Multivariable chain rule, simple version

Google Classroom

The chain rule for derivatives can be extended to higher dimensions. Here we see what that looks like in the relatively simple case where the composition is a single-variable function.

Background

- Single variable chain rule
- The gradient
- Derivatives of vector valued functions

What we're building to

- Given a multivariable function $f(x, y)$, and two single variable functions $x(t)$ and $y(t)$, here's what the multivariable chain rule says:

$$\underbrace{\frac{d}{dt} f(x(t), y(t))}_{\text{Derivative of composition function}} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$$

- Written with vector notation, where $\vec{\mathbf{v}}(t) = \left[\begin{array}{c} x(t) \\ y(t) \end{array} \right]$, this rule has a very elegant form in terms of the gradient of f and the vector-derivative of $\vec{\mathbf{v}}(t)$.

$$\underbrace{\frac{d}{dt} f(\vec{\mathbf{v}}(t))}_{\text{Derivative of composition function}} = \overbrace{\nabla f \cdot \vec{\mathbf{v}}'(t)}^{\text{Dot product of vectors}}$$

A more general chain rule

As you can probably imagine, the multivariable chain rule generalizes the chain rule from single variable calculus. The single variable chain rule tells you how to take the derivative of the composition of two functions:

$$\frac{d}{dt} f(g(t)) = \frac{df}{dg} \frac{dg}{dt} = f'(g(t))g'(t)$$

What if instead of taking in a one-dimensional input, t, the function f took in a two-dimensional input, (x, y)?

$$f(x, y) = \ldots \text{some expression of } x \text{ and } y \ldots$$

Well, in that case, it wouldn't make sense to compose it with a scalar-valued function $g(t)$. Instead, let's say there are two separate scalar-valued functions $x(t)$ and $y(t)$, and we plug these in as the coordinates of f. The overall composition will be a single variable function, with a single-number input t, and a single-number output $f(x(t), y(t))$, as shown in this diagram:

$$\begin{array}{rccc}
\text{One final output} & & f(x(t), y(t)) & \\
 & \nearrow & & \nwarrow \\
\text{Two intermediate outputs} & x(t) & & y(t) \\
 & \nwarrow & & \nearrow \\
\text{One input} & & t &
\end{array}$$

There is still a chain rule that lets you compute the derivative of this new single-variable function $f(x(t), y(t))$, and it involves the partial derivatives of f:

$$\underbrace{\frac{d}{dt}}_{\substack{\text{This is an ordinary derivative,} \\ \text{not a partial derivative } \frac{\partial}{\partial t} \\ \text{because the total composition} \\ \text{has one input and one output.}}} f(x(t), y(t)) = \underbrace{\overbrace{\frac{\partial f}{\partial x}}^{\substack{\text{How } f \text{ changes} \\ \text{due to a tiny} \\ \text{change in } x}} \overbrace{\frac{dx}{dt}}^{\substack{\text{How } x \text{ changes} \\ \text{due to a tiny} \\ \text{change in } t}}}_{\substack{\text{Total change in } f \\ \text{due to the influence} \\ t \text{ has on } x}} +$$

Keep in mind, an expression like $\frac{\partial f}{\partial x} \frac{dx}{dt}$ is shorthand for

$$\frac{\partial f}{\partial x}(x(t), y(t)) \frac{dx}{dt}(t)$$

That is, both are functions of t, but $\frac{\partial f}{\partial x}$ is evaluated via the intermediate functions $x(t)$ and $y(t)$.

Written with vector notation

Rather than thinking of $x(t)$ and $y(t)$ as being separate functions, it's common to package them together into a single, vector-valued function:

$$\vec{\mathbf{v}}(t) = \left[\begin{array}{c} x(t) \\ y(t) \end{array}\right]$$

Then instead of writing the composition as $f(x(t), y(t))$, you can write it as $f(\vec{\mathbf{v}}(t))$.

With this notation, the multivariable chain rule can be written more compactly as a dot product between the gradient of f and the vector-derivative of $\vec{\mathbf{v}}(t)$:

$$\frac{d}{dt} f(\vec{\mathbf{v}}(t)) = \underbrace{\frac{\partial f}{\partial x}(\vec{\mathbf{v}}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{\mathbf{v}}(t))\frac{dy}{dt}}_{\text{Rewrite this sum as a dot product}}$$

$$= \underbrace{\left[\begin{array}{c} \frac{\partial f}{\partial x}(\vec{\mathbf{v}}(t)) \\ \\ \frac{\partial f}{\partial y}(\vec{\mathbf{v}}(t)) \end{array}\right]}_{\nabla f(\vec{\mathbf{v}}(t))} \cdot \underbrace{\left[\begin{array}{c} \frac{dx}{dt} \\ \\ \frac{dy}{dt} \end{array}\right]}_{\vec{\mathbf{v}}'(t)}$$

$$= \nabla f(\vec{\mathbf{v}}(t)) \cdot \vec{\mathbf{v}}'(t)$$

Written like this, the analogy with the single-variable derivative is clearer.

$$\frac{d}{dt} f(g(t)) = f'(g(t))g'(t) = \frac{df}{dg} \cdot \frac{dg}{dt}$$

The gradient ∇f plays the role of the derivative of f, and the vector derivative $\vec{\mathbf{v}}'(t)$ plays the role as the ordinary derivative of g.

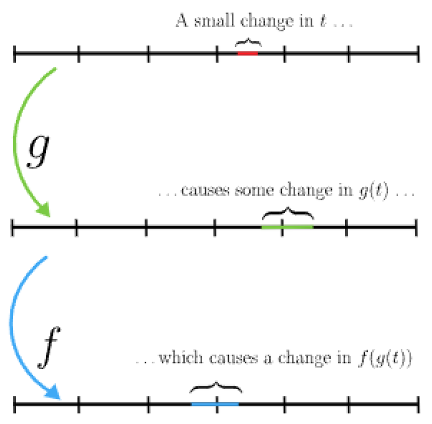
Intuition for why the chain rule works

As a warm up, consider the single variable chain rule for a composition like $f(g(t))$. Here's how I like to understand that composition:

- First, g maps a point t on the number line to another point $g(t)$ the number line.

- Then f comes in and maps the point $g(t)$ to yet another point on the number line, $f(g(t))$

Understanding the derivative of $f(g(t))$ requires understanding how a tiny change in t changes the final output.



So let's dive into what the chain rule is really saying.

$$\frac{d}{dx} f(g(t)) = \frac{df}{dg} \cdot \frac{dg}{dt}$$

- The term $\frac{dg}{dt}$ represents how a tiny change in t influences the intermediate output, $g(t)$. [Hide explanation]

  $$\frac{dg}{dt}(t_0) = \lim_{h \to 0} \frac{g(t_0 + h) - g(t_0)}{h}$$

  If we move from an input t_0 to an input $t_0 + h$ for a tiny value of h, the output of g will change by about

  $$\left(\frac{dg}{dt}(t_0)\right) \cdot h$$

  You can imagine multiplying both sides by h in the limit above. The smaller h is, the smaller the error between this value and the real change $g(t_0 + h) - g(t_0)$.

- The term $\frac{df}{dg}$ represents how a tiny change in g influences the final output $f(g(t))$. [Hide explanation]

$$\frac{df}{dg}(g(t_0)) = \lim_{k \to 0} \frac{f(g(t_0)+k) - f(g(t_0))}{k}$$

If we move from a point $g(t_0)$ to $g(t_0) + k$ for some small k, the output of f will change by about

$$\left(\frac{df}{dg} g(t_0)\right) \cdot k.$$

- The total change in f due to a small change in t is then the product of both these influences. [Hide explanation]

Combining the previous two explanations, if the change in g was caused by a change of h to t_0, then this change k to $g(t_0)$ is about

$$k \approx \left(\frac{dg}{dt}(t_0)\right) \cdot h.$$

Plugging in this value for k, the resulting change to f that we are interested in is $\left(\frac{df}{dg} g(t_0)\right) \cdot \left(\frac{dg}{dx}(t_0)\right) \cdot h$

The smaller h is, the more true this estimate is.

Challenge question: Make this an even tighter argument by evaluating

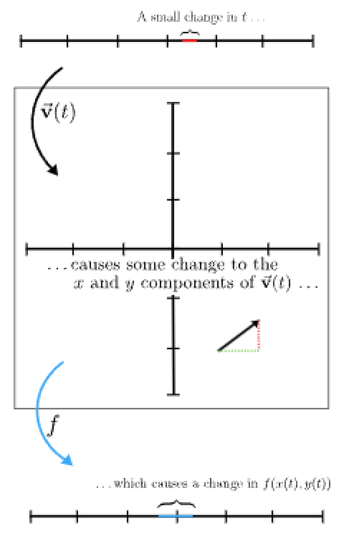
$$\lim_{h \to 0} \frac{f(g(t_0+h)) - f(g(t_0))}{h}$$

using the limit-based definitions of $\frac{df}{dg}$ and $\frac{dg}{dt}$.

(Hint: Start by replacing $g(t_0 + h)$ with a certain limiting value).

Extend this intuition to more dimensions

The intuition is similar for the multivariable chain rule. You can think of $\vec{\mathbf{v}}$ as mapping a point on the number line to a point on the xy-plane, and $f(\vec{\mathbf{v}}(t))$ as mapping that point back down to some place on the number line. The question is, how does a small change in the initial input t change the total output $f(\vec{\mathbf{v}}(t))$?



Let's break down what the multivariable chain rule is saying, spelling it out in terms of the component functions $x(t)$ and $y(t)$:

$$\frac{d}{dt} f(\vec{\mathbf{v}}(t)) = \frac{d}{dt} f(x(t), y(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

- The term $\frac{dx}{dt}$ represents how a tiny change in t influences the intermediate output $x(t)$. [Hide explanation]

> For small values of h,
>
> $$x(t_0 + h) - x(t_0) \approx \left(\frac{dx}{dt}(t_0)\right) h$$

- Likewise the term $\frac{dy}{dt}$ represents how a tiny change in t influences the second intermediate output $y(t)$. [Hide explanation]

> For small values of h,
>
> $$y(t_0 + h) - y(t_0) \approx \left(\frac{dy}{dt}(t_0)\right) h$$

- The term $\frac{\partial f}{\partial x}$ represents how a tiny change **to the x-component** of an input to f influences its output, and similarly the term $\frac{\partial f}{\partial y}$ accounts for how a small change **to the y-component** of the input changes f. [Hide explanation]

> For small values of k,
>
> $$f(x_0 + k, y_0) - f(x_0, y_0) \approx \left(\frac{\partial f}{\partial x}(x_0, y_0)\right) k$$
> $$f(x_0, y_0 + k) - f(x_0, y_0) \approx \left(\frac{\partial f}{\partial y}(x_0, y_0)\right) k$$

- One way a small change to t influences $f(x(t), y(t))$ is that it first changes $x(t)$, which in turn changes f. This effect is captured in the product $\frac{\partial f}{\partial x}\frac{dx}{dt}$. [Hide explanation]

> The change in f due to the small change in $x(t)$, which in turn is caused by a small change t, is expressed like this:
>
> $$f(x(t_0+h), y(t_0)) - f(x(t_0), y(t_0))$$
>
> We know that $x(t_0+h)$ is well approximated as
>
> $$x(t_0+h) \approx x(t_0) + \left(\frac{dx}{dt}(t_0)\right) h$$
>
> So the quantity we care about now becomes
>
> $$f\left(x(t_0) + \left(\frac{dx}{dt}(t_0)\right) h, y(t_0)\right) - f(x(t_0), y(t_0))$$
>
> But this expression is just asking how a certain small change to the x coordinate of the input to f influences the output, which we know to be well approximated using the partial derivative with respect to x:
>
> $$\left(\frac{\partial f}{\partial x}(x(t_0), y(t_0))\right) \overbrace{\left(\frac{dx}{dt}(t_0)\right) h}^{\text{Small change to } x}$$

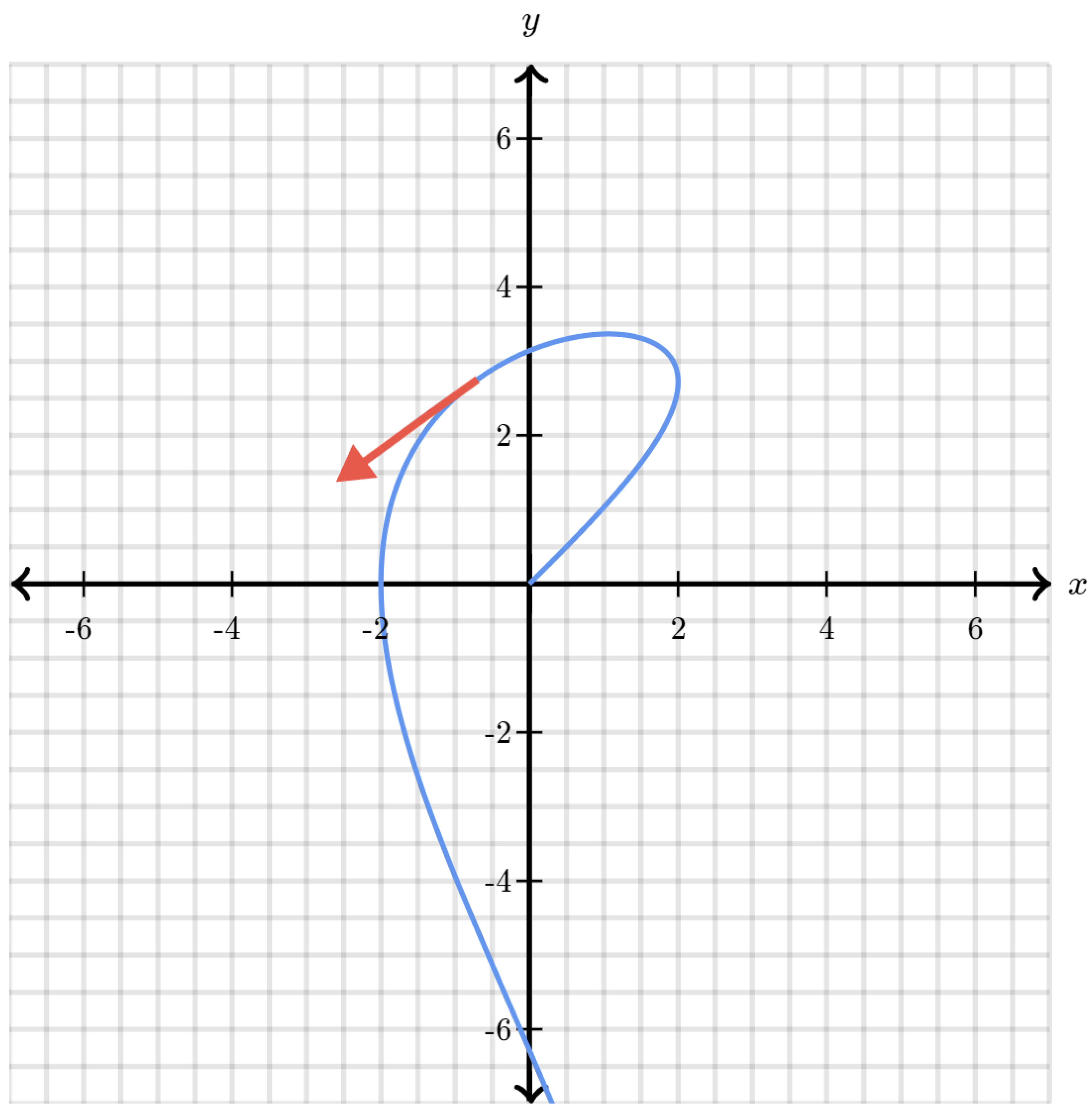
- The other way a change to t changes the output of $f(x(t), y(t))$ is by first changing the second intermediate output $y(t)$, which in turn affects the output of f. This effect is captured in the product $\frac{\partial f}{\partial y}\frac{dy}{dt}$.

- Adding these two products gives the total change in f.

Connection with directional derivative

You might notice that the dot product expression for the multivariable chain rule looks a lot like a directional derivative:

$$\nabla f(\vec{\mathbf{v}}(t)) \cdot \vec{\mathbf{v}}'(t)$$



In fact, that's exactly what it is! The derivative $\vec{\mathbf{v}}'(t_0)$ at a particular value t_0 gives a vector in the input space of f:

$$\vec{\mathbf{v}}'(t_0) = \begin{bmatrix} x'(t_0) \\ y'(t_0) \end{bmatrix}$$

If $\vec{\mathbf{v}}(t)$ is interpreted as a parametric path inside this space, perhaps thought of as the trajectory of a particle, the derivative at a particular point in time t_0 gives the velocity vector of this particle at that time.

With this interpretation, the chain rule tells us that the derivative of the composition $f(\vec{\mathbf{v}}(t))$ is the **directional derivative of f along the derivative of $\vec{\mathbf{v}}(t)$**.

This should make sense, because a tiny change by "dt" to t should, by the meaning of the derivative, cause a tiny change $d\vec{\mathbf{v}}$ to the output of $\vec{\mathbf{v}}(t)$. And the point of the directional derivative is that a tiny change $d\vec{\mathbf{v}}$ to the input of f should cause a change df as determined by $\dfrac{\partial f}{\partial \vec{\mathbf{v}}} = \nabla_{\vec{\mathbf{v}}} f$.

Example 1: With and without the new chain rule

Define $f(x, y)$ like this:

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

And define $\vec{\mathbf{v}}(t)$ like this:

$$\vec{\mathbf{v}}(t) = \begin{bmatrix} \cos(t) \\ \\ \sin(t) \end{bmatrix}$$

Find the derivative $\frac{d}{dt} f(\vec{\mathbf{v}}(t))$.

Solution without chain rule:

Before throwing our fancy new tool at the problem, it's worth pointing out that this is something we can solve by first writing out the composition as a single variable function of t:

$$f(\vec{\mathbf{v}}(t)) = f(\cos(t), \sin(t))$$

$$= \cos(t)^2 \sin(t)$$

Now you can take the ordinary derivative:

$$\frac{d}{dt} \cos(t)^2 \sin(t)$$

$$= \cos(t)^2 (\cos(t)) + 2\cos(t)(-\sin(t))\sin(t)$$

$$= \boxed{\cos^3(t) - 2\cos(t)\sin^2(t)}$$

But of course, the purpose of this example is to get a feel for what the chain rule feels like.

Solution using chain rule:

First, let's explicitly state the component functions of $\vec{\mathbf{v}}(t)$:

$$x(t) = \cos(t)$$

$$y(t) = \sin(t)$$

According to the chain rule,

$$\frac{d}{dt} f(\vec{\mathbf{v}}(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Taking the partial derivatives of $f(x, y) = x^2y$ and the ordinary derivatives of $x(t) = \cos(t)$, $y(t) = \sin(t)$, we get

$$\frac{\partial}{\partial x}(x^2y)\frac{d}{dt}(\cos(t)) + \frac{\partial}{\partial y}(x^2y)\frac{d}{dt}(\sin(t))$$

$$= (2xy)(-\sin(t)) + (x^2)(\cos(t))$$

We want everything in terms of t, so we plug in $x = \cos(t)$ and $y = \sin(t)$.

$$(2xy)(-\sin(t)) + (x^2)(\cos(t))$$

$$(2\cos(t)\sin(t))(-\sin(t)) + (\cos(t)^2)\cos(t)$$

$$= \boxed{-2\cos(t)\sin^2(t) + \cos^3(t)}$$

Reassuringly, this is the same as the answer we got without using the chain rule. You might be thinking that this new chain rule makes things unnecessarily complicated, and the dirty little secret is that for concrete computations like this one, it is often not needed.

However, it is useful for writing equations in terms of an unknown function, as the next example shows.

Example 2: Unknown function

Suppose the temperature across a two-dimensional region varies according to a function $T(x, y)$, which we do not know. You wander throughout this region, sampling temperatures as you go, and your x and y coordinates as functions of time are

$$x(t) = 30\cos(2t)$$

$$y(t) = 40\sin(3t)$$

In taking your measurements, you notice that the temperature never changes along your path. What can you say about the partial derivatives of T?

[Hide explanation]

The temperature you experience, as a function of time, is

$$T(x(t), y(t)) = T(30\cos(2t), 40\sin(3t))$$

The fact that the temperature you experience never changes means the derivative of this temperature, as a function of time, is 0. Using the multivariable chain rule, we can write the derivative of this function in terms of the partial derivatives of T:

$$\begin{aligned}\frac{d}{dt}T(x(t), y(t)) &= \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} \\ &= \frac{\partial T}{\partial x} \cdot 30(2)(-\sin(2t)) + \frac{\partial T}{\partial y} \cdot 40(3)\cos(3t) \\ &= -\frac{\partial T}{\partial x} \cdot 60\sin(2t) + \frac{\partial T}{\partial y} \cdot 120\cos(3t)\end{aligned}$$

More accurately, we should be evaluating these partial derivatives at $(x, y) = (30\cos(2t), 40\sin(3t))$, so the full expression for what we know about the unknown temperature function T is

$$\begin{aligned}0 = -&\frac{\partial T}{\partial x}(30\cos(2t), 40\sin(3t)) \cdot 60\sin(2t) + \\ &\frac{\partial T}{\partial y}(30\cos(2t), 40\sin(3t)) \cdot 120\cos(3t)\end{aligned}$$

Summary

- Given a multivariable function $f(x, y)$, and two single variable functions $x(t)$ and $y(t)$, here's what the multivariable chain rule says:

$$\underbrace{\frac{d}{dt}f(x(t), y(t))}_{\text{Derivative of composition function}} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

- Written with vector notation, where $\vec{\mathbf{v}}(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$, this rule has a very elegant form in terms of the gradient of f and the vector-derivative of $\vec{\mathbf{v}}(t)$.

$$\underbrace{\frac{d}{dt}f(\vec{\mathbf{v}}(t))}_{\text{Derivative of composition function}} = \overbrace{\nabla f \cdot \vec{\mathbf{v}}'(t)}^{\text{Dot product of vectors}}$$