t + 3$ and $y(t) = -50t + 2.65$, find $\left.\frac{dz}{dt}\right|_{t=0.045}$.

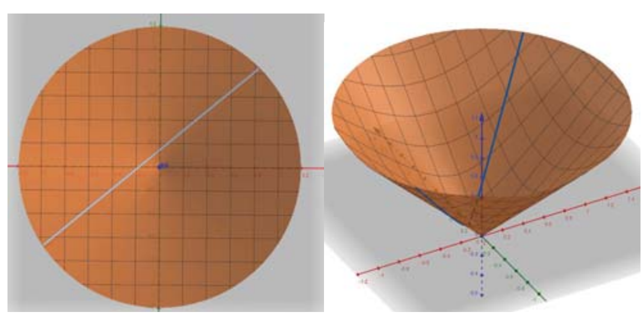

Figure 10.1.1: Graph of $z = \sqrt{x^2 + y^2}$, where $x(t) = -60t + 3$ and $y(t) = -50t + 2.65$
View in Geogebra:
https://www.geogebra.org/3d/ptjmrryu

Notice that in Figure 10.1.1 we can see how to interpret having a function in two variables with each variable being a function of another variable. Looking down on the xy-plane we see the plane curve described by the parametric equations $x(t) = -60t+3$ and $y(t) = -50t+2.65$. The z-value is then taken on the surface of the cone as the x and y-values travel along the plane curve. Thus dz/dt is the rate of change of z with respect to t as you move along the space curve shown.

Exercise 10.1.1 Let $z = x^2 \sin(y)$, $x(t) = t^2 + 1$, and $y = e^t$. Find $\left.\dfrac{dz}{dt}\right|_{t=0}$.

Proof of Theorem 10.1.1: In Lesson 9.2 it was covered that the change in z, given a function $z = f(x, y)$, is

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y$$

where ϵ_1 and ϵ_2 go to 0 as Δx and Δy go to zero.

Now, $\Delta x = g(t + \Delta t) - g(t)$ is the change in x induced by a change, Δt in t and $\Delta y = h(t + \Delta t) - h(t)$ is the change in y induced by Δt. Thus, if $\Delta t \to 0$, then Δx and Δy both go to zero as well. In turn, if $\Delta t \to 0$, then ϵ_1 and ϵ_2 both go to zero as well. Therefore:

$$\begin{aligned}
\frac{dz}{dt} &= \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\dfrac{\partial f}{\partial x}\Delta x + \dfrac{\partial f}{\partial y}\Delta y + \epsilon_1 \Delta x + \epsilon_2 \Delta y}{\Delta t} \\
&= \lim_{\Delta t \to 0} \left[\frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \epsilon_1 \frac{\Delta x}{\Delta t} + \epsilon_2 \frac{\Delta y}{\Delta t}\right] \\
&= \frac{\partial f}{\partial x} \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y} \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} + \left(\lim_{\Delta t \to 0} \epsilon_1\right) \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} + \left(\lim_{\Delta t \to 0} \epsilon_2\right) \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} \\
&= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + 0 \cdot \frac{dx}{dt} + 0 \cdot \frac{dy}{dt} \\
&= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}
\end{aligned}$$

■

10.2 The Chain Rule, Case 2

We now want to look at, keeping $z = f(x,y)$ as a function in two variables, look at how to handle when $x = g(s,t)$ and $y = h(s,t)$ are themselves functions of two variables. Here we can find *partial* derivatives of z with respect to s and t individually. In a case like this, if we plug in $g(s,t)$ for x and $h(s,t)$ for y into an expression for z, we would still have z as a function of two variables so we wouldn't be able to get a regular derivative out, only the partial derivatives.

Theorem 10.2.1

Suppose $z = f(x,y)$ is differentiable in x and y while $x = g(s,t)$ and $y = h(s,t)$ are differentiable in s and t. Then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

Example 10.2.1 Let $z = x^2\sin(y)$, $x = st^2$, and $y = e^{st}$. Use the chain rule to find $\dfrac{\partial z}{\partial t}$ and $\dfrac{\partial z}{\partial s}$. What is $\left.\dfrac{\partial z}{\partial t}\right|_{t=.1 \text{ and } s=-0.5}$?

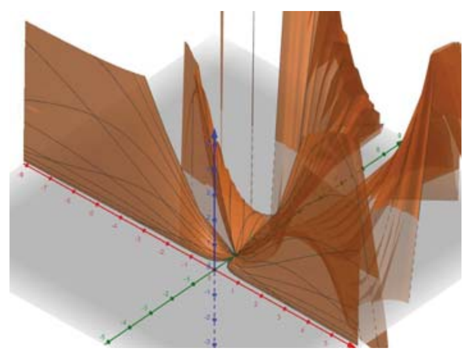

Figure 10.2.1: Graph of $z = x^2\sin(y)$, where $x(s,t) = st^2$ and $y(s,t) = e^{st}$
View in Geogebra:
https://www.geogebra.org/3d/squvuymr

If we want to interpret these derivatives graphically, note that $z = f(x,y)$ will give us a surface defined for all x and y in $\mathbb{R}$. However, looking at $x(s,t) = st^2$ and $y(s,t) = e^{st}$ you can start to realize that the points $(x,y) \in \mathbb{R}^2$ that can be reached via s and t is not the entire plane. So $z = f(g(s,t),h(s,t))$ is a subset of the surface obtained from when we look at $z = f(x,y)$ where x and y can be any real numbers. When we graph $f(g(s,t),h(s,t))$ we see the gridlines created by s and t and the partial derivatives of z with respect to s and t are thus the rates at which z is changing as we travel along these respective gridlines.

For your exercise, you are given a function $w = w(x, y, z)$ in three variables while $x = f(s, t)$, $y = g(s, t)$, and $z = h(s, t)$ are each functions of two variables. You are asked to find the partial derivatives of w with respect to s and t. While this is slightly more complicated than the previous example, the methodology is no different. Here the generalized chain rule is given in case you would like to look it over before doing the next exercise, but it shouldn't be necessary.

Theorem 10.2.2

Suppose the w is a function of n-variables, x_1, x_2, ... , x_n and that each x_i is a function of m-variables t_1, t_2, ... , t_m. Then the partial derivatives of w with respect to each t_i individually are

$$\frac{\partial w}{\partial t_i} = \frac{\partial w}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial w}{\partial x_2}\frac{\partial x_2}{\partial t_i} + ... + \frac{\partial w}{\partial x_n}\frac{\partial x_n}{\partial t_i}.$$

Exercise 10.2.1 Suppose $w = x^2 + y^2 - z^2$ while $x(s, t) = \sin(s)$, $y(s, t) = \cos(s)$, and $z(s, t) = t$. Determine $\frac{\partial w}{\partial s}$, and $\frac{\partial w}{\partial t}$.

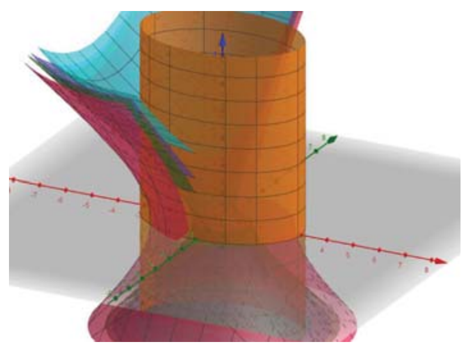

Figure 10.2.2: Graph of level surfaces of $w = x^2 + y^2 - z^2$ and $\mathbf{r}(s, t) = \langle \sin(s), \cos(s), t\rangle$
View in Geogebra: https://www.geogebra.org/3d/sguvec6v

To interpret these partial derivatives, take a look at the graph in figure 10.2.2. Here we show a cutout of level surfaces of $w = x^2 + y^2 - z^2$ where each level surface represents a different value of w. The circular cylinder is the surface generated by the parametric equations in two variables $x(s, t) = \sin(s)$, $y(s, t) = \cos(s)$, and $z(s, t) = t$. The partial derivatives of w with respect to s and t are the rates at which we move from one level surface of w to another along the gridlines of the circular cylinder given by s and t.

10.3 Implicit Differentiation

In single-variable calculus we found the derivatives of implicit equations in x and y by assuming y is a function of x and taking the derivative of both sides with respect to x to find $\frac{dy}{dx}$. See the introduction example 10.1.0 above. We now find a formula and often easier way to find $\frac{dy}{dx}$.

Suppose we have an equation in x and y so that x and y are related to each other. This equation could be an explicit or implicit equation, but either way we could set the equation to 0 by moving everything to one side. We then have an equation of the form $f(x, y) = 0$. Looking for $\frac{dy}{dx}$ we take the derivative of both sides with respect to x and apply the chain rule from theorem 10.1.1 to the lefthand side to obtain

$$\frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0.$$

Noting that $dx/dx = 1$, we solve for $\frac{dy}{dx}$ and get

Theorem 10.3.1

Given an equation $f(x, y) = 0$ where f is partially differentiable in x and y, then

$$\frac{dy}{dx} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} = -\frac{f_x}{f_y}$$

so long as $\frac{\partial f}{\partial y} \neq 0$.

Example 10.3.1 If $\cos(x-y) = xe^y$, find $\frac{dy}{dx}$ both the way that was shown in single-variable calculus and using theorem 10.3.1.

This same technique can be used for equations of more than two variables. Say we have an equation in three variables, say x, y, and z, and we set it equal to zero by moving everything to one side then we get an equation of the form $f(x, y, z) = 0$. Supposing we want $\dfrac{\partial z}{\partial x}$. We can take this partial derivative of both sides and apply the chain rule to the left to obtain

$$\frac{\partial f}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial f}{\partial x}\frac{\partial z}{\partial x} = 0$$

and then solve for $\dfrac{\partial z}{\partial x}$ to obtain

$$\frac{\partial z}{\partial x} = -\frac{\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}} = -\frac{f_x}{f_z}$$

since $\dfrac{\partial x}{\partial x} = 1$ and $\dfrac{\partial y}{\partial x} = 0$. Note $f_z \neq 0$ for this formula to be applicable.

Theorem 10.3.2

Given an equation $f(x, y, z) = 0$ where f is partially differentiable in x, y, and z then

$$\frac{\partial z}{\partial x} = -\frac{f_x}{f_z} \quad \text{and} \quad \frac{\partial z}{\partial y} = -\frac{f_y}{f_z}$$

so long as $f_z \neq 0$.

Exercise 10.3.1 Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ if $yz = \ln(x + z)$ using theorem 10.3.2.